DOCUMENT RESUME

ED 382 642 TM 023 072

AUTHOR	De Champlain, Andre
TITLE	An Overview of Nonlinear Factor Analysis and Its Relationship to Item Response Theory.
PUB DATE	Apr 95
NOTE	33p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Educational Research; *Factor Analysis; *Item Response Theory; *Measurement Techniques; Simulation
IDENTIFIERS	Item Parameters; *Nonlinear Models

ABSTRACT

An approach that is currently gaining popularity in educational measurement is one that treats item response theory (IRT) as a special case of nonlinear factor analysis (NLFA). A brief overview is provided of some of the research that has examined the relationship between IRT and NLFA. Three NLFA models are outlined, emphasizing their major strengths and weaknesses. These are: (1) McDonald's polynomial approximation to a normal ogive model; (2) the factor analytic model for dichotomous variables of Christoffersson and Muthen (1975, 1984); and the full-information factor analytic model of Bock and Aitkin (1981) and Bock, Gibbons, and Muraki (1988). Although the full-information factor analytic model appears to be the strongest of the approaches described, it appears that a greater number of empirical studies should be undertaken to compare the various NLFA models with respect to how accurately they can recover simulated parameter values. An appendix presents an IRT-NLFA proof. (Contains 61 references.) (SLD)

An overview of nonlinear factor analysis and its relationship to item response theory

Andre De Champlain
Law School Admission Council

Paper presented at the meeting of the
American Educational Research Association

San Francisco, CA
April, 1995

Abstract

Item response theory (IRT) models have been used extensively to address educational measurement and psychometric concerns pertaining to a host of areas such as differential item functioning, equating and computer-adaptive testing. The many advantages of IRT models (e.g., item and ability parameter invariance), have contributed to their use in a wide number of areas by practitioners and researchers alike.

Another approach which is currently gaining popularity in educational measurement is the one that treats item response theory as a special case of nonlinear factor analysis (NLFA). Several authors have shown that these models are mathematically equivalent (Balassiano & Ackerman, 1995a; 1995b; Goldstein & Wood, 1989; Knol & Berger, 1991; McDonald, 1967; 1985; 1989; in press). It would therefore appear reasonable to make use of NLFA models to examine a multitude of educational measurement problems which had been, until quite recently, looked at solely from an IRT perspective.

The purpose of this paper is twofold:

- First, to provide a brief overview of some of the research that has examined the relationship between IRT and NLFA;
- Second, to outline three NLFA models, emphasizing their major strengths and weaknesses. More precisely, McDonald's (1967; 1982b) polynomial approximation to a normal ogive model, Christoffersson's (1975)/ Muthén's (1984) factor analytic model for dichotomous variables as well as Bock and Aitkin's (1981)/Bock, Gibbons and Muraki's (1988) full-information factor analytic model, will be summarized.

Introduction

Over the past three decades, the educational measurement and psychometric literatures have been replete with studies focusing on item response theory (IRT) models. The numerous textbooks that have been written centering primarily and, in some instances, exclusively on IRT attest to the importance of these models in the development and analysis of tests and items (Baker, 1992; Hambleton, 1983; 1989; Hambleton & Swaminathan; 1985; Hulin, Drasgow, & Parsons, 1983, Warm, 1978). The use of IRT models has been widespread in both testing organizations and departments of education for a variety of purposes such as item analysis (Baker, 1985; Mislevy & Bock, 1990; Wingersky, Patrick, & Lord, 1991; Thissen, 1993), score equating (Cook & Eignor, 1983; Lord, 1977; 1980; 1982; Petersen, Kolen, & Hoover, 1989; Skaggs & Lissitz, 1986), differential item functioning (Thissen, Steinberg, & Wainer, 1993) and computer adaptive testing (Hambleton, Zaal, & Pieters, 1993; Kingsbury & Zara, 1991; Wainer et al., 1990), to name a few. The many properties of IRT models, among them, that "sample-free" item parameter estimates and "test-free" ability estimates can be obtained, have generated considerable interest in their use to solve a host of measurement-related problems.

Another approach which is currently gaining popularity in educational measurement is the one that treats item response theory as a special case of nonlinear factor analysis (NLFA). Several authors have shown that these models are mathematically equivalent (Balassiano & Ackerman, 1995a; 1995b; Goldstein & Wood, 1989; Knol & Berger, 1991; McDonald, 1967; 1985; 1989; in press). Muthén (1978, 1983, 1984) has also demonstrated that commonly used models in IRT (e.g. the two-parameter normal ogive model) are really specific cases of a more general factor analytic model for categorical variables with multiple indicators (i.e. response categories). McDonald (1982b), starting from Spearman's common factor model, also shows that IRT models are a special case of NLFA and provides a general framework which includes unidimensional/

multidimensional, linear/nonlinear models as well as dichotomous and polychotomous models.

Takane and De Leeuw (1987) have also established that IRT models are mathematically equivalent to NLFA. These authors have provided a systematic series of proofs that show the equivalence of these models with dichotomous as well as polychotomous item responses.

Thus, it appears as though IRT and NLFA models represent two equivalent formulations of a more general latent trait model. Indeed, the two terms are often used interchangeably. For example, the model proposed by Bock and Aitkin (1981) has been synonymously referred to as full-information factor analysis (Bock, Gibbons, & Muraki, 1988) and multidimensional IRT (McKinley, 1988). Given the equivalence of IRT and NLFA, it would appear reasonable to make use of the latter models to examine a multitude of educational measurement problems which had been, until quite recently, looked at solely from an IRT perspective. Several nonlinear factor analytic models, with potential applications to measurement and psychometric issues, have been proposed in the literature (Bock & Aitkin, 1981; Bock, Gibbons, & Muraki, 1988; Bock & Lieberman, 1970; Christoffersson, 1975; McDonald, 1967; 1982b; Muthen, 1978; 1984).

The first part of this paper will consist in providing a brief overview of some of the research that has examined the relationship between common IRT models and NLFA.

Three NLFA models that have been used to address measurement related issues will be presented in the second part of this paper. Specifically, McDonald's (1967; 1982b) polynomial approximation to a normal ogive model, Christoffersson's (1975)/ Muthén's (1978) factor analytic model for dichotomous variables as well as Bock and Aitkin's (1981)/Bock, Gibbons and Muraki's (1988) full-information factor analytic model, will be summarized. In addition, some of the strengths and weaknesses of the models will be highlighted.

The relationship between common IRT models and nonlinear factor analysis

A considerable body of research has been dedicated to examining the relationship between common IRT models, e.g., logistic and normal ogive functions, and NLFA (Bartholomew, 1983; Goldstein & Wood, 1989; Knol & Berger, 1991; McDonald, 1967; 1989; Takane & De Leeuw, 1987).

Bartholomew (1983) has provided a general latent trait model on which several IRT as well as factor analytic functions for dichotomous variables are founded. The author states that common factor analytic models, such as those proposed by Bock and Aitkin (1981), Christoffersson (1975) and Muthen (1978) are special cases of this general latent trait model. The model is of the form,

$$G(\pi_i(\mathbf{y}))=\alpha_{i0}+\sum_{j=1}^{q}\alpha_{ij}H(y_j), \quad i=1,2\ldots,p. \tag{1}$$

Bartholomew states that the models outlined by Bock and Aitkin (1981), Christoffersson (1975) and Muthen (1978) use the probit function, $\{G(u)=\Phi^{-1}(u)\}$ for both G and H. Lord and Novick (1968), whose discussion on IRT models is restricted to the q=1 (i.e., unidimensional) case, treat y_{js} as parameters and use the logit for G and the probit for H. "Translated" in the unidimensional IRT vernacular, the terms in equation 1 would correspond to the following:

$G(\pi_i)$ = the response function outlining the probability of obtaining a correct response to item i;

$(\mathbf{y})$ = a vector of ability (in this case, a scalar, given that q=1);

α_{i0}= a parameter related to the *difficulty* of item i;

α_{ij} = a parameter related to the *discrimination* of item i on latent trait j;

$H(y_j)$= The density function for a given latent trait j.

Fraser and McDonald (1988) and McDonald (1981; in press) also examined the relationship between common item response functions (IRF) and NLFA. McDonald (1994) states that the unidimensional normal ogive model given by,

$$P(Y_i=1|\theta_j)=c_i+(1-c_i)N[a_i(\theta_j-b_i)]\,, \tag{2}$$

where,

$\theta_j=$ the latent variable;

$b_i =$ the θ value at the point of inflexion of the item response function;

$a_i =$ the slope of the IRF at its point of inflexion;

$c_i =$ the lower asymptote value of the IRF;

$N\{.\} =$ the normal distribution function;

can be re-expressed using the latent trait parameterization as,

$$P(y_i=1|\theta_j)=c_i+(1-c_i)N[f_{i0}+f_{i1}\theta_j]\,, \tag{3}$$

with $f_{i0} = -a_ib_i$ and $f_{i1} = a_i$; f_{i1} corresponding to the factor loading of factor 1 on item i. Function (3) can be generalized to the multidimensional case,

$$P(y_i=1|\underline{\theta})=c_i+(1-c_i)N[f_{i0}+\underline{f}'_i\underline{\theta}]\,. \tag{4}$$

Fraser and McDonald (1988) and McDonald (1981; in press) also demonstrated that the latent trait model shown in (4) could be derived (c.f. Christoffersson, 1975) in the form,

$$P(i=1|\underline{\theta})=c_i+(1-c_i)N[t_{i0}+\underline{h}'_i\underline{\theta}/m_i]\,. \tag{5}$$

The parameters in models (4) and (5) are related by,

$$t_{i0}=f_{i0}/(1+\underline{f}'_i P \underline{f}_i)^{1/2},$$

$$h_i=f_i/(1+\underline{f}'_i P \underline{f}_i)^{1/2},$$

$$m_i^2 = 1 / s_i^2,$$

$$s_i = (1+\underline{f}'_i P \underline{f}_i)^{1/2},$$

where (i=1, ... , m) and P is the m x m matrix containing the correlations among the dimensions (assuming the latent traits have been standardized). A detailed discussion of this relationship is found in McDonald (1985).

Knol and Berger (1991) examined the relationship between several NLFA models and logistic IRT functions. More precisely, the authors focused their attention on comparing Bock and Aitkin's (1981) full-information factor analytic model and McDonald's (1967) polynomial approximation to a normal ogive model, to the two-parameter logistic IRT function.

Bock and Aitkin's model (1981) uses a marginal maximum likelihood procedure in the estimation of item parameters. In the model, $\mathbf{X} = (X_1, \ldots, X_n)'$ corresponds to a random response pattern vector to n binary variables, where each X_i(i=1,...,n) value is defined as 1, if the item is correctly answered and 0, if incorrectly answered. Under the assumption of local independence, the marginal probability of the response vector $\mathbf{X} = \mathbf{x}$ is given by,

$$P(\boldsymbol{X}=\boldsymbol{x}) = \int \prod_{i=1}^{n} [p_i(\theta)]^{x_i} [1-p_i(\theta)]^{1-x_i} g(\theta)\, d\theta, \qquad \textbf{(6)}$$

where $p_i(\theta)$ corresponds to the item characteristic function of item i, $g(\theta)$ is the density function of the latent m-component random vector of abilities θ, and the integration is taken over the entire multidimensional ability space.

Knol and Berger (1991) state that if θ is assumed to be multivariate normal distributed with a mean equal to 0 and a covariance matrix $\mathbf{I}$, the multidimensional two-parameter normal ogive model ICF for item i (i=1,...,n) is given by,

$$p_i(\theta) = F(\alpha_i'\theta - \beta_i), \tag{7}$$

where,

α_i = the m x 1 vector of item discrimination parameters;

β_i = the item difficulty parameter;

$F(.)$ = the cumulative standard normal distribution.

Knol and Berger (1991) also examined the relationship between McDonald's polynomial approximation to a normal ogive model (McDonald, 1967; 1982b) and the two-parameter logistic IRT function. McDonald, using harmonic analysis, proposed a NLFA model that is based on the pairwise joint-proportion of the item responses. The ICFs for this model are approximated by a third degree Hermite-Tchebycheff polynomial. The pairwise probabilities $\pi_{ij} = P(X_i=1, X_j=1)$ are estimated by minimizing the unweighted least-squares function,

$$f(\mathbf{A}, \beta) = \sum_{i<j}\sum [p_{ij} - \hat{\pi}_{ij}(\alpha_i, \beta_i, \alpha_j, \beta_j)]^2, \tag{8}$$

where,

$\mathbf{A} = (\alpha_1, \ldots, \alpha_n)'$, described in (7);

$\beta = (\beta_1, \ldots, \beta_n)'$, also defined in (7)

p_{ij} = the observed joint-proportions.

As Knol and Berger (1991) state, the relationship between the logistic distribution function $L(.)$ and the cumulative standard normal distribution function $F(.)$ given by (Mood, Graybill, & Boes, 1974),

$$|F(z) - L(1.7z)| < .01 \tag{9}$$

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

$$p_i(\theta_j) = \frac{e^{[1.7(\alpha_i'\theta_j-\beta_i)]}}{1+e^{[1.7(\alpha_i'\theta_j-\beta_i)]}} = L[1.7(\alpha_i'\theta_j-\beta_i)]\,. \tag{10}$$

Takane and De Leeuw (1987) also showed that common IRT models and NLFA models are formally equivalent. These authors have provided a proof that demonstrates the equivalence of Bock and Aitken's (1981) full-information factor analytic model to Christoffersson's(1975)/Muthén's(1984) generalized least-squares factor analytic model for dichotomous variables. This proof is presented in Appendix A for the reader's benefit.

Summary

The purpose of the first part of the paper was to briefly outline past research that has investigated the relationship between common IRT models and NLFA. These studies have shown that logistic and normal ogive functions are formally equivalent to McDonald's (1967; 1982b) polynomial approximation to a normal ogive model, Christoffersson's (1975)/Muthén's (1984) factor analytic model for dichotomous variables, and the full-information factor analytic approach advocated by Bock and Aitkin (1981) as well as Bock, Gibbons and Muraki (1988). Hence, based on the IRT-NLFA relationship, it would appear that these latter models might provide a useful framework with which common measurement and psychometric problems can be addressed. A summary of these three NLFA models is provided in the next section of the paper, emphasizing some of the advantages and limitations of each approach for the practitioner.

A polynomial approximation to a normal ogive model

McDonald (1967; 1982a; 1982b, 1989; in press) and McDonald and Ahlawat (1974) have provided a general framework that enables the organization of existing unidimensional as well as multidimensional IRT models based on a more general NLFA approach. Specifically, generalizing from Spearman's common factor model, McDonald (1982b) has presented three classes of models which can be used in educational measurement, that is,

i. models that are strictly linear in both their coefficients and latent traits;

ii. models that are linear in their coefficients but not in their latent traits;

iii. models that are strictly nonlinear.

McDonald's distinctive contribution to the area, however, lies with the second class of models presented. McDonald and Ahlawat (1974) have proposed a group of regression functions that are linear in their coefficients (i.e. their item parameters) but nonlinear in their latent traits, of the general form,

$$f_i(x_1, \ldots, x_t) = a_{i0} + \sum_{l=1}^{t} \sum_{p=1}^{s} a_{ilp} h_p(x_l) \quad (i=1, \ldots n), \tag{11}$$

where,

$f_i(x_1, \ldots, x_t)$ = a function that represents the probability that an examinee with latent trait values $x_1, \ldots, x_t$ will correctly respond to the ith binary item;

a_{i0} = An intercept parameter of the regression function for item i;

a_{ilp} = A regression coefficient for item i on latent trait l of the p-th polynomial degree;

$h_p(x_l)$ = a general polynomial function of the form,

$$f_{i1}\theta_j + f_{i2}\theta_j^2 + \ldots + f_{ik}\theta_j^k. \tag{12}$$

An IRT model which describes the probability that a randomly selected examinee j of ability θ_j will correctly answer an item is the two-parameter normal ogive model. The item characteristic curve (ICC) for the model is given by,

$$P_i(\theta_j) = \int_{-\infty}^{Z_{ij}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \tag{13}$$

where t is the normal deviate. One common parameterization of Z_{ij} for item i is,

$$Z_{ij} = a_i(\theta_j - b_i), \tag{14}$$

where b_i and a_i have been previously defined in (2). McDonald (1967), using harmonic analysis, has shown that the normal ogive model could also be approximated as closely as desired by a polynomial series of the general form,

$$f_{i0} + f_{i1}\theta_j + f_{i2}\theta_j^2 + \ldots f_{ik}\theta_j^k \tag{15}$$

where, f_{ik} is the factor loading of factor k on item i.

The unweighted least squares (ULS) function that is minimized to enable the estimation of the pairwise probabilities $\pi_{ij} = P(X_i=1, X_j=1)$ is,

$$f(\mathbf{A}, \beta) = \sum_{i<j}\sum [p_{ij} - \hat{\pi}_{ij}(\alpha_i, \beta_i, \alpha_j, \beta_j)]^2, \tag{16}$$

with $\mathbf{A}$, β and p_{ij} previously defined in (7) and (8). As was stated earlier, the ICFs for this model are approximated by a third-degree Hermite-Tchebycheff polynomial. A few advantages and limitations of the model are presented in the next section of the paper.

Advantages and limitations of McDonald's polynomial approximation to a normal ogive model

As was previously stated, McDonald's (1967) approach to NLFA uses ULS estimation of the model parameters. ULS estimation is quite economical as compared to generalized least-squares and maximum likelihood procedures and hence has the practical advantage of allowing for the analysis of tests with a fairly large number of items and/or dimensions.

Also, McDonald's model has been implemented in the computer program NOHARM (Fraser & McDonald, 1988). The program enables the user to fit confirmatory or exploratory unidimensional and multidimensional models to item response matrices. The output from a typical NOHARM run includes the results for the latent trait parameterization, the common factor model reparameterization as well as, in the unidimensional case, Lord's parameterization (i.e., a vector of discrimination parameters, **a**, and difficulty parameters, ***b***, are provided). In addition, a residual joint-proportions matrix is included in the output which can be useful to assess the fit of a given model.

However, the greater degree of computational efficiency associated with the ULS estimation procedure is achieved at the sacrifice of information (Mislevy, 1986). That is, only the information in the one-way marginals (percent-corrects) and two-way marginals (joint percent-corrects) is utilized by NOHARM in the estimation of parameters, thus explaining why it is often referred to as a "limited" or "bivariate" factor analytic method. However, McDonald (in press) and Muthen (1978) have suggested that one should not lose too much information in the absence of higher-order marginals. Also, Knol and Berger (1991) compared NOHARM parameter estimates to those obtained based on a full-information factor analytic model (i.e., using TESTFACT; Wilson, Wood & Gibbons, 1987) and generally found only slight differences between the two procedures with respect to their ability in recovering (simulated) factor analytic parameters. However, these findings were based on a limited number of replications (10) and should be interpreted cautiously. Nonetheless, from a practical perspective, it would seem that there might not be much to be gained in using full-information methods. Balassiano and Ackerman (1995b) have also shown that the overall performance of NOHARM, with respect to recovering simulated item parameter values, was satisfactory, even with small sample sizes (N = 200).

Another limitation of the model, again attributable to the ULS estimation procedure, is the absence of standard errors for the parameter estimates and a fit statistic for the given model. However, McDonald (1994) and Balassiano and Ackerman (1995b) have suggested criteria (e.g., the inverse of the square root of the sample size) that may be used as approximate standard errors for the parameters of the model. Also, two approximate χ^2 statistics, based on the residuals obtained after fitting a NLFA (NOHARM) model to an item response matrix, were proposed and investigated by De Champlain (1992) and Gessaroli and De Champlain (1995). Results obtained with a variety of simulated data sets showed that the approximate χ^2 statistics were quite accurate in correctly determining the number of factors underlying simulated item responses. This would suggest that these procedures might be useful as practical guides for the assessment of model fit, even though they are perhaps not the theoretically preferred statistics due to the ULS estimation method on which they're based. However, further research needs to be undertaken in order to evaluate the behavior of these approximate χ^2 in a larger number of conditions before making any definite statements about their usefulness.

Finally, some authors have noted that a problem with McDonald's model is the absence of an index that would indicate the appropriate number of polynomials to retain in a series (Hambleton & Rovinelli, 1986). Findings pertaining to this question, however, seem to indicate that terms beyond the cubic can generally be dismissed (McDonald, 1982b, Nandakumar, 1991).

A factor analytic model for dichotomous variables

Christoffersson (1975) and Muthén (1978) proposed a factor analytic model for dichotomous variables in which it is postulated that response variables X_i are accounted for by the latent continuous variables Y_i and threshold variables λ_i such that,

$$X_i = 1, \text{ if } Y_i > \lambda_i$$
$$X_i = 0, \text{ otherwise,}$$

where,

$$\boldsymbol{Y} = \Lambda\theta + \boldsymbol{E}, \tag{17}$$

and $\mathbf{Y} = (Y_i, \ldots, Y_n)'$. The model outlined in (29) is identical to the common factor model with the exception that $\mathbf{Y}$ is unobserved. Assuming that $\theta \sim \mathrm{MVN}(0,\mathbf{I})$, $\mathbf{E} \sim \mathrm{MVN}(0,\Psi^2)$, where Ψ^2 is a diagonal matrix of residual variances, and $\mathrm{cov}(\theta,\mathbf{E}) = 0$, the covariance matrix $\mathbf{\Sigma}$ among the $\mathbf{Y}$ latent variables can expressed as,

$$\Sigma(\boldsymbol{Y}) = \Lambda\Phi\Lambda' + \Psi. \tag{18}$$

Therefore,

$$\boldsymbol{Y} \sim MVN(0, \Lambda\Phi\Lambda' + \Psi^2), \tag{19}$$

The probability of a correct response based on Christoffersson's model is given by,

$$P(y_i=1) = \int_{h_i}^{\infty} \frac{1}{(2\pi)^{1/2}} e^{-x^2/2} dx. \tag{20}$$

The probability of correctly answering a pair of items is given by,

$$P(y_i=1,\ Y_j=1) = \int_{h_i}^{\infty}\int_{h_j}^{\infty} \frac{1}{2\pi|\Sigma_{ij}|^{1/2}} e^{-x'\Sigma_{ij}^{-1} x/2} dx. \tag{21}$$

Christoffersson (1975), using the tetrachoric expansion (Kendall, 1941) reexpresses (21) as,

$$P(y_i=1,\ y_j=1)=\sum_{s=0}^{\infty}\sigma_{ij}^{s}\ \tau_s(h_i)\tau_s(h_j)\ , \tag{22}$$

where τ_s is the s-th tetrachoric function. Given the rapid convergence of the series, Christoffersson (1975) states that we may cut the expansion after L terms and use,

$$\hat{P}(y_i=1,\ y_j=1)=\sum_{s=0}^{L-1}\sigma_{ij}^{s}\ \tau_s(h_i)\tau_s(h_j)\ . \tag{23}$$

The parameters of Christoffersson's (1975) model can be estimated using a generalized least-squares (GLS) estimation procedure that minimizes the fit function,

$$F=(\boldsymbol{p}-\boldsymbol{P})'S_e^{-1}(\boldsymbol{p}-\boldsymbol{P})\ , \tag{24}$$

where,

S_e= a consistent estimator of Σ_e, the residual covariance matrix;

P= a vector of expected item proportions correct P_j and joint item proportions P_{jk};

p= a vector of observed item proportions correct p_j and joint item proportions p_{jk};

Muthén (1978; 1983; 1984; 1988) has proposed a GLS estimator that is equivalent to that outlined by Christoffersson (1975) but computationally more efficient. According to Muthén (1978), the parameters of the factor analytic model for dichotomous variables can be estimated by minimizing the weighted least-squares fit function,

$$F=\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\sigma})'W_\delta^{-1}(\boldsymbol{s}-\boldsymbol{\sigma})\ , \tag{25}$$

where,

σ= Population threshold and tetrachoric correlation values;

s= Sample estimates of the threshold and tetrachoric correlation values;

W_δ= A consistent estimator of the asymptotic covariance matrix of **s**, multiplied by the total sample size.

This approach, also referred to as GLS estimation using a full-weight matrix approach (Muthén, 1988), is asymptotically equivalent to Christoffersson's solution and slightly less demanding in terms of computational requirements. It is referred to as a full-weight matrix approach because, as Muthén (1988) states, the GLS estimator utilizes a weight matrix of size p^* x p^*, where p^* corresponds to the total number of elements in the **s** vector.

<u>Advantages and limitations of Christoffersson's / Muthén's factor analytic model for dichotomous variables</u>

The GLS estimation procedure, unlike ULS, utilizes not only terms from the one-way and two-way margins but also from the three-way and four-way margins, that is, the joint proportions correct for three and four items taken at the same time. As Mislevy (1986) states, the use of a greater amount of information in the estimation procedure is especially advantageous when one attempts to extract more from the data, that is, with solutions that contain fewer items, examinees or more factors (with other conditions held constant).

Also, statistical tests of model fit are readily available. The F function minimized in the GLS solution (c.f. equations 33 + 34) asymptotically follows a chi-square distribution, with $df = k(k-1)/2 - t$, where k is equal to the number of items and t, the number of parameters estimated in the model. In addition, standard errors for the parameters estimated in the model can be obtained quite easily.

Finally, Muthén's solution is incorporated in the computer program LISCOMP (Muthén, 1988). As was the case with NOHARM (Fraser & McDonald, 1988), LISCOMP (Muthén, 1988) enables the user to fit both exploratory and confirmatory unidimensional or multidimensional models. Also, the output from a typical LISCOMP run contains the common factor model parameter estimates and standard errors as well as a residual correlation matrix and a chi-square statistic which allows the user to assess the degree of fit of a given model or competing models.

However, the GLS estimation is computationally very intensive. Although Muthén's (1978) solution is more efficient than Christoffersson's (1978), the procedure, as implemented in LISCOMP (Muthén, 1988), is still impractical using a personal computer with tests containing more than 25 items (Mislevy, 1986; Muthén, 1988).

Also, though GLS makes use of more information to fit the one- and two-way margins than does ULS, it still ignores higher level interactions and, in that sense, does not fully utilize all of the available information. However, as was the case for ULS estimation, it is quite possible that this loss of information is inconsequential.

Full-information item factor analysis

Bock and Aitkin (1981) proposed, based on the following m-factor model (for dichotomous data),

$$y_{ji}=\lambda_{i1}\theta_{1j}+\lambda_{i2}\theta_{2j},\ \ldots\ ,\lambda_{im}\theta_{mj}+\varepsilon_{ji}, \tag{26}$$

that an unobservable response process Y_{ji} for person j to item i is a linear function of m normally distributed latent variables $\theta_j = [\theta_{1j}, \theta_{2j}, \ldots, \theta_{mj}]$ and factor loadings $\boldsymbol{\lambda}_i = [\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{im}]$. This latent response process, y_{ji} is related to the binary (observed) item response x_{ji} through a threshold parameter, γ_i for item i, in the following fashion:

$$\text{if } y_{ji} \geq \gamma_i, \text{ then } x_{ji} = 1,$$

$$\text{if } y_{ji} < \gamma_i, \text{ then } x_{ji} = 0.$$

The probability that examinee j with abilities $\theta_j = [\theta_{1j}, \theta_{2j}, \ldots, \theta_{mj}]$ will correctly answer item i is given by the function,

$$P(x_{ji}=1|\theta_j)=\Phi(\gamma_i-\sum_{k=1}^{m}\lambda_{ik}\theta_{kj}\ /\ \sigma_i), \tag{27}$$

where Φ corresponds to unit normal cumulative distribution and σ_i is the standard deviation of the unobserved random variable $\varepsilon_{ji} \sim N(0,\sigma^2{}_i)$.

Bock and Aitkin (1981) proposed a marginal maximum likelihood (MML) procedure to estimate the parameters in the model based on Dempster, Laird and Rubin's (1977) EM algorithm. The threshold and factor loadings are estimated so as to maximize the following function,

$$L_m=P(X)=\frac{N!}{r_1!,r_2!,\ldots,r_s!}\ \tilde{P}_1^{r_1},\tilde{P}_2^{r_2},\ldots,\tilde{P}_s^{r_s}, \tag{28}$$

where, r_s is the frequency of response pattern s and $\tilde{P}_s$ is the marginal probability of the response pattern based on the item parameter estimates. The function outlined in (27), with the MML parameter estimates by means of the EM algorithm, is commonly referred to as *full-information item factor analysis* (Bock, Muraki, & Gibbons, 1988) and has been implemented in the computer program TESTFACT (Wilson, Wood, & Gibbons, 1987).

Advantages and limitations of Bock and Aitkin's (1981) / Bock, Gibbons and Muraki's (1988) full-information item factor analysis

One of the key advantages of full-information item factor analysis (FIFA) is that it utilizes all available information in the estimation procedure. Contrary to the two least-squares models previously outlined, which are restricted to lower-order marginals, FIFA is based on the estimation of item response vectors and hence uses all available information in the data.

Also, the procedure is implemented in the computer program TESTFACT (Wilson, Wood, & Gibbons, 1987). The output from a TESTFACT analysis contains, among other things, classical item statistics and factor analytic parameter estimates as well as their associated standard errors. In addition, a likelihood-ratio chi-square test is provided to help the user determine the fit of a model, or of competing models.

However, the use of all information contained in the 2^p item vectors, where p is equal to the number of items, by FIFA requires that there should be no empty cells which is usually not feasible unless some collapsing is done. In addition, as Mislevy (1986) and Berger & Knol (1990) have noted, the G^2 goodness-of-fit statistic computed by TESTFACT will be very unreliable with data sets containing more than 10 items due to the small expected number of examinees per cell. More precisely, Mislevy (1986) states that the approximation to the chi-square distribution might be poor in this instance. Wilson, Wood and Gibbons (1987) also caution against relying on the G^2 fit statistic when a large number of cells have expected frequencies near zero. In that instance, the authors recommend using the G^2 difference test (comparing two specific models) given that it follows a chi-square distribution in large samples, even in the presence of a sparse frequency table.

Conclusion

IRT models have been used extensively in the past few decades not only in the development and analysis of educational test items but also in a host of other applications such as for the equating of alternate test forms and the detection of differentially functioning items.

Several researchers have suggested, however, that common IRT models are really specific cases of a more general NLFA model (Goldstein & Wood, 1989; Knol & Berger, 1991; McDonald, 1967; in press; Takane & De Leeuw, 1987). The research conducted by the latter authors clearly shows that familiar IRT models, such as the normal ogive and logistic functions, can easily be expressed with a factor analytic parameterization. The findings obtained in these studies would therefore seem to suggest that NLFA might provide a useful framework with which to address measurement-related issues that had been primarily investigated using IRT models.

Three factor analytic models were briefly outlined. More precisely, McDonald's (1967; 1982b) polynomial approximation to a normal ogive model, Christoffersson's (1975)/Muthén's (1978) factor analysis model for dichotomous variables and Bock and Aitkin's (1981)/Bock, Gibbons & Muraki's (1988) full-information factor analytic model, were described. In addition, the major strengths and weaknesses of each model were delineated. Based on this information, are there any conditions that might dictate the use of one model over another?

The main advantage associated with McDonald's polynomial approximation to a normal ogive model, that is, the relative economy of the ULS estimator, also constitutes its primary shortcoming. In other words, as Mislevy (1986) stated, the higher degree of computational efficiency is achieved at the sacrifice of information. The model utilizes lower-order marginals in the estimation process and consequently ignores higher-order relationships among the data. However, there is some empirical evidence to suggest that "limited-information" factor analytic parameter estimates do not differ substantially from those obtained using the theoretically sounder "full-information" method as implemented in the computer program TESTFACT (Wilson, Wood, & Gibbons, 1987; Knol & Berger, 1991).

Also, the absence of standard errors for the estimated factor analytic parameters and of a fit statistic to gauge the overall adequacy of a model, are major disadvantages of McDonald's model, as implemented in the computer program NOHARM (Fraser & McDonald, 1988). Nonetheless, approximate standard errors have been proposed as useful guides in assessing parameter estimation accuracy (Balassiano & Ackerman, 1995b; McDonald, 1994). Also, two approximate chi-square statistics, based on the residual matrix obtained after fitting an *m*-factor model to an item response matrix using NOHARM (Fraser & McDonald, 1988) proved to be very accurate with respect to correctly identifying the number of dimensions underlying simulated data sets in specific conditions. Of course, these chi-square statistics are weak in their theoretical foundation due to the fact that they're based on ULS estimation. However, Browne (1977) has indicated that in many cases, these chi-square statistics are formally equivalent to those derived from a GLS estimation and that they differ only slightly. Therefore, from a practical perspective, these approximate chi-square statistics might be useful tools to those interested in fitting McDonald's model to item response matrices.

The factor analytic model for dichotomous variables proposed by Christoffersson (1975) and amended by Muthén (1978) is, from a theoretical standpoint, superior to McDonald's approach in that the GLS estimation procedure yields a valid chi-square goodness-of-fit statistic as well as legitimate standard errors for the estimated parameters. Nonetheless, the model is still based on "limited information" in that it ignores higher-level interactions in the data. Also, the computational requirements of Muthén's GLS solution as implemented in LISCOMP (Muthén, 1988) are quite exacting: they increase proportionally to the number of factors and with the fourth power of the number of items. This led Mislevy (1986) to suggest that Muthén' solution might be adopted with tests that have a relatively small item to factor ratio.

Finally, the full-information factor analytic model proposed by Bock and Aitkin (1981) and Bock, Gibbons, an Muraki (1988) is, based on theoretical grounds, the strongest of the approaches outlined, given that it does, as the name implies, make use of all available information contained in the 2^p item response vectors. However, in most applications, the use of the full-information is usually not feasible unless collapsing of cells is undertaken. Also, the computational requirements associated with the MML estimation procedure implemented in TESTFACT, increase geometrically with the number of factors specified in the model but only linearly with the number of items and response vectors. Hence, Mislevy (1986) advises using this procedure with longer tests and more parsimonious models.

In summary, it would appear as though a greater number of empirical studies should be undertaken to compare the various NLFA models with respect to how accurately they can recover simulated parameter values, before making any definite recommendations as to which procedure should be favored given a specific set of conditions, i.e., test length, sample size, factor model, etc. It is possible that the theoretically sounder models, e.g., full-information factor analysis, might not yield substantially more accurate parameter estimates than methods that are based on lower-order marginals, e.g., McDonald's (1967; 1982b) and Muthén's (1978; 1988) approaches. There is some preliminary evidence to support this claim (Boulet & Gessaroli, 1992; Gibbons, 1984; Knol & Berger, 1991). Some authors have even suggested that factor loadings obtained from a linear factor analysis of phi and tetrachoric correlation matrices did not differ noticeably from those derived using LISCOMP (Muthén, 1988; Parry & McArdle, 1991) or TESTFACT (Wilson, Wood, & Gibbons, 1987; Knol & Berger, 1991). However, more studies are needed in this area to clearly identify the conditions in which one method might outperform another.

In addition, the usefulness of these models in addressing common measurement-related problems should be investigated. For example, LISCOMP (Muthén, 1988) and TESTFACT (Wilson, Wood, & Gibbons, 1987) provide chi-square goodness-of-fit statistics in order to aid the practitioner in determining which model best accounts for the item response probabilities. Also, similar fit statistics have been proposed to accompany McDonald's (1967; 1982b) NLFA model (De Champlain, 1992; Gesaroli & De Champlain, 1995). Given that unidimensionality of the latent ability space is one of the main postulates underlying most IRT models, it would seem important to evaluate the degree of accuracy with which each of these fit statistics is able to correctly identify or reject this assumption under a variety of simulated conditions. Similarly, it might be interesting to assess the degree of effectiveness of these fit statistics in detecting violation of local independence. A frequent problem that confronts practitioners is how to best model item response data that contain sets that is, where several items refer to a common stem, e.g., a reading comprehension passage. The factor analytic framework might provide the means of effectively dealing with local item dependence through items loading on a secondary dimension, for example. The usefulness of the NLFA framework in addressing these types of issues will be illustrated in the next three presentations of this symposium.

An overview of the three symposium presentations

The first paper will be centered on outlining methods available for the assessment of dimensionality that are based on NLFA. Examples of how to use these procedures to test for specific dimensional structures will be illustrated using data from a national testing program.

The next paper will compare the degree of accuracy of parameter estimates when based on "limited-information" (NOHARM) and "full-information" (TESTFACT) factor analytic models for simulated unidimensional and multidimensional data sets. In addition, the use of these methods will be depicted with actual achievement test data.

The final paper will focus on explaining how the factor analytic framework might be useful in dealing with local item dependence (LID). Specifically, the identification of LID using NLFA will once more be illustrated with data from a national testing program. Also, methods of obtaining "purified" estimates of reliability and standard errors of measurement as well as ability (i.e., without LID contamination) will be outlined.

It is hoped that these presentations will underscore the usefulness of NLFA in addressing the above mentioned problems with actual achievement test data and stimulate discussion with respect to these areas, thus hopefully fostering future research.

References

Baker, F.B. (1985). The basics of item response theory. Portsmouth, NH: Heinemann.

Baker, F.B. (1992). Item response theory: Parameter estimation techniques. New-York, NY: Marcel Dekker Inc.

Balassiano, M., & Ackerman, T. (1995a). An in-depth analysis of the NOHARM estimation algorithm and implications for modeling the multidimensional latent ability space. Unpublished manuscript, University of Illinois at Urbana-Champaign, Faculty of Education, Urbana-Champaign.

Balassiano, M., & Ackerman, T. (1995b). An evaluation of NOHARM estimation accuracy with a two-dimensional latent space. Unpublished manuscript, University of Illinois at Urbana-Champaign, Faculty of Education, Urbana-Champaign.

Bartholomew, D.J. (1983). Latent variable models for ordered categorical data. Journal of Econometrics, 22, 229-243.

Berger, M.P.F., & Knol, D.L. (1990, April). On the assessment of dimensionality in multidimensional Item Response Theory models. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Bock, R.D. (1984). Full information item factor analysis. Paper presented at the meeting of the Psychometric Society, Santa Barbara, CA.

Bock, D.R., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. Psychometrika, 4, 443-459.

Bock, D.R., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. Applied Psychological Measurement, 12(3), 261-280.

Bock, R.D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. Psychometrika, 35, 179-197.

Boulet, J.R., & Gessaroli, M.E. (1993, June). The effect of nonnormal ability distributions on IRT parameter estimation using "full-information" and "limited-information" methods. Paper presented at the meeting of the Canadian Educational Research Association, Ottawa, ONT.

Browne, M.W. (1977). The analysis of patterned correlation matrices by generalized least squares. British Journal of Mathematical and Statistical Psychology, 30, 113-124.

Christoffersson, A. (1975). Factor analysis of dichotimized variables. Psychometrika, 40(1), 5-32.

Cook, L.L., & Eignor, D.R. (1983). Practical considerations regarding the use of item response theory to equate tests. In R.K. Hambleton (Ed.), Applications of item response theory, Vancouver, B.C.: Educational Research Institute of British Columbia, 175-195.

De Champlain, A. (1992). Assessing test dimensionality using two approximate chi statistics. Unpublished doctoral dissertation, University of Ottawa, Ottawa.

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.

Fraser, C., & McDonald, R.P. (1988). NOHARM: Least squares item factor analysis. Multivariate Behavioral Research, 23, 267-269.

Gessaroli, M.E. (in press). The assessment of dimensionality via local and essential independence: A comparison in theory and practice. To be published in D. Laveault, B.D. Zumbo, M.E. Gessaroli, & M.W. Boss (Eds.), Modern Theories in Measurement: Problems and Issues. Ottawa, Ontario: Edumetrics Research Group, University of Ottawa.

Gessaroli, M.E., & De Champlain, A. (1995). Using an approximate chi-square statistic to test for the number of dimensions underlying the responses to a set of items. Manuscript submitted for publication.

Gibbons, R.D. (1984). Multivariate probit analysis: A general model. Paper presented at the meeting of the Psychometric Society.

Goldstein, H., & Wood, R. (1989). Five decades of Item Response Modelling. British Journal of Mathematical and Statistical Psychology, 42, 139-167.

Hambleton, R.K. (1983). Applications of item response theory. Vancouver, BC: Educational Research Institute of British Columbia.

Hambleton, R.K. (1989). Principles and selected applications of item response theory. In R.L. Linn (Ed.), Educational Measurement (pp. 147-200). New York, NY: American Council on Education and Macmillan Publishing Company.

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

Hambleton, R.K., & Swaminathan, H. (1985). Item response theory: Principles and applications, Boston, MA: Kluwer-Nyjhoff.

Hambleton, R.K., Zaal, J., & Pieters, J.P.M. (1993). Computer adaptive testing: Theory, Applications, and standards. In R.K. Hambleton and J.N. Zaal (Eds.), Advances in educational and psychological testing: Theory and applications (pp. 341-366). Boston, MA: Kluwer Academic Publishers.

Hulin, C.L., Drasgow, F. & Parsons, L.K. (1983). Item response theory. Homewood, IL: Dow-Jones Irwin.

Kendall, M.G. (1941). Relations connected with the tetrachoric series and its generalization. Biometrika, 32, 196-198.

Kingsbury, G.G., & Zara, A.R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. Applied Measurement in Education, 4, 241-261.

Knol, D.L., & Berger, M.P.F. (1991). Empirical comparison between factor analysis and multidimensional item response models. Multivariate Behavioral Research, 26, 457-477.

Lord, F.M. (1977). Practical applications of characteristic curve theory. Journal of Educational Measurement, 14, 117-138.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum, Inc. Lord, F.M. (1982). Item response theory and equating - A technical summary. In P.W. Holland and D.B. Rubin (Eds.), Test equating (pp. 141-148). New York, NY: Academic Press.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

McDonald, R.P. (1967). Nonlinear factor analysis. Psychometrika Monograph No. 15, 32(4, Pt. 2).

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

McDonald, R.P. (1982a). Some alternative approaches to the improvement of measurement in education and psychology: Fitting latent trait models. In D. Spearitt (Ed.), The improvement of measurement in education and psychology (pp. 213-233). Hawthorn, VI: Australian Council for Educational Research.

McDonald, R.P. (1982b). Linear versus nonlinear models in item response theory. Applied Psychological Measurement, 6, 379-396.

McDonald, R.P. (1985). Factor analysis and related methods. Hillsdale, NJ: Lawrence Erlbaum Associates.

McDonald, R.P. (1989). Future directions for item response theory. International Journal of Educational Research, 13, 205-220.

McDonald, R.P. (in press). Testing for approximate dimensionality. Paper presented at the International Symposium on Modern Theories in Measurement: Problems and Issues, Montebello, PQ.

McDonald, R.P. (In press). Normal ogive multidimensional model.

McDonald, R.P., & Ahlawat, K.S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82-99.

McKinley, R.L. (1988, April). Assessing dimensionality using confirmatory multidimensional IRT. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

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

Mislevy, R.J., & Bock, R.D. (1990). BILOG 3: Item analysis and test scoring with binary logistic models. Mooresville, In.: Scientific Software, Inc.

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

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

Muthén, B. (1983). Latent variable structural equation modelling with categorical data. Journal of Econometrics, 22, 43-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Muthén, B. (1988). LISCOMP. Mooresville, IN: Scientific Software.

Nandakumar, R. (1991a, April). Assessing the dimensionality of a set of items - Comparison of different approaches. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Il.

Parry, C.D.H., & McArdle, J.J. (1991). An applied comparison of methods for least-squares factor analysis of dichotomous variables. Applied Psychological Measurement, 1, 35-46.

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming and equating. In R.L. Linn (Ed.), Educational Measurement (pp. 221-262). New York: American Council on Education and Macmillan Publishing Company.

Skaggs, G., & Lissitz, R.W. (1986). IRT test equating: Relevant issues and a review of recent research. Review of Educational Research, 4, 495-529.

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

Thissen, D. (1993). MULTILOG 6. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Thissen, D., Steiberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P.W. Holland and H. Wainer (Eds.), Differential item functioning (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R., Steinberg, L., & Thissen, D. (1990). Computerized Adaptive Testing: A primer. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Warm, T.A. (1978). A primer of Item Response Theory (Tech. Rep. No. OG-941278) Oklahoma City, OK: U.S. Coast Guard Institute.

Wilson, D., Wood, R., & Gibbons, R.D. (1987). TESTFACT: Test scoring, item statistics, and item factor analysis. Mooresville, IN: Scientific, Software, Inc.

Wingersky, M.S., Patrick, R., & Lord, F.M. (1991). LOGIST VI user's guide. Princeton, NJ: Educational Testing Service.

Appendix A

Takane and De Leeuw's (1987) IRT-NLFA proof

Let $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ be a random vector of response patterns to n binary items on a test. Each $\tilde{x}_i$ is assigned a value of 1, if the examinee correctly answers the item, or 0, if there is an incorrect response. Let $\tilde{\mathbf{u}}$ be an m-component random vector of abilities ($m \leq n$) with its density function denoted by $g(\mathbf{u})$. $\tilde{\mathbf{u}}$ is unobservable directly, but is assumed to follow a multivariate normal distribution with mean **0** and covariance $\boldsymbol{I}$ (identity matrix); that is $\tilde{\mathbf{u}} \sim \boldsymbol{N}(\mathbf{0}, (\boldsymbol{I}))$. The domain of $\tilde{\mathbf{u}}$ (denoted by U) is the multidimensional region defined by the direct product of $(-\infty, \infty)$. In IRT, the two-parameter normal ogive model specifies the marginal probability that $\tilde{\mathbf{x}} = \mathbf{x}$ (Bock & Aitkin, 1981; Bock & Lieberman, 1970) as,

$$Pr(\tilde{\boldsymbol{x}}=x) = \int_U Pr(\tilde{\boldsymbol{x}}=x|\mathbf{u})\, g(\mathbf{u})\, d\mathbf{u}, \tag{29}$$

where $\Pr(\tilde{\mathbf{x}}=\mathbf{x}|\mathbf{u})$ is the conditional probability of observing response pattern $\mathbf{x}$ given $\tilde{\mathbf{u}} = \mathbf{u}$. Also, it is assumed that,

$$Pr(\tilde{\boldsymbol{x}}=\boldsymbol{x}|\mathbf{u}) = \prod_i^n (p_i(\mathbf{u}))^{x_i}(1-p_i(\mathbf{u}))^{1-x_i}, \tag{30}$$

(that is, local independence) with,

$$p_i(\mathbf{u}) = \int_{-\infty}^{\mathbf{a}'\mathbf{u}+b} \phi(z)\, dz = \boldsymbol{\Phi}(\mathbf{a}'\mathbf{u}+b), \tag{31}$$

where ϕ is the density function of the standard normal distribution and Φ, the normal ogive function (i.e., the cumulative distribution function of the standard normal distribution).

On the other hand, Takane and De Leeuw (1987) state that in the factor analytic model proposed by Christoffersson (1975), the marginal probability of response pattern $\mathbf{x}$ is specified as,

$$Pr(\tilde{\boldsymbol{x}}=\boldsymbol{x}) = \int_R h(\boldsymbol{y})\, d\boldsymbol{y}, \tag{32}$$

where R is the multidimensional region of integration and

$$\tilde{\boldsymbol{y}} = C\tilde{\mathbf{u}} + \tilde{\boldsymbol{e}}. \tag{33}$$

Equation 15 corresponds to the common factor analytic model with **C** being the matrix of factor loadings, $\tilde{\mathbf{u}}$, the vector of factor scores (abilities in an IRT framework) and $\tilde{\mathbf{e}}$, the random vector of uniqueness components distributed as $N(\mathbf{0},Q^2)$ where Q^2 is further assumed to be diagonal (linear local independence), and $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{e}}$ are independent of each other. It follows that,

$$\tilde{\boldsymbol{y}} \sim N(\mathbf{0}, CC' + Q^2), \tag{34}$$

(marginal distribution of $\tilde{\mathbf{y}}$) and

$$\tilde{\boldsymbol{y}}|\mathbf{u} \sim N(C\mathbf{u}, Q^2), \tag{35}$$

(conditional distribution of $\tilde{\mathbf{y}}$ given $\tilde{\mathbf{u}} = \mathbf{u}$). The continuous random variables, $\tilde{\mathbf{y}}$ are dichotomized by $\tilde{x}_i = 1$, if $\tilde{y}_i \geq r_i$ or $\tilde{x}_i = 0$, if $\tilde{y}_i < r_i$ for $i=1,\ldots,n$, where r_i is the threshold parameter for variable i. Therefore, R, the region of integration above, is the multidimensional parallelopiped defined by the direct product of intervals, $R_i = (r_i, \infty)$ if $\tilde{\mathbf{x}}_i=1$ and $R_i=(-\infty, r_i)$ if $\tilde{\mathbf{x}}_i = 0$. Now (11) including (12) and (13) is equivalent to (14) with $\tilde{\mathbf{y}}$ defined in (15). We first prove that (14) $\rightarrow$ (11). The authors show that from (14) we have

$$Pr(\tilde{\boldsymbol{x}}=\boldsymbol{x}) = \int_R h(\boldsymbol{y})\, d\boldsymbol{y}$$

$$= \int_R \left(\int_U f(\boldsymbol{y}|\mathbf{u})\, g(\mathbf{u})\, d\mathbf{u}\right) d\boldsymbol{y}$$

$$= \int_U g(\mathbf{u}) \left(\int_R f(\tilde{\boldsymbol{y}}|\mathbf{u})\, d\boldsymbol{y}\right) d\mathbf{u}, \tag{36}$$

where $f(\tilde{\mathbf{y}}|\mathbf{u})$ is the conditional density of $\tilde{\mathbf{y}}$ given $\tilde{\mathbf{u}}=\mathbf{u}$. But because of (17), it can be shown that,

$$\int_R f(\boldsymbol{y}|\boldsymbol{u})\,d\boldsymbol{y}=\prod_i \int_{R_i} f_i(y_i|\boldsymbol{u})\,dy_i$$

$$=\prod_i \left(\int_{r_i}^{\infty} f_i(y_i|\boldsymbol{u})\,dy_i\right)^{x_i}\left(1-\int_{r_i}^{\infty} f_i(y_i|\boldsymbol{u})\,dy_i\right)^{1-x_i}, \tag{37}$$

where,

$$\int_{r_i}^{\infty} f_i(y_i|\boldsymbol{u})\,dy_i=\boldsymbol{\Phi}\left(\frac{c_i^{\prime}\boldsymbol{u}-r_i}{q_i}\right), \tag{38}$$

for $i = 1,\ldots,n$. In this instance q^2_i is the i-th diagonal element of Q^2. Equation (19) is thus equivalent to (13) by setting

$$a_i=\frac{c_i}{q_i} \tag{39}$$

and

$$b_i=-\frac{r_i}{q_i} \tag{40}$$

for $i = 1,\ldots,n$.

Takane and De Leeuw (1987) state that it might appear as though factor analysis with c_i, r_i and $q_i (i=1,\ldots,n)$ has more parameters than IRT with only a_i and $b_i (i=1,\ldots n)$. However, according to the authors, when the data are dichotomous, the variance of $\tilde{y}_i$ cannot be estimated due to the lack of relevant information in the data, and thus, q_i can be set to an arbitrary value. Hence, the effective number of parameters is identical in both models. In conclusion, the authors mention that the equivalence of marginal probabilities in IRT and FA models holds approximately with logistic (IRT) models also, as long the logistic distribution provides a good approximation of the normal distribution (i.e. normal ogive).