DOCUMENT RESUME

ED 309 185 TM 013 656

AUTHOR	Knol, Dirk L.; Berger, Martijn P. F.
TITLE	Empirical Comparison between Factor Analysis and Item Response Models. Project Psychometric Aspects of Item Banking No. 37. Research Report 88-11.
INSTITUTION	Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE	Jul 88
NOTE	46p.
AVAILABLE FROM	Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Comparative Analysis; Computer Simulation; Computer Uses in Education; *Factor Analysis; Foreign Countries; *Latent Trait Theory; Mathematical Models; Multidimensional Scaling; Statistical Analysis; *Test Items
IDENTIFIERS	*Empirical Methods; Item Parameters; Tetrachoric Correlation

ABSTRACT

Many multidimensional item response theory (IRT) models have been proposed. A comparison is made between the so-called full information models and the models that use only pairwise information. Three multidimensional models described are: (1) the compensatory model of R. D. Bock and M. Aitken (1981) using the computer program TESTFACT; (2) a model based on R. P. McDonald's (1985) harmonic analysis using the program NOHARM; and (3) the computer program MAXLOG of R. L. McKinley and M. D. Reckase (1983). Five factor analysis procedures for dichotomous items are discussed. A simulation study was conducted to compare the various methods. The item parameters of four different sets of items were used with numbers of subjects set at 250, 500, and 1,000. Ten replications were generated for each set of item parameters and each sample size. All models were compared with respect to estimates of IRT and factor analysis parameters using six criteria in terms of mean squared differences between the known and estimated item parameters. The most striking result of the simulation study was that common factor analysis programs outperformed the more complex programs TESTFACT, MAXLOG, and NOHARM. It was apparent that a common factor analysis in the matrix of tetrachoric correlations yielded the best estimates. A procedure based on the mean squared residuals of the correlation matrix was also presented for assessing the dimensionality of the model. Nine tables present the data from the simulation study. A 45-item list of references is included. (SLD)

Empirical Comparison between Factor Analysis and Item Response Models

Research Report 88-11

Dirk L. Knol
Martijn P.F. Berger

Department of EDUCATION

Division of Educational Measurement and Data Analysis

University of Twente

Project Psychometric Aspects of Item Banking No.37

Colofon:
Typing: Dirk L. Knol
Cover design: Audiovisuele Sectie TOLAB Toegepaste Onderwijskunde
Printed by: Centrale Reproductie-afdeling

Empirical Comparison between Factor Analysis and Item Response Models

Dirk L. Knol

Martijn P.F. Berger

Empirical comparison between factor analysis and item response models / Dirk L. Knol & Martijn P.F. Berger - Enschede : University of Twente, Department of Education, July, 1988. - 38 pages

Abstract

Many multidimensional item response models have been proposed in literature. The models and various methods for estimating the item parameters are reviewed briefly. In a simulation study these methods are compared with respect to their estimates of the item parameters. It is concluded that a common factor analysis on the matrix of tetrachoric correlations yields the best estimates.

Additionally, a procedure based upon the mean squared residuals of the correlation matrix is presented for the assessment of the dimensionality of the model.

Key words: Common Factor Analysis, Dichotomous Variables, Item Response Theory, Multidimensionality, Tetrachoric Correlations.

Empirical Comparison between Factor Analysis and Item Response Models

Introduction

One of the main problems in constructing Rasch (1960) scales from a large number of dichotomously scored items is the multidimensionality of the itempool. Usually, the Rasch model does not fit the whole itempool. Procedures for constructing Rasch scales which start from the entire itempool are not very promising (cf. Knol, 1987b). A more promising procedure is to identify the main dimensions of the itempool and to start an (iterative) procedure on the different subsets of items separately. Verhelst (1983) and Knol (1987b) describe such iterative procedures.

To identify the main dimensions of an itempool, a multidimensional representation of the items can be useful. Many multidimensional models can be used for that purpose. Roughly, the models can be distinguished to the extent in which they make use of the information of the data matrix. For continuous variables, a classical common factor analysis (FA) on the matrix of product-moment correlations can be used. However, for dichotomous items, the matrix of pairwise (tetrachoric) correlations is not sufficient (Mood, Graybill & Boes, 1974, pp. 299-314). Therefore, several models have been proposed, which do use all the available information contained in the response vectors. Because these so-called full-information (cf. Bock, Gibbons & Muraki, 1985) models suffer from numerical difficulties, an approximation has been proposed by McDonald (1985).

The purpose of the present paper is to compare the so-called full-information models with the models which use only pairwise information. For simulated data, various estimations of the item parameters will be compared. Furthermore a procedure to estimate the dimensionality of the models will also be presented.

Firstly, a short review of multidimensional item response theory (IRT) models will be given. Then the various FA models for dichotomous variables are described.

Multidimensional IRT Models

Several multidimensional IRT models for dichotomous data have been proposed (Bock & Aitkin, 1981; Bock & Lieberman, 1970; McDonald, 1985; Mulaik, 1972; Rasch, 1961 Reckase, 1973; Sympson, 1978; Whiteley, 1980). Generally, the models can be classified into so-called compensatory models, which allow high ability values on one dimension to compensate for low abilities on other dimensions, and noncompensatory models. These last mentioned models (Sympson, 1978; Whiteley, 1980) do not allow high ability to compensate for low ability on other dimensions. Apart from the psychological meaningfulness of these models, the most important practical disadvantage of noncompensatory models is that no efficient algorithms for the parameter estimates are available.

The compensatory model of Bock and Aitkin (1981) is relatively simple and a marginal maximum likelihood (MML) procedure for the estimation of item parameters has been developed. Let $\underline{X} = (X_1,\ldots,X_n)'$ be a random vector of response patterns to n dichotomous items, where each X_i $(i = 1,\ldots,n)$ is defined as

(1) $$X_i = \begin{cases} 1, & \text{if item } i \text{ is correctly answered} \\ 0, & \text{otherwise} \end{cases}.$$

Under the (usual) assumption of local independence the marginal probability of the response vector $\underline{X} = \underline{x}$ is given by

(2) $$P(\underline{X} = \underline{x}) = \int \prod_{i=1}^{n} [p_i(\underline{\theta})]^{x_i} [1 - p_i(\underline{\theta})]^{1-x_i} g(\underline{\theta})d\underline{\theta},$$

where $p_i(\underline{\theta})$ is the item characteristic function (ICF) of item i, $g(\underline{\theta})$ is the density function of the unobserved m-component random vector of abilities $\underline{\theta}$, and the integration is taken over the entire multidimensional ability space. It is assumed that $\underline{\theta}$ is multivariate normally (MVN) distributed with mean $\underline{0}$ and covariance matrix I. In the multidimensional two-parameter normal ogive (M2PNO) model the ICF of item i is, given by

(3) $$p_i(\underline{\theta}) = F(\underline{\alpha}_i'\underline{\theta} - \beta_i),$$

where $\underline{\alpha}_i$ is the m x 1 vector of discrimination parameters for item i, β_i is the difficulty parameter for item i ($i=1,\ldots,n$) and F(.) is the cumulative standard normal distribution. An iterative procedure to obtain MML estimates of the item parameters via the EM algorithm (Dempster, Laird & Rubin, 1977) has been implemented in the computer program TESTFACT (Wilson, Wood & Gibbons, 1984).

An IRT model that uses only information contained in the pairwise proportions is based upon McDonald's (1985) harmonic analysis. A computer program NOHARM II (Fraser, 1988) is available in which the pairwise proportions $\pi_{ij} = P(X_i=1, X_j=1)$ are approximated by minimizing the unweighted least squares function

$$(4) \qquad f(A, \underline{\beta}) = \sum_{i<j}\sum [p_{ij} - \hat{\pi}_{ij}(\underline{\alpha}_i, \beta_i, \underline{\alpha}_j, \beta_j)]^2 ,$$

where $A = (\underline{\alpha}_1,\ldots,\underline{\alpha}_n)'$, $\underline{\beta} = (\beta_1,\ldots,\beta_n)'$, p_{ij} are the sample proportions and the ICF's are approximated by a third degree Hermite–Tchebycheff polynomial.

Because of the well-known relationship between the logistic distribution function L (cf. Mood, Graybill & Boes, 1974, p. 542) and the cumulative standard normal distribution function F

$$(5) \qquad |F(z) - L(1.7z)| < 0.01$$

for all z (Haley, 1952), it is possible to approximate the normal ogive ICF (3) by the logistic ICF

$$(6) \qquad p_i(\underline{\theta}) = \frac{\exp[c(\underline{\alpha}_i'\underline{\theta} - \beta_i)]}{1 + \exp[c(\underline{\alpha}_i'\underline{\theta} - \beta_i)]} = L[c(\underline{\alpha}_i'\underline{\theta} - \beta_i)] \; .$$

The computer program MAXLOG (McKinley & Reckase, 1983) yields estimations of the parameters of the multidimensional two-parameter logistic (M2PL) model. Because the program uses the method of joint ML estimation, problems such as the so-called drift of the discrimination parameters may be encountered and estimation may be cumbersome when the number of subjects N is large.

In all three programs mentioned above the numbers of variables and dimensions are limited. This makes the programs not very useful for large scale applications.

FA for Dichotomous Items

In FA for dichotomous variables (Christoffersson, 1975; Muthén, 1978), the response variables X_i are assumed to be governed by the unobserved continuous variables Y_i and thresholds τ_i as

$$(7) \qquad X_i = \begin{cases} 1, & \text{if } Y_i > \tau_i \\ 0, & \text{otherwise} \; , \end{cases}$$

where

(8) $\underline{Y} = \Lambda\underline{\theta} + \underline{E}$.

and $\underline{Y} = (Y_1, \ldots, Y_n)'$. Model (8) is the usual random factors FA model, the only difference being that Y is unobserved. Under the assumptions

(9a) $\underline{\theta} \sim MVN\ (0,\ I)$,

(9b) $\underline{E} \sim MVN\ (0,\ \Psi^2)$,

where Ψ^2 is a diagonal matrix with positive diagonal elements, and

(9c) $cov(\underline{\theta},\ \underline{E}) = \underline{0}$,

the covariance matrix Σ among $\underline{Y}$ variables is given by

(10) $\Sigma = \Lambda\Lambda' + \Psi^2$.

Hence,

(11) $\underline{Y} \sim MVN(\underline{0},\ \Lambda\Lambda' + \Psi^2)$.

In FA for dichotomous variables the marginal probability of response pattern $\underline{X} = \underline{x}$ is

(12) $$P(\underline{X} = \underline{x}) = \int_Z h(\underline{Y})d\underline{Y} ,$$

where $h(\cdot)$ is the MVN $(\underline{0}, \Lambda\Lambda' + \Psi^2)$ density and Z is the multidimensional integration region defined by the Cartesian product of Z_i, such that $Z_i = (\tau_i, \infty)$ if $X_i = 1$ and $Z_i = (-\infty, \tau_i)$ if $X_i = 0$.

Takane and De Leeuw (1987) showed the formal equivalence of the marginal likelihood (2) of the M2PNO model with $\underline{\theta} \sim MVN(\underline{0}, I)$ and the likelihood (12) of FA for dichotomous variables. The parameters of the IRT formulation $\underline{\alpha}_i$ and β_i ($i=1,\ldots,n$) can be expressed in terms of the parameters of the FA formulation $\underline{\lambda}_i$, τ_i and ψ_i as

(13a) $$\underline{\alpha}_i = \underline{\lambda}_i/\psi_i$$

and

(13b) $$\beta_i = \tau_i/\psi_i$$

(Takane & De Leeuw, 1987), where $\underline{\lambda}_i'$ denotes row i of Λ and ψ_i^2 is the i-th diagonal element of Ψ^2. Reversely, the FA parameters can be expressed in terms of the IRT parameters as

(14a) $$\underline{\lambda}_i = (1 + \underline{\alpha}_i'\underline{\alpha}_i)^{-\frac{1}{2}}\underline{\alpha}_i ,$$

(14b) $$\tau_i = (1 + \underline{\alpha}_i'\underline{\alpha}_i)^{-\frac{1}{2}}\beta_i$$

and

(14c) $\psi_i = (1 + \underline{\alpha}_i'\underline{\alpha}_i)^{-1/2}$,

cf. also Knol (1987a).

The parameters of the FA formulation can be estimated with the program LISCOMP (Muthén, 1985), using the method of generalized least squares (GLS). Since LISCOMP is not yet available for a VAX computer the FA model for dichotomous items will not be treated throughout this paper.

Common to the models treated above is the usage of all available information from the data matrix. If we are willing to use only information of the one-way marginals (percents-correct) and the two-way marginals p_{ij}, it is possible to approximate the above models by more classical models, e.g. models in the realm of common FA.

If the latent continuous response variables $\underline{Y}$, underlying the manifest dichotomous response variables $\underline{X}$, are MVN distributed, then the ML estimator of the product-moment correlation between the (bivariate normal distributed) variables Y_i and Y_j is given by the tetrachoric correlation between X_i and X_j. Hence it seems reasonable to perform a common FA on the matrix of estimated tetrachoric correlations in order to obtain estimates for the FA parametrization of model (3). Estimates of the IRT parametrization of the M2PNO model can be obtained by the transformations (14). There are, however, some problems connected with this approach. As already noted, the matrix of sample tetrachoric correlations is not sufficient, and the estimates become unstable when the proportions of the 2 x 2 table are extreme, or the number of observations is low. Furthermore the matrix of tetrachoric correlations is not necessarily positive definite, and this makes the matrix inappropriate for the ML and GLS FA methods. Finally, the possible occurence of one or more unique variances approximately equal to zero, i.e. Heywood (1931) cases, may be encountered. See Mislevy (1986) for an excellent review of these problems.

Various FA programs are available. In SPSS[x] (1986), iterative principal FA (Harman & Jones, 1966), minimum residuals or unweighted least squares FA (Harman & Jones, 1966), generalized least squares FA (Jöreskog & Goldberger, 1972), maximum likelihood FA (Jöreskog, 1967), and alpha FA (Kaiser & Caffrey, 1965) are implemented. These methods will be denoted by IPFA, ULS, GLS, ML and ALPHA, respectively. In LISREL VI (Jöreskog & Sörbom, 1984) ULS, GLS and ML methods are available. Additionally, an adjusted minimum residuals (MINRES) FA method (Harman & Jones, 1966; Zegers & Ten Berge, 1983), in which arbitrarily lower bounds can be set on the unique variances (see also Knol, 1987a), has been used in the simulation study. An advantage of MINRES is the possibility to avoid Heywood cases.

For each method estimations of the parameters $\Lambda = (\lambda_1,\ldots,\lambda_n)'$, $\Psi^2 = (\psi_1^2,\ldots,\psi_n^2)'$, $\tau = (\tau_1,\ldots,\tau_n)'$, $A = (\alpha_1,\ldots,\alpha_n)'$ and $\beta = (\beta_1,\ldots,\beta_n)'$ can be obtained by either the transformations (13) for the FA models or (14) for the IRT models.

A Simulation Study Comparing Methods

To compare the various methods, data matrices were generated with known discrimination and difficulty parameters. The item parameters of four different sets of items are given in Table 1, where the groups of items which have the same discrimination parameters, have difficulties −2, −1, 0, 1 and 2, respectively.

Insert Table 1 about here

Multidimensional abilities $\underline{\theta}$ have been drawn from the MVN ($\underline{0}$,I) distribution using the procedure G05EZF of the NAG (1984) program library. The binary response of observation v (v=1,...,N) on item i was obtained by

$$(15) \qquad x_{vi} = \begin{cases} 1, & \text{if } p_i(\underline{\theta}_v) > u_{vi} , \\ 0, & \text{if } p_i(\underline{\theta}_v) \leq u_{vi} , \end{cases}$$

where u_{vi} is randomly drawn from the uniform [0,1] distribution and $p_i(\underline{\theta}_v)$ is given by (3). To verify the effects due to the sample size, the number of subjects was set equal to 250, 500 and 1000. For each set of itemparameters and each sample size, 10 replications were generated.

The sample tetrachoric correlations were computed by means of the procedure BECTR from the IMSL (1984) program library. In BECTR, a tetrachoric correlation is computed as the root of a sixth-degree equation. If there was no solution, a solution was obtained by adding an observation to each cell of the 2 x 2 frequency table (cf. Mislevy, 1986). In the case of multiple roots, the root with the smallest absolute value was used.

In the next section six criteria to assess differences among known and estimated item parameters will be given.

Six Criteria

All models will be compared with respect to estimates of both IRT and FA parameters. The criteria will be in terms of mean squared differences between the known and estimated item parameters.

In the case of orthogonal abilities $\underline{\theta}$, the n x m matrix Λ of factor loadings is determined up to an orthogonal transformation T. If Λ_o, Ψ_o^2, $\underline{\tau}_o$, $\underline{A}_o$ and $\underline{\beta}_o$ are the known item parameters, where the dimensionality m is known, then the first criterion is given by

(16a) $$g_1(\Lambda, T) = \{(nm)^{-1}\mathrm{tr}(\Lambda T - \Lambda_o)'(\Lambda T - \Lambda_o)\}^{½} ,$$

where (16a) is minimized as a function of T under the constraint that T is orthogonal. If $P\Delta Q' = \Lambda'\Lambda_o$ denotes the singular values decomposition of the matrix $\Lambda'\Lambda_o$, then the minimum of (16a) is attained for $T = PQ'$ (Green, 1952). The criteria for the unique variances $\underline{\psi}^2$ and the thresholds $\underline{\tau}$ are

(16b) $g_2(\underline{\psi}^2) = \{n^{-1}(\underline{\psi}^2 - \underline{\psi}_o^2)'(\underline{\psi}^2 - \underline{\psi}_o^2)\}^{½}$

and

(16c) $g_3(\underline{\tau}) = \{n^{-1}(\underline{\tau} - \underline{\tau}_o)'(\underline{\tau} - \underline{\tau}_o)\}^{½}$,

respectively. In the case of orthogonal $\underline{\theta}$ the n x m matrix A of discriminations is also determined up to an orthogonal transformation, and the first criterion for the IRT parametrization is

(16d) $g_4(A, T) = \{(nm)^{-1}tr(AT - A_o)'(AT - A_o)\}^{½}$,

where (16d) is also minimized as a function of T under the constraint that T is orthogonal. The criterion for the difficulties $\underline{\beta}$ is

(16e) $g_5(\underline{\beta}) = \{n^{-1}(\underline{\beta} - \underline{\beta}_o)'(\underline{\beta} - \underline{\beta}_o)\}^{½}$.

The last criterion is given by the mean squared difference between the values of the generated and estimated ICF's

(16f) $g_6(A, \underline{\beta}) = \{(Nn)^{-1} \sum_{v=1}^{N} \sum_{i=1}^{n} [p_i(\underline{\theta}_v) - p_{io}(\underline{\theta}_v)]^2\}^{1/2}$

where $p_{io}(\underline{\theta}_v) = F(\underline{\alpha}_{io}'\underline{\theta}_v - \beta_i)$ and $\underline{\theta}_v$ (v=1,...,N) are assumed to be known.

Results

In order to apply the GLS and ML common FA methods, the matrix of tetrachoric correlations has to be positive definite. If the matrix of tetrachoric correlations R is indefinite or ill-conditioned with respect to inversion, a smoothing procedure has to be used. Let R = KDK' be the eigendecomposition of R, where D is an n x n diagonal matrix containing the eigenvalues d_i of R in descending order and K is the matrix of corresponding normalized eigenvectors. Then a nonnegative definite correlation matrix $R^+(\delta)$ can be obtained by

(17) $R^+(\delta) = (\text{Diag } KD^+K')^{-1/2} KD^+K' (\text{Diag } KD^+K')^{-1/2}$,

where d_i^+ is diagonal element i of D^+ (i=1,...,n) with $d_i^+ = \max(d_i, \delta)$ and $\delta \geq 0$. Note that if $\delta = 0$, KD^+K' is the least-squares approximation to R of rank r, where r is the number of positive eigenvalues of R (cf. Rao, 1973). Note also that $R^+(0)$ coincides with Frane's (1978) smoothing procedure.

To investigate the effects of the smoothing procedure the common FA methods that do not require positive definiteness of R (i.e., IPFA, ULS, MINRES and ALPHA) were applied to 10 indefinite matrices obtained from dataset 2 (cf. Table 1) with N = 250 and various values of δ. The results are given in Table 2.

Insert Table 2 about here

Only for MINRES a slight increase is observed. From the results in Table 2 it can be inferred that the effect of smoothing is negligible. Additionally a small decrease of the total number of variables with estimated unique variances smaller than .2 is observed for increasing values of δ. Therefore it was decided to perform all common FA on the same smoothed tetrachoric correlation matrix $R^+(.005)$, ensuring that the matrix to be analyzed is sufficiently well conditioned.

In Table 3 the mean values of the six criteria over 10 replications for the different methods are given for generated data corresponding to the unidimensional data set 1.

Insert Table 3 about here

Because the ULS procedures of $SPSS^x$ and LISREL and IPFA of $SPSS^x$ gave the same results, only the IPFA procedure of $SPSS^x$ has been reported in the tables. Only the GLS procedure from LISREL is reported, because it gives consistently better results then the correspondent $SPSS^x$ procedure. The ML procedure of $SPSS^x$ procedure was chosen because the corresponding LISREL procedure often did not converge to a proper solution.

From the results in Table 3 it can be concluded that MAXLOG performs very badly. The GLS and ML procedures also perform badly. As expected TESTFACT is the best procedure and NOHARM also performs quite good. The procedures IPFA and MINRES give approximately the same results.

The results of the various methods obtained from the multidimensional datasets 2, 3 and 4 are given in the Tables 4, 5 and 6, respectively.

Insert Tables 4–6 about here

It has to be noted that the GLS procedure applied to data set 4 never converged to a proper solution; hence, the outcomes are not reported in Table 6.

Essentially the same conclusions can be drawn from the results for these multidimensional datasets. However, for the three-dimensional datasets 3 and 4 the performances of NOHARM and TESTFACT decrease, and in fact become even worse than the simple common FA procedures IPFA and MINRES.

The Dimensionality of Binary Scored Items

The dimensionality of binary items has been a source of debate in educational and psychological literature, and various aspects have been discussed by Goldstein (1980), McDonald (1981) and Hattie (1985), among others.

The increasing interest in IRT and the widespread use of the one, two- and three-parameter IRT models, which all assume a unidimensional ability space, has increased the need for a clear definition of dimensionality. Moreover, reliable indices to assess the dimensionality of a set of binary scored items are needed.

Both Goldstein (1980) and McDonald (1981) discuss the dimensionality of binary items in relation with the principle of local independence. The formal requirement of the independence of the item responses $(x_1,\ldots,x_n)$ is that the joint distribution of the responses given a vector of abilities $\underline{\theta}$ is equal to the product of the marginal distributions of the items given $\underline{\theta}$, i.e.:

$$(19) \quad P(\underline{X} = \underline{x}|\underline{\theta}) = \prod_{i=1}^{m} P(X_i = x_i|\underline{\theta})$$

Goldstein (1980) states that a distinction should be made between the conditional distribution of the item responses, given $\underline{\theta}$ and the conditional distribution of item responses given both $\underline{\theta}$ and the responses to the other items. A unidimensional model can be assumed with or without local independence. The assumption of local independence is really very strong, and Goldstein (1980) doubts whether this assumption is actually met in real life situations.

On the other hand, McDonald (1981) argues that the principle of local independence and the definition of dimensionality are related to each other. If a subject from a given population is completely characterized by one or more abilities $\underline{\theta} = (\theta_1, \ldots, \theta_m)$, then the scores of that subject with the abilities $\underline{\theta}$ on the n items are mutually statistically independent. This means that, if these abilities span the complete ability space in the population, all mutual statistical dependencies among the n items are explained by these abilities $(\theta_1, \ldots, \theta_m)$. If, however, a model is specified with a number of abilities, which do not span the complete ability space, then there will still remain mutual dependencies among the items. An adequate method to specify the dimensionality of a set of binary items is therefore needed. Unfortunately, no all-round index to identify the dimensionality of binary items is available.

Dimensionality Indices

The most frequently applied procedure to verify the dimensionality of binary items is to compute tetrachoric correlations and to inspect the eigenvalues of the corresponding matrix. Sometimes even phi-coefficients are used (Hambleton & Rovinelli, 1986), but it is well-known that these coefficients are affected by the difficulties of the items. As already mentioned above, the use of tetrachoric correlations has some disadvantages.

Dimensionality indices obtained from linear factor analysis will not be optimal, since IRT models for binary data are intrinsically nonlinear, i.e. nonlinear in the item parameters and in the ability parameters.

Up to date, there are no widely accepted tests of fit for models formulated on binary items which are comparable with the χ^2-test and the residual analysis in common FA and research is needed on the assessment of misfit of multidimensional IRT models. Perhaps, the use of a formal χ^2-test should be avoided, because of the distributional problems connected with the χ^2-test for small samples and because the use of a test statistic is never in itself a sufficient justification for the acceptance or rejection of a certain model.

Hambleton and Rovinelli (1986) compared some methods for the determination of the dimensionality c a set of items and concluded that linear factor analysis based on phi-coefficients tended to overestimate the number of underlying abilities and that inspection of the residuals obtained from a nonlinear factor analysis would be a promising approach to assess dimensionality.

Hattie (1985) reviewed the rationale of various methods and concluded that too many indices were developed on an ad hoc base. Hattie (1984) reported that indices based on residuals obtained from nonlinear factor analysis could very well distinguish a unidimensional set of items from a set with more than one dimension and recommended the use of the mean squared or mean absolute residuals as a suitable loss function.

Tucker, Humphreys, Lloyd, and Roznowski (1986) compared some indices based on the eigenvalues of the tetrachoric correlation matrix with some indices based on the local-independence principle. Their preliminary results seem to indicate that the indices based on the eigenvalues do not work very well.

A Simulation Study Assessing Dimensionality

Following the suggestion made by Hambleton and Rovinelli (1986) and Hattie (1984, 1985), the residuals were used as a measure for dimensionality.

If Λ_k is the estimated matrix of factor loadings from a solution with k estimated common factors, then the matrix of residuals $R^* = [r_{ij}^*]$ is

(19) $R^* = R - \Lambda_k \Lambda_k'$

and the mean squared and mean absolute residuals are

(20a) $e_1 = 2[n(n-1)]^{-1} \sum_{i<j}\sum (r_{ij}^*)^2$

and

(20b) $e_2 = 2[n(n-1)]^{-1} \sum_{i<j}\sum |r_{ij}^*|$.

respectively.

For each of the first three datasets given in Table 1, with one, two and three dimensions, respectively, 5 datamatrices were generated for sample sizes 250, 500 and 1000. In the Tables 7, 8 and 9 the mean squared residuals are given for the three datasets, after an analysis was performed with assumed dimensionality ranging from one to five.

Insert Tables 7–9 about here

Since the mean absolute residuals and the mean squared residuals lead to the same conclusions, only the mean squared residuals are given. To verify the sizes of the obtained residuals, the mean squared residuals for datasets with all correlations among the variables equal to zero, i.e. no common factors, are also given. In the Tables 7 through 9 only the results for MINRES, NOHARM and TESTFACT are given. Although the results of an ordinary principal components analysis (PCA) will generally not give a satisfactory fit when applied to binary data, the pattern of residuals might give an adequate indication of dimensionality. Therefore the residuals obtained after a PCA are also given in the Tables.

From the Tables 7 through 9 relatively high values of the mean squared residuals e_1 can be observed when datasets have been analyzed with a smaller dimensionality than the generated dimensionality. This applies to all methods. Also, a large drop of e_1 can be observed between the analyses with m-1 and m dimensions (where m is the generated dimensionality of the dataset). No such drops of e_1 are observed for analyses with higher assumed dimensionality. Hence, it seems that the dimensionality of a dataset can be assessed by inspecting the mean squared residuals obtained from different assumed dimensionalities of the methods.

Discussion

The most striking result of the simulation study in which various IRT and FA programs were compared, is that the common FA methods outperformed the more complex programs TESTFACT, MAXLOG and NOHARM, despite their theoretical advantages. Also, quite remarkable is the failure of the ML and GLS FA procedures compared with IPFA, ULS or MINRES. Of course this may be due to the implementation of the specific methods. Nevertheless, since no other programs are available, it is advised to use IPFA, ULS or MINRES. An additional advantage is that these programs can handle relatively large numbers of variables and factors. Because IPFA has some algorithmic drawbacks (cf. Gorsuch, 1974, p. 98) and MINRES (or ULS) perform equally well as IPFA, it is advised to avoid the usage of IPFA. An advantage of MINRES compared to ULS is that MINRES avoids Heywood cases. Therefore, in large-scale applications it is advised to use MINRES on the (possibly smoothed) matrix of tetrachoric correlations.

A possible drawback of MINRES could be that no statistical goodness of fit measure is available, hence the estimation of the dimensionality of the model can be problematic. Therefore, a non-statistical procedure for assessing the dimensionality of the model is proposed, leading to essentially the same results as TESTFACT and NOHARM.

References

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 40, 443–459.

Bock, R.D., Gibbons, R., & Muraki, E. (1985). Full-Information Item Factor Analysis (MRC Report 85-1 [Revised]). Chicago, IL: University of Chicago, National Opinion Research Center.

Bock, R.D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. Psychometrika, 35, 283–319.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. Psychometrika, 40, 5–32.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39, 1–38.

Frane, J.W. (1978). Missing data and BMDP: Some pragmatic approaches. Proceedings of the Statistical Computing Section (pp. 27–33). Washington, DC: American Statistical Association.

Fraser, C. (1988). NOHARM II. A Fortran program for fitting unidimensional and multidimensional normal ogive models of latent trait theory. Armidale, Australia: The University of New England, Centre for Behavioural Studies.

Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. British Journal of Mathematical and Statistical Psychology, 33, 234–246.

Gorsuch, R.L. (1974). Factor Analysis. Philadelphia, PA: Saunders.

Green, B.F. (1952). The orthogonal approximation of an oblique structure in factor analysis. Psychometrika, 17, 429–440.

Haley, D.C. (1952). Estimation of the dosage mortality relationship when the dose is subject to error (Technical Report 15). Stanford, CA: Stanford University, Applied Mathematics and Statistics Laboratory.

Hambleton, R.K., & Rovinelli, R.J. (1986). Assessing the dimensionality of a set of test items. Applied Psychological Measurement, 10, 287–302.

Harman, H.H., & Jones, W.H. (1966). Factor analysis by minimizing residuals (MINRES). Psychometrika, 31, 351–368.

Hattie, J.A. (1984). An empirical study of various indices for determining unidimensionality. Multivariate Behavioral Research, 19, 49–78.

Hattie, J.A. (1985). Methodology review: Assessing unidimensionality of tests and items. Applied Psychological Measurement, 9, 139–164.

Heywood, H.B. (1931). On finite sequences of real numbers. Proceedings of the Royal Society, Series A, 134, 486–501.

IMSL (1984). Library (9th ed.). Houston, TX: International Mathematical and Statistical Libraries.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika, 32, 443-482.

Jöreskog, K.G., & Goldberger, A.S. (1972). Factor Analysis by generalized least squares. Psychometrika, 37, 243-260.

Jöreskog, K.G., & Sörbom, D. (1984). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood. Mooresville, IN: Scientific Software.

Kaiser, H.F., & Caffrey, J. (1965). Alpha factor analysis. Psychometrika, 30, 1-14.

Knol, D.L. (1987a). Het verband tussen itemresponstheorie en factoranalyse voor dichotome items (Rapport 87-2) [The relationship between item response theory and factor analysis for dichotomous items]. Enschede, The Netherlands: University of Twente, Department of Education.

Knol, D.L. (1987b). Stapsgewijze itemselectieprocedures in het Raschmodel (Rapport 87-6) [Stepwise item selection procedures in the Rasch model]. Enschede, The Netherlands: University of Twente, Department of Education.

McDonald, R.P. (1981). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 100-117.

McDonald, R.P. (1985). Unidimensional and multidimensional models for item response theory. In D.J. Weiss (Ed.), Proceedings of the 1982 Computerized Adaptive Testing Conference (pp. 127-148). Minneapolis: University of Minnesota, Department of Psychology, Psychometrics Methods Program.

McKinley, R.L., & Reckase, M.D. (1982). The use of the general Rasch model with multidimensional item response data (Research Report ONR 82-1). Iowa City, IA: The American College Testing Program.

McKinley, R.L., & Reckase, M.D. (1983). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. Behavior Research Methods & Instrumentation, 15, 389-390.

Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Mood, A.M., Graybill, F.A., & Boes, D.C. (1974). Introduction to the Theory of Statistics (3rd ed.). New-York: McGraw-Hill.

Mulaik, S.A. (1972). A mathematical investigation of some multidimensional Rasch models for psychological tests. Paper presented at the annual meeting of the Psychometric Society, Princeton, NJ.

Muthén, B. (1978). Contributions to factor analysis of dichotomized variables. Psychometrika ,43, 551-560.

Muthén, B. (1985). LISCOMP. Mooresville, IN: Scientific Software.

NAG (1984). Library (Mark 11). Oxford, UK: Numerical Algorithms Group.

Rao, C.R. (1973). Linear statistical inference and its applications (2th ed.). New-York: Wiley.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: The Danish Institute for Educational Research.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), Proceedings of the Fourth Berkeley Symposium On Mathematical Statistics and Probability (Vol. 4, pp. 321–333). Berkeley: University of California Press.

Reckase, M.D. (1973). Development and application of a multivariate logistic latent trait model. Dissertation Abstracts International, 33.

SPSSx (1986). User's Guide (2nd ed.). New-York: McGraw-Hill.

Sympson, J.B. (1978). A model for testing with multidimensional items. In D.J. Weiss (Ed.), Proceedings of the 1977 Computerized Adaptive Testing Conference (pp. 82–98). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analyses of discretized variables. Psychometrika, 52, 393–408.

Tucker, L.R., Humphreys, L.S., Lloyd, G., & Roznowski, M.A. (1986). Comparative accuracy of five indices of dimensionality. Urbana: University of Illinois, Department of Psychology.

Verhelst, N. (1983). Automatische itemselectieprocedure in het Rasch-model [Automatic item selection procedures in the Rasch model]. Utrecht, The Netherlands, University of Utrecht, Department of Psychology.

Whiteley, S.E. (1980). Multicomponent latent trait models for ability tests. Psychometrika, 45, 497–494.

Wilson, D.T., Wood, R., & Gibbons, R.T. (1984). TESTFACT. Test scoring, item statistics, and item factor analysis. Mooresville, IN: Scientific Software.

Zegers, F.E., & Ten Berge, J.M.F. (1983). A fast and simple computational method of minimum residual factor analysis. Multivariate Behavioral Research, 18, 331–340.

Table 1
Discrimination parameters of the four different data sets (each group of five items with the same discrimination parameters has difficulty parameters -2, -1, 0, 1, and 2)

Set	n	m		A_O		
1	15	1	(5x)	1.00		
			(5x)	1.25		
			(5x)	1.50		
2	15	2	(5x)	1	1	
			(5x)	1	0	
			(5x)	0	1	
3	15	3	(5x)	1	1	0
			(5x)	1	0	1
			(5x)	0	1	1
4	30	3	(5x)	1	1	0
			(5x)	1	0	1
			(5x)	0	1	1
			(5x)	1	0	0
			(5x)	0	1	0
			(5x)	0	0	1

Table 2

The total number of estimated unique variances <.2 and the mean values of two FA criteria for four CFA methods for data set 2 and various $R^{+}(\delta)$ with N = 250

		δ					
Crit	Method	d_n	0	.001	.005	.01	.05
$\#\psi_i^2<.2$	IPFA	4	3	3	3	3	3
	ULS	4	3	3	3	3	3
	MINRES	3	1	0	1	2	2
	ALPHA	9	7	7	6	6	4
Λ	IPFA	.096	.096	.096	.096	.096	.096
	ULS	.096	.096	.096	.096	.096	.096
	MINRES	.095	.096	.096	.096	.096	.096
	ALPHA	.104	.103	.103	.103	.103	.103
ψ^2	IPFA	.125	.125	.125	.125	.125	.125
	ULS	.125	.125	.125	.125	.125	.125
	MINRES	.121	.124	.124	.124	.124	.124
	ALPHA	.144	.141	.141	.141	.141	.140

Table 3

Mean values of the six criteria for different N and various methods for data set 1

N	Method	Criterion					
		Λ	Ψ^2	I	A	β	p
250	IPFA	.091	.127	.097	.296	.231	.047
	ALPHA	.096	.133	.097	.311	.239	.048
	ML	.092	.128	.097	.294	.226	.047
	GLS	.122	.166	.097	.370	.284	.053
	MINRES	.091	.127	.097	.296	.231	.047
	NOHARM	.067	.098	.097	.264	.245	.043
	TESTFACT	.064	.093	.105	.244	.226	.045
	MAXLOG	.081	.124	.128	.454	.440	.056
500	IPFA	.051	.075	.058	.188	.152	.030
	ALPHA	.053	.078	.058	.196	.157	.030
	ML	.050	.073	.058	.183	.151	.029
	GLS	.071	.102	.058	.244	.195	.034
	MINRES	.051	.075	.058	.188	.152	.030
	NOHARM	.047	.071	.058	.198	.188	.029
	TESTFACT	.044	.065	.058	.169	.166	.028
	MAXLOG	.066	.105	.101	.410	.414	.044
1000	IPFA	.031	.047	.044	.122	.108	.021
	ALPHA	.032	.048	.044	.126	.109	.021
	ML	.031	.046	.044	.121	.108	.021
	GLS	.036	.054	.044	.140	.127	.022
	MINRES	.031	.047	.044	.122	.108	.021
	NOHARM	.029	.045	.044	.130	.120	.021
	TESTFACT	.028	.042	.044	.110	.104	.020
	MAXLOG	.048	.079	.079	.340	.352	.034

Table 4

Mean values of the six criteria for different N and various methods for data set 2

		Criterion					
N	Method	Λ	Ψ^2	τ	A	β	p
250	IPFA	.103	.130	.093	.239	.237	.057
	ALPHA	.103	.138	.093	.254	.238	.059
	ML	.119	.149	.093	.280	.288	.063
	GLS	.119	.166	.093	.283	.271	.063
	MINRES	.103	.130	.093	.239	.237	.057
	NOHARM	.099	.136	.093	.284	.344	.059
	TESTFACT	.093	.122	.101	.258	.296	.059
	MAXLOG	.119	.227	.149	.724	.571	.102
500	IPFA	.072	.093	.061	.181	.173	.041
	ALPHA	.073	.094	.061	.186	.175	.042
	ML	.078	.101	.061	.193	.186	.043
	GLS	.080	.109	.061	.192	.189	.042
	MINRES	.072	.091	.061	.179	.173	.041
	NOHARM	.070	.092	.061	.192	.210	.042
	TESTFACT	.070	.089	.065	.167	.181	.041
	MAXLOG	.104	.210	.106	.655	.496	.088
1000	IPFA	.048	.057	.050	.113	.102	.030
	ALPHA	.048	.057	.050	.114	.102	.030
	ML	.059	.073	.050	.139	.129	.035
	GLS	.052	.063	.050	.115	.115	.030
	MINRES	.048	.057	.050	.113	.102	.030
	NOHARM	.047	.058	.050	.120	.121	.030
	TESTFACT	.048	.061	.054	.116	.112	.031
	MAXLOG	.099	.205	.087	.604	.488	.082

Table 5

Mean values of the six criteria for different N and various methods for data set 3

N	Method	Criterion					
		Λ	Ψ^2	I	A	B	p
250	IPFA	.095	.116	.086	.237	.218	.067
	ALPHA	.097	.121	.086	.245	.225	.069
	ML	.120	.148	.086	.289	.258	.081
	GLS	.135	.204	.086	.335	.309	.089
	MINRES	.095	.112	.086	.235	.218	.067
	NOHARM	.087	.108	.086	.243	.283	.066
	TESTFACT	.094	.119	.118	.324	.318	.078
	MAXLOG	.288	.415	.497	.739	.739	.222
500	IPFA	.063	.083	.069	.183	.189	.051
	ALPHA	.063	.083	.069	.184	.191	.052
	ML	.085	.110	.069	.228	.213	.064
	GLS	.089	.127	.069	.243	.243	.065
	MINRES	.062	.081	.069	.182	.189	.051
	NOHARM	.060	.079	.069	.182	.220	.051
	TESTFACT	.066	.098	.080	.207	.225	.055
	MAXLOG	.260	.397	.498	.694	.871	.218
1000	IPFA	.045	.057	.050	.126	.135	.037
	ALPHA	.045	.058	.050	.127	.134	.037
	ML	.045	.059	.050	.128	.140	.037
	GLS	.053	.076	.050	.148	.164	.041
	MINRES	.045	.058	.050	.126	.135	.037
	NOHARM	.043	.055	.050	.122	.136	.036
	TESTFACT	.049	.082	.065	.147	.154	.043
	MAXLOG	.287	.444	.467	.698	.682	.213

Table 6

Mean values of the six criteria for different N and various methods for data set 4

N	Method	Criterion					
		Λ	Ψ²	I	A	B	P
250	IPFA	.087	.113	.090	.183	.218	.057
	ALPHA	.089	.118	.090	.192	.223	.059
	ML	.087	.113	.090	.185	.224	.057
	GLS						
	MINRES	.087	.112	.090	.183	.218	.057
	NOHARM	.083	.101	.090	.202	.263	.058
	TESTFACT	.092	.126	.121	.209	.285	.067
	MAXLOG	.159	.211	.187	.747	.454	.140
500	IPFA	.063	.075	.066	.138	.154	.043
	ALPHA	.063	.076	.066	.140	.156	.044
	ML	.063	.077	.066	.140	.157	.044
	GLS						
	MINRES	.063	.075	.066	.137	.154	.043
	NOHARM	.063	.079	.066	.155	.193	.045
	TESTFACT	.071	.103	.089	.161	.172	.053
	MAXLOG	.154	.194	.159	.670	.393	.130
1000	IPFA	.043	.055	.045	.104	.118	.031
	ALPHA	.044	.056	.045	.108	.120	.031
	ML	.043	.055	.045	.104	.120	.031
	GLS						
	MINRES	.043	.055	.045	.104	.118	.031
	NOHARM	.043	.054	.045	.107	.131	.031
	TESTFACT	.056	.095	.075	.137	.126	.044
	MAXLOG	.103	.165	.107	.552	.376	.094

Table 7

Mean squared residuals for different generated and assumed dimensionality of the data and various methods with N = 250

Method	Generated Dimensionality	Assumed Dimensionality 1	2	3	4	5
PCA	0	.0184	.0163	.0147	.0125	.0101
	1	.0111	.0081	.0061	.0048	.0039
	2	.0354	.0102	.0072	.0056	.0041
	3	.0347	.0165	.0060	.0045	.0034
MINRES	0	.0154	.0111	.0081	.0056	.0040
	1	.0098	.0064	.0043	.0031	.0022
	2	.0338	.0083	.0053	.0037	.0027
	3	.0331	.0148	.0046	.0032	.0024
NOHARM	0	.0072	.0042	.0029	.0022	.0015
	1	.0038	.0024	.0013	.0007	.0004
	2	.0169	.0023	.0015	.0010	.0006
	3	.0264	.0112	.0027	.0017	.0011
TESTFACT	0	.0150	.0107	.0077	.0054	.0039
	1	.0072	.0050	.0036	.0031	.0022
	2	.0288	.0079	.0045	.0029	.0020
	3	.0256	.0131	.0057	.0037	.0039

Table 8

Mean squared residuals for different generated and assumed dimensionality of the data and various methods with N = 500

Method	Generated Dimensionality	Assumed Dimensionality 1	2	3	4	5
PCA	0	.0129	.0129	.0121	.0118	.0112
	1	.0073	.0054	.0045	.0037	.0033
	2	.0350	.0064	.0048	.0038	.0031
	3	.0275	.0135	.0038	.0029	.0023
MINRES	0	.0096	.0073	.0050	.0036	.0024
	1	.0061	.0038	.0027	.0019	.0013
	2	.0334	.0046	.0029	.0020	.0014
	3	.0260	.0119	.0024	.0016	.0012
NOHARM	0	.0038	.0026	.0016	.0011	.0006
	1	.0018	.0011	.0007	.0005	.0003
	2	.0183	.0013	.0008	.0005	.0003
	3	.0218	.0094	.0011	.0007	.0004
TESTFACT	0	.0095	.0069	.0050	.0036	.0026
	1	.0061	.0039	.0025	.0015	.0010
	2	.0303	.0057	.0034	.0021	.0015
	3	.0225	.0104	.0031	.0021	.0012

Table 9

Mean squared residuals for different generated and assumed dimensionality of the data and various methods with N = 1000

Method	Generated Dimension-ality	Assumed Dimensionality				
		1	2	3	4	5
PCA	0	.0080	.0090	.0100	.0106	.0113
	1	.0042	.0040	.0039	.0036	.0033
	2	.0307	.0044	.0037	.0031	.0029
	3	.0298	.0153	.0027	.0021	.0018
MINRES	0	.0049	.0036	.0026	.0018	.0012
	1	.0030	.0021	.0018	.0011	.0008
	2	.0290	.0026	.0017	.0011	.0008
	3	.0283	.0136	.0014	.0009	.0006
NOHARM	0	.0017	.0012	.0008	.0006	.0004
	1	.0009	.0005	.0003	.0002	.0001
	2	.0159	.0008	.0005	.0003	.0002
	3	.0225	.0105	.0008	.0004	.0003
TESTFACT	0	.0050	.0036	.0026	.0018	.0013
	1	.0026	.0019	.0015	.0011	.0008
	2	.0283	.0027	.0017	.0011	.0008
	3	.0245	.0124	.0024	.0014	.0010

Titles of recent Research Reports from the Division of Educational Measurement and Data Analysis, University of Twente, Enschede, The Netherlands.

RR-87-1 R. Engelen, Semiparametric estimation in the Rasch model

RR-87-2 W.J. van der Linden (Ed.), IRT-based test construction

RR-87-3 R. Engelen, P. Thommassen, & W. Vervaat, Ignatov's theorem: A new and short proof

RR-87-4 E. van der Burg, & J. de Leeuw, Use of the multinomial jackknife and bootstrap in generalized non-linear canonical correlation analysis

RR-87-5 H. Kelderman, Estimating a quasi-loglinear models for the Rasch table if the number of items is large

RR-87-6 R. Engelen, A review of different estimation procedures in the Rasch model

RR-87-7 D.L. Knol & J.M.F. ten Berge, Least-squares approximation of an improper by a proper correlation matrix using a semi-infinite convex program

RR-87-8 E. van der Burg & J. de Leeuw, Nonlinear canonical correlation analysis with k sets of variables

RR-87-9 W.J. van der Linden, Applications of decision theory to test-based decision making

RR-87-10 W.J. van der Linden & E. Boekkooi-Timminga, A maximin model for test design with practical constraints

RR-88-1 E. van der Burg & J. de Leeuw, Nonlinear redundancy analysis

RR-88-2 W.J. van der Linden & J.J. Adema, Algorithmic test design using classical item parameters

RR-88-3 E. Boekkooi-Timminga, A cluster-based method for test construction

RR-88-4 J.J. Adema, A note on solving large-scale zero-one programming problems

RR-88-5 W.J. van der Linden, Optimizing incomplete sample designs for item response model parameters

RR-88-6 H.J. Vos, The use of decision theory in the Minnesota Adaptive Instructional System

RR-88-7 J.H.A.N. Rikers, Towards an authoring system for item construction

RR-88-8 R.J.H. Engelen & W.J. van der Linden, Item information in the Rasch model

RR-88-9 W.J. van der Linden & T.J.H.M. Eggen, The Rasch model as a model for paired comparisons with an individual tie parameter

RR-88-10 H. Kelderman & G. Macready, Loglinear-latent-class models for detecting item bias

RR-88-11 D.L. Knol & M.P.F. Berger, Empirical Comparison between Factor Analysis and Item Response Models

Research Reports can be obtained at costs from Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

partment of
EDUCATION

A publication by
the Department of Education
of the University of Twente
P.O. Box 217
7500 AE Enschede