Journal of Informatics and Mathematical Sciences
Vol. 11, No. 1, pp. 51–79, 2019
ISSN 0975-5748 (online); 0974-875X (print)
Published by RGN Publications
DOI: 10.26713/jims.v11i1.951

http://www.rgnpublications.com

Research Article

Effect of Measurement Scales on Results of Item Response Theory Models and Multivariate Statistical Techniques

B. K. Nkansah [1], A. Zakaria [2,*] and N. K. Howard [3]

Department of Statistics, University of Cape Coast, Cape Coast, Ghana
***Corresponding author:** zakaria.arimiyaw@ucc.edu.gh

Abstract. The study investigates the effects of response scales of items on results of item response theory models and multivariate techniques. A total of sixty-four datasets have been simulated under various conditions such as item response format, number of dimensions underlying response scales, and sample size using R package *mirt* command: $simdata(a,d,N,itemtype)$. Two main statistical techniques – Item Response Theory (IRT) models and Factor Analysis – are employed. We find that there is a direct relationship between parameters of IRT and those of factor models, particularly item discrimination and factor loadings. The results also show that the overall fitness of the item response model increases with increasing scale points for higher dimensionality and sample size 150 and higher. The fitness deteriorates over increasing scale points for small sample sizes for unidimensional model. Again, the number of influential indicators on factors increases with increasing scale-points which improves the fitness of the model. The study suggests that a five-point response scale gives most reasonable results among various scales examined. IRT analysis is recommended as a preliminary process to ascertain the observed features of items. The study also finds a sample size of 150 as adequate for a most plausible factor solution, under various conditions.

Keywords. Item response theory; Factor model; Scale points; Dimensionality

MSC. 62Hxx

Received: June 22, 2018 **Accepted:** January 7, 2019

1. Introduction

A scale is a continuum, consisting of the highest point and the lowest point along with several intermediate points (thresholds) between these two points. These scale-point positions are related in a manner that when the first point happens to be the lowest, the second point indicates a higher degree of a particular characteristic, followed by the third and so on. Scales may be classified in various ways, one of which is in respect of number of dimensions. On this basis, scales can be classified as unidimensional and multidimensional. Under the unidimensional scale, only one characteristic (ability) of the person is measured, whereas multidimensional scaling considers that an individual might be better described by several dimensions, rather than a single-dimension continuum [13]. Scales are employed in questionnaires to measure the characteristics of respondents such as abilities, opinions, preferences, and attitudes that are not directly observable [5]. These person characteristics, due to their covert nature, are referred to as latent. On a questionnaire, a scale is composed of response format and number of items (questions) which are indications of latent abilities. A person's response to a set of items is influenced by the characteristics of the individual and by the characteristics of the items. Modelling the relationship between item responses and the characteristics of persons falls under the realm of item response theory (IRT) models. The IRT models are quite useful in the construction of scales (e.g., Likert scale) for measuring latent constructs of persons.

An important issue to consider when designing Likert scale items is the optimal number of response categories. Considering reliability and validity, Jacoby and Matell attempted to determine the number of response alternatives to use in the construction of Likert-type scales [11]. They indicated that both reliability and validity are independent of the number of scale points used for Likert-type items. The authors suggested that two or three-point Likert scales are good enough. Martin studied the effects of varying the number of scale points on the correlation coefficient using the bivariate normal distribution [15]. Martin argued that the correlation coefficient generally decreases as the number of response categories becomes smaller, and suggested the use of ten to twenty points on a scale.

IRT analysis has been found to be highly influenced by sample size. Notably, the problem of estimation of item parameters has a link with sample size. In other words, how large a sample to be used in IRT analysis will depend on how many item parameters to be estimated. For complex IRT models that require estimation of more parameters, sample size should increase accordingly. The task of determining minimum sample size has been attempted by some researchers through simulation studies. Reise and Yu estimated the parameters of the graded response (GR) model, and recommended that a sample size of at least 500 is required to achieve adequate estimation [24]. For Rasch item response model, useful information can be obtained from samples as small as 100 and sample sizes of 500 are more than adequate in estimating item parameters [6]. Under two-parameter logistic (2PL) model, Stone found that with sample size of 500 or more and 20 or more items, both item difficulty and discrimination parameters are generally stable and precise [27]. Smith, Schumacker and Bush [18] examined the fitness of items using the mean square (MSQ) statistic and provided the following guidelines for sample size: misfit is

evident when MSQ values are larger than 1.3 for samples less than 500, 1.2 for samples between 500 and 1,000, and 1.1 for samples larger than 1,000 respondents.

Factor analysis, undoubtedly, an important multivariate statistical technique, is also widely applied in analysing questionnaire items. Within the context of the technique, individual items typically represent indicator variables, and the latent abilities that the questionnaire seeks to measure represent the factors. The factor analysis model is based on three basic assumptions about the indicator variables - normality, constant variance and linearity. The indicator variables are also considered to be measured on at least the interval scale When these assumptions are satisfied, the usual Pearson product-moment correlation coefficient provides a reliable measure of the extent of correlation between each pair of indicator variables, and the linear factor model reasonably fits the data.

However, a major concern in the literature (e.g., see [30]) has to do with the factor analysis of Likert-type data. Item responses give categorical data, which suggests a violation of the continuous nature of the indicator variables. The implication is that the Pearson correlations between pairs of indicator variables in this case are less reliable and is a potential source of distortions in the factor structure. The severity of the distortions tend to increase as the number of response categories on the items decreases [4]. As a remedy, Ferrando and Lorenzo-Seva recommended the use of tetrachoric correlations for factor analysis of dichotomous response data [9]. For factor analysis of ordered polytomous data, it is recommended to use polychoric correlations.

Another issue to consider when conducting factor analysis is the characteristics of the sample from which the measurements of the indicator variables are taken. Obviously, an aspect of the sample that is worth considering is how large the sample should be in order to perform factor analysis. Correlations are less reliable when estimated from small samples [28]. Gorsuch put it bluntly that "no one seems to know exactly where a large n begins and a small n leaves off" [10]. Comrey and Lee noted that as the sample size increases, the reliability of the obtained correlations increases [4]. They found that samples of size 50 give very inadequate reliability of correlation coefficients, while samples of size 1000 are more than adequate for factor analysis. With regards to evaluating the adequacy of the sample size, Comrey and Lee provided some guidelines: 50 is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1000 or greater is excellent [4]. Other researchers are of the view that under optimal conditions (communalities of 0.70 or greater and 3 to 5 indicator variables loading on each factor), a sample of size 100 can be adequate; under moderately good conditions (communalities of 0.40 to 0.70 and at least 3 indicators loading on each factor), a sample of at least 200 should suffice; and under poor conditions (communalities lower than 0.40 and some factors with only two indicator variables on them), samples of at least 400 might be necessary [8, 14, 28].

This research is another attempt at examining the influence of the number of points on the response scales of items on the results of IRT and how it translates into suitable factor structure. In addition, we will investigate the effect of sample size on the factor structure. In the next section, we study the theoretical connection between the two methods. Subsequent section will consider the simulation studies and the analysis and results.

2. Methods

Two broad analytical techniques have been considered. These are item response theory and factor analysis. Factor analysis is used in the study since it is based on correlations coefficients and identifies underlying constructs among a set of variables.

2.1 Item response models

Item response theory (IRT) models express the association between an individual's response to an item and the underlying latent variable (ability) being measured by the instrument (questionnaire) [22]. IRT uses latent characterisations of individuals and items as predictors of observed responses. The IRT describes, in probabilistic terms, how a person with higher ability level is likely to provide a response in a different response category in relation to a person with a low ability level [6, 19]. Each item is characterised by one or more model parameters: discrimination (α), difficulty (δ), and guess (c) parameters.

The item response theory models may be classified broadly in three essential ways. Firstly, in terms of the item characteristics or parameters that are included in the models. In this regard, some models are designed to account for one parameter (mostly, the difficulty parameter), while other more complex models account for two or more parameters. Secondly, IRT models can also differ in terms of the response option format. Along these lines, some models are designed to be used for dichotomous items, whereas others are designed for items with more than two response options (i.e. polytomous items), such as Likert scale items. Examples of dichotomous item response models are the Rasch model, one-parameter logistic (1PL), two-parameter logistic (2PL), and three-parameter logistic (3PL) models. Examples of polytomous item response models are the partial credit, the rating scale, the graded response, and the nominal models. Thirdly, IRT models are classified in terms of the number of dimensions that defines the person ability parameter. In this case, each of the dichotomous and polytomous item response models is either unidimensional or multidimensional.

The two-parameter logistic (2PL) model, in unidimensional sense, is defined as

$$p\left(X_{ij}=1|\theta,\alpha,\delta\right)=\frac{1}{1+\exp[-1.702\alpha_i(\theta-\delta_i)]} \tag{1}$$

In this model, probability of response to a dichotomous item is determined by the individual's ability level (θ) and two item parameters – discrimination (α) and difficulty (δ), and items have different discrimination powers, α_i. High values of α_i result in steeper item characteristic curve (a graphical representation of probabilities of response to an item). For 2PL model, the guess parameter (c) is set to zero, since Likert-type items are not scored as right or wrong [18]. Thus, when the value of c is constrained to zero, it would facilitate the comparison of the results of two-point and higher-point scales. In Equation (1), the value 1.702 is a scaling factor that ensures the value of the item discriminating parameter in logistic models compares to a normal-ogive model. This scaling is important for linking IRT parameters with factor analysis results [23].

Multidimensional equivalent of Equation (1), known as multidimensional two-parameter logistic (M2PL) model, is given as [21]

$$p\left(X_{ij}=1|\boldsymbol{\theta}_j,\boldsymbol{\alpha}_i,d_i\right)=\frac{1}{1+\exp\left[-1.702(\boldsymbol{\alpha}_i'\boldsymbol{\theta}_j+d_i)\right]} \tag{2}$$

where $\boldsymbol{\alpha}_i$ is a vector of discrimination parameters for item i and $d_i=-\mathbf{1}'\boldsymbol{\alpha}_i\delta_i$ is a scalar parameter that is related to the item's difficulty. The exponent in Equation (2) can be expanded as

$$\boldsymbol{\alpha}_i'\boldsymbol{\theta}_j+d_i=\sum_{l=1}^{m}\alpha_{il}\theta_{jl}+d_i \tag{3}$$

which indicates how the elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ vectors interact. Equation (3) shows that the exponent in the M2PL model is a linear combination of the elements of $\boldsymbol{\theta}$. This feature reflects the compensatory nature of the M2PL model. Equation (3) defines a line in an m-dimensional space. If the exponent is set to some constant, k, that is

$$k=\boldsymbol{\alpha}_i'\boldsymbol{\theta}_j+d_i \tag{4}$$

then all θ-vectors satisfying Equation (4) will fall along the same straight line with the same probability of a favourable response for the model.

Categorical data can be described effectively in terms of the number of categories into which data can be placed. For ordered polytomous items, the response categories have an explicit rank ordering with respect to the ability. Ordered categories are defined by boundaries that separate the categories. Intuitively, there is always one less boundary than there are categories. For instance, a five-point Likert-type item requires four boundaries to separate the five possible response categories [19]. In general, each response variable X_{ij}, $i=1,2,\ldots,p$; $j=1,2,\ldots,n$, has r_i+1 response categories represented by category scores $k=\{0,1,2,\ldots,g,\ldots,r_i\}$ and r_i boundaries denoted by $h=\{1,2,\ldots,g,\ldots,k\}$.

Polytomous item response models result in a general expression for the probability of a person responding in a given response category. Mathematically, the various polytomous models for ordered response categories differ in terms of the expressions that are used to represent the location parameter (δ) of the category boundaries.

For polytomous items, we consider the generalised partial credit (GPC) model. The GPC model applies 2PL concept to ordered categorical responses. In GPC model, the probability of observing a response in category g over category $g-1$ for item i is given as [17]

$$P\left(X_{ij}=g|\theta,\alpha_i,\delta_{ih}\right)=\frac{\exp\left[\sum_{h=0}^{g}\alpha_h(\theta-\delta_{ih})\right]}{\sum_{k=0}^{r_i}\exp\left[\sum_{h=0}^{k}\alpha_h(\theta-\delta_{ih})\right]} \tag{5}$$

where, α_h denotes the discrimination associated with response category h on item i.

2.2 Factor analysis models

Factor analysis is a multivariate statistical technique that is employed to discover which variables (indicators) in a set form meaningful subsets that are relatively independent of one

another. Variables that are correlated with one another but largely independent of other subsets of variables are combined into factors (abilities in IRT). These factors are thought to reflect underlying processes that have created the correlations among the variables [28]. In this paper, for convenience of keeping track of the factor, we will consider in most places a one-factor model. A one-factor model uses only one factor to explain the correlations among the indicators. It seeks to model the relationship between the underlying factor θ and the continuous response variable Y_i, $i = 1,2,\ldots,p$. That is,

$$Y_i = \lambda_i\theta + \epsilon_i, \quad i = 1,2,\cdots,p \tag{6}$$

where λ_i is the loading of y_i on θ.

The preceding concept can easily be extended to a factor model that contains m factors. Suppose that a random sample $\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_n$ from a homogeneous population with mean vector, $\boldsymbol{\mu}$ and covariance matrix, $\boldsymbol{\Sigma}$. For this sample, the m-factor model is given by

$$Y_i = \lambda_{i1}\theta_1 + \lambda_{i2}\theta_2 + \cdots + \lambda_{il}\theta_l + \cdots + \lambda_{im}\theta_m + \varepsilon_i; \quad i = 1,2,\cdots,p \tag{7}$$

where $\theta_1,\theta_2,\cdots,\theta_m$ are the common factors (latent abilities in IRT); the coefficients λ_{il} are the loadings; and the error terms $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_m$ are unique factors. The loadings λ_{il} can be used in the interpretation of the factors. For instance, θ_m may be interpreted by examining its loadings $\lambda_{1m},\lambda_{2m},\cdots,\lambda_{pm}$ and noting the x's that have large loadings on θ_m.

In matrix notation, Equation (7) can be written as

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{8}$$

where $\mathbf{y} = \left(Y_1,Y_2,\cdots,Y_p\right)'$, $\boldsymbol{\theta} = (\theta_1,\theta_2,\cdots,\theta_m)'$, $\boldsymbol{\varepsilon} = \left(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_p\right)'$, and

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \cdots & \lambda_{pm} \end{pmatrix} \tag{9}$$

Factor analysis model is effectively utilised when its assumptions are satisfied. The assumptions are

$$\mathrm{E}(\mathbf{y}) = \mathbf{0}_{(p\times 1)}, \quad \mathrm{E}(\boldsymbol{\theta}) = \mathbf{0}_{(m\times 1)}, \quad \mathrm{var}(\boldsymbol{\theta}) = \mathbf{I}_{(m\times m)}, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}_{(p\times 1)}$$

$$\mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\psi}_{(p\times p)} = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}, \text{and} \quad \mathrm{cov}(\boldsymbol{\varepsilon},\boldsymbol{\theta}) = \mathbf{0}_{(p\times m)} \tag{10}$$

The assumptions show that the means of the common factor and unique factors are zero. The assumptions for ε_i are similar to those of θ_l except that each ε_i is allowed to have different variance, ψ_i. Again the unique factors are uncorrelated among themselves or with the common factor.

Based on the assumptions, the variance-covariance matrix of the observed variables, $\mathbf{\Sigma}$, can be expressed in terms of the factor loadings and the unique factors. From Equation (8),

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \boldsymbol{\psi} \tag{11}$$

Thus, Equation (11) represents a simplified structure for $\mathbf{\Sigma}$, in which the covariances are modelled by the λ_{il}s alone since $\boldsymbol{\psi}$ is diagonal. From Equation (11),

$$\text{var}(Y_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{ip}^2 + \psi_i \tag{12}$$

That is, the variance of Y_i is partitioned into a part that is due to the common factors (i.e. the communality) given by

$$\text{communality} = h_i^2 = \sum_{l=1}^{m} \lambda_{il}^2,$$

and a part unique to Y_i, referred to as specific variance, ψ_i (see notable text e.g., [12,25]). Again, Equation (11) shows that

$$\text{cov}(Y_i, Y_k) = \sum_{l=1}^{m} \lambda_{il}\lambda_{kl}.$$

The covariances of the Y's with the θ's can also be found in terms of the λ's, and given as

$$\text{cov}(\mathbf{y}, \boldsymbol{\theta}) = \mathbf{\Lambda} \tag{13}$$

If standardised variables are used, Equation (11) is replaced by a model for the correlation matrix, $\mathbf{R}$. Thus,

$$\text{cov}(Y_i, \theta_l) = \lambda_{il} \tag{14}$$

is the $(i-l)$ element of $\mathbf{\Lambda}$ and represents the correlations of the variables with the factors.

2.3 Item-factor relations

Item response models and factor analysis techniques have widely been applied in analysing questionnaire survey data, which are mainly item responses. In what follows, we present the relationship between the parameters of factor analysis and item response models [29] under various conditions such as item response format (dichotomous and polytomous) and dimensionality of the underlying factor/ability (unidimensional and multidimensional).

When performing factor analysis, it is assumed that both the underlying latent factor θ and the response variables Y_i, $i = 1,2,\ldots,p$ are continuous. Suppose that θ and Y_i possess a joint normal probability distribution. Then their density function, $f(Y,\theta)$, is defined by

$$f(Y,\theta) = \frac{1}{2\pi\sigma_Y\sigma_\theta\sqrt{1-\rho_{Y,\theta}^2}} \exp\left[-\frac{1}{2\left(1-\rho_{Y,\theta}^2\right)}\left\{\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho_{Y,\theta}\left(\frac{y-\mu_Y}{\sigma_Y}\right)\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right) + \left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2\right\}\right], \tag{15}$$

where μ_Y and σ_Y are, respectively, the mean and standard deviation of Y_i, μ_θ and σ_θ are, respectively, the mean and standard deviation of θ, and $\rho_{Y,\theta}^2$ measures the correlation between Y_i and θ.

The distribution of θ is assumed to be normal and defined as

$$g(\theta) = \frac{1}{\sigma_\theta \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta - \mu_\theta}{\sigma_\theta}\right)^2\right] \tag{16}$$

The conditional distribution of Y_i given θ, $f(Y|\theta) = \frac{f(Y,\theta)}{g(\theta)}$, is given by

$$f(Y|\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_Y^2(1-\rho_{Y,\theta}^2)}} \exp\left[-\frac{1}{2\left(1-\rho_{Y,\theta}^2\right)}\left\{\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho_{Y,\theta}\left(\frac{y-\mu_Y}{\sigma_Y}\right)\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right) + \left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2\right\} + \frac{1}{2}\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2\right]$$

A little simplification gives

$$f(Y|\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_Y^2(1-\rho_{Y,\theta}^2)}} \exp\left[-\frac{1}{2\sigma_Y^2\left(1-\rho_{Y,\theta}^2\right)}\left\{y - (\mu_Y + \rho_{Y,\theta}\frac{\sigma_Y}{\sigma_\theta}(\theta-\mu_\theta))\right\}^2\right] \tag{17}$$

Equation (17) is a density function of a normal random variable with mean

$$\mu_Y + \rho_{Y,\theta}\frac{\sigma_Y}{\sigma_\theta}\left(\theta - \mu_\theta\right) \tag{18}$$

and variance

$$\sigma_Y^2\left(1-\rho_{Y,\theta}^2\right) \tag{19}$$

Therefore, the conditional distribution of Y_i given θ is normally distributed. That is,

$$Y|\theta \sim N\left[\mu_Y + \rho_{Y,\theta}\frac{\sigma_Y}{\sigma_\theta}\left(\theta-\mu_\theta\right), \quad \sigma_Y^2\left(1-\rho_{Y,\theta}^2\right)\right]. \tag{20}$$

For the one-factor model (Equation (6)), we assume that

$$\theta \sim N(0,1) \quad \text{and} \quad \epsilon_i \sim N\left(0,\psi_i\right). \tag{21}$$

For a given indicator variable i,

$$Y_i \sim N\left(0, \lambda_i^2 + \psi_i\right) \tag{22}$$

From the assumptions and using Equations (20)-(22), we obtain

$$Y|\theta \sim N\left(\lambda_i\theta, 1-\lambda_i^2\right) \tag{23}$$

In factor analysis the response variable, Y is assumed to be continuous and normally distributed. However, responses to close-ended items in questionnaires are categorical and, for that matter, result in categorical data. Many researchers have described the relationship between item responses to be non-linear and declared the standard factor models in Equations (6) and (7) as inappropriate [2, 9]. In order to apply factor analysis model to item response data, it is assumed that the continuous response variable, Y is discretised to yield the categorical response variable, X. This means that the continuous response variable, Y underlies each categorical response variable, X_i. Specifically, for binary items, each response score X (0 and 1) is considered to arise from an arbitrary dichotomisation of the continuous underlying response variable Y. Figure 1 illustrates the relationship between observed item response X and underlying response variable Y [9, 16].

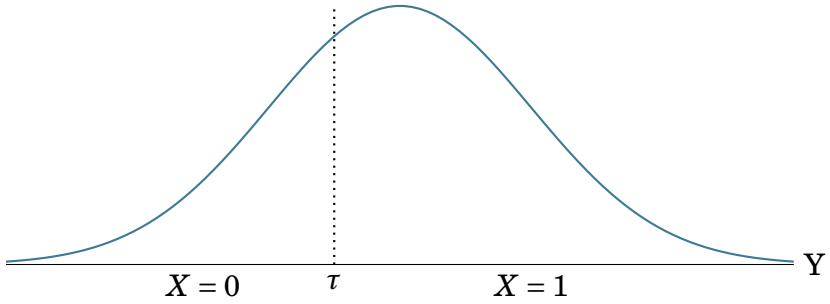

Figure 1. Pictorial representation of a Normal response variable underlying the observed dichotomous response variable

Figure 1 indicates that the relationship between continuous underlying response variable, Y and the dichotomous observed response variable, X is defined by

$$X_i = \begin{cases} 1, & \text{if } Y \geq \tau \\ 0, & \text{if } Y < \tau \end{cases} \tag{24}$$

where, τ denotes threshold between the two response categories. The threshold is estimated to link the underlying normal response variable Y to the observed response categorical variable X. This implies that, to obtain a positive response (i.e. Yes, represented by $X_i = 1$), then

$$\begin{aligned} p(X_i = 1|\theta) &= p(Y \geq \tau) \\ &= p\left[\frac{Y - \lambda_i\theta}{\sqrt{1-\lambda_i^2}} \geq \frac{\tau - \lambda_i\theta}{\sqrt{1-\lambda_i^2}}\right] \\ &= p\left[Z \geq \frac{\tau - \lambda_i\theta}{\sqrt{1-\lambda_i^2}}\right] \\ &= \Phi\left[\frac{\lambda_i}{\sqrt{1-\lambda_i^2}}\left(\theta - \frac{\tau}{\lambda_i}\right)\right], \end{aligned} \tag{25}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Now, we consider that the random variable X which is logistically distributed with parameters μ and $\sigma\,(>0)$, with a density function defined [1] by

$$f(x;\mu,\sigma) = \frac{\pi}{\sigma\sqrt{3}} \cdot \frac{\exp\left\{-\frac{\pi}{\sigma\sqrt{3}}(x-\mu)\right\}}{\left[1+\exp\left\{-\frac{\pi}{\sigma\sqrt{3}}(x-\mu)\right\}\right]^2} \tag{26}$$

By letting $Z = \frac{1}{\sigma}(x-\mu)$, which has a standard Normal distribution with mean 0 and variance 1, Equation (26) becomes

$$f(z;0,1) = 1.702 \cdot \frac{\exp(-1.702z)}{[1+\exp(-1.702z)]^2}, \tag{27}$$

which is the standard logistic distribution function. The cumulative distribution function of Z, $F(z;0,1)$, is given by

$$F(z;0,1) = \int_{-\infty}^{z} f(t;0,1)\,\mathrm{d}t$$

That is,

$$F(z;0,1) = 1.702\int_{-\infty}^{z} \frac{\exp(-1.702t)}{[1+\exp(-1.702t)]^2}\,\mathrm{d}t \tag{28}$$

Let $u = 1+\exp(-1.702t)$, differentiating u with respect to t gives

$$\mathrm{d}u = -1.702\exp(-1.702t)\,\mathrm{d}t \quad \text{or} \quad \mathrm{d}t = \frac{-1}{1.702(u-1)}\,\mathrm{d}u$$

Substituting $\mathrm{d}t$ into Equation (28) yields

$$\begin{aligned} F(z;0,1) &= 1.702\int_{\infty}^{1+\exp(-1.702z)} \frac{(u-1)}{u^2}\cdot\frac{-1}{1.702(u-1)}\,\mathrm{d}u \\ &= \int_{\infty}^{1+\exp(-1.702z)} -\frac{1}{u^2}\,\mathrm{d}u \\ &= \left[\frac{1}{u}\right]_{\infty}^{1+\exp(-1.702z)} \\ &= \frac{1}{1+\exp(-1.702z)}, \quad z\in\Re \end{aligned} \tag{29}$$

which is the cumulative distribution function of the logistic distribution.

The unidimensional 2PL model

$$p\left(X_{ij}=1|\theta,\alpha,\delta\right) = \frac{1}{1+\exp[-1.702\alpha_i(\theta-\delta_i)]}$$

has the form of the logistic cumulative distribution function in Equation (29) evaluated at $\alpha_i(\theta-\delta_i)$. Thus, for the jth individual,

$$p(X_{ij}=1|\theta) = \Phi[\alpha_i(\theta-\delta_i)] \tag{30}$$

Therefore, appropriately equating the probabilities in Equation (25) and Equation (30) yields

$$\alpha_i = \frac{\lambda_i}{\sqrt{1-\lambda_i^2}}, \quad |\lambda_i|<1 \tag{31}$$

and

$$\delta_i = \frac{\tau}{\lambda_i} \tag{32}$$

Equations (31) and (32) show the relationship between IRT and factor analysis for dichotomous data [29]. Equation (31) indicates that α_i is directly a function of λ_i. This means that, an item that greatly discriminates between individuals at lower and higher ability levels will be highly influential in the formation of the corresponding factor. However, if the item has poor discriminatory power then, it will not contribute significantly to the formation of the factor. Equation (32) shows that an item's difficulty δ_i is a function of its category threshold value (τ) and λ_i. In this case, there is no clear relationship between the difficulty parameter and that of the factor model.

Sometimes the responses to a set of items in a questionnaire is not characterised by only one ability, but a combination of several abilities of the respondent. To this end, we determine the relationship between the parameters of multidimensional item response and m-factor models.

Considering the m-factor model (see Equation (8)), each X_i can be written as

$$X_i = \boldsymbol{\lambda}_i'\boldsymbol{\theta} + \epsilon_i \tag{33}$$

where, $\boldsymbol{\lambda_i} = (\lambda_{i1}, \lambda_{i2}, \cdots, \lambda_{im})'$. The distribution of X_i (in Equation (22)) becomes

$$X_i \sim N\left(0, \sum_{j=1}^{m} \lambda_{ij}^2 + \psi_i\right) \tag{34}$$

Also, the conditional distribution of X_i given $\boldsymbol{\theta}$ (see Equation (23)) is given by

$$X_i|\boldsymbol{\theta} \sim N\left(\boldsymbol{\lambda}_i'\boldsymbol{\theta}, 1 - \sum_{j=1}^{m} \lambda_{ij}^2\right) \tag{35}$$

Following similar algebraic steps required in Equation (25), we determine that

$$p(X_i = 1|\theta) = \Phi\left(\frac{\boldsymbol{\lambda}_i'\boldsymbol{\theta} - \tau}{\sqrt{1 - \sum_{j=1}^{m} \lambda_{ij}^2}}\right) = \Phi\left(\frac{\boldsymbol{\lambda}_i}{\sqrt{1 - \sum_{j=1}^{m} \lambda_{ij}^2}}\boldsymbol{\theta} - \frac{\tau}{\sqrt{1 - \sum_{j=1}^{m} \lambda_{ij}^2}}\right). \tag{36}$$

Using the M2PL model and following Equation (30), it follows that

$$p(X_{ij} = 1|\boldsymbol{\theta}) = \Phi\left(\boldsymbol{\alpha}_i'\boldsymbol{\theta} + d_i\right) \tag{37}$$

Comparing Equation (36) and Equation (37) gives

$$\boldsymbol{\alpha}_i = \frac{\boldsymbol{\lambda}_i}{\sqrt{1 - \sum_{j=1}^{m} \lambda_{ij}^2}} \tag{38}$$

and

$$d_i = \frac{-\tau}{\sqrt{1 - \sum_{j=1}^{m} \lambda_{ij}^2}} \tag{39}$$

In the case of ordered polytomous data, the observed item response variable, X arises as a result of a categorisation of an underlying continuous response variable, Y by means of a series of thresholds, $\tau_h, h = 1, 2, \ldots, g, g+1, \ldots, k$. Schematically, the relationship between an underlying normal response variable and the observed categorical response variable is illustrated in Figure (2) [9, 16].

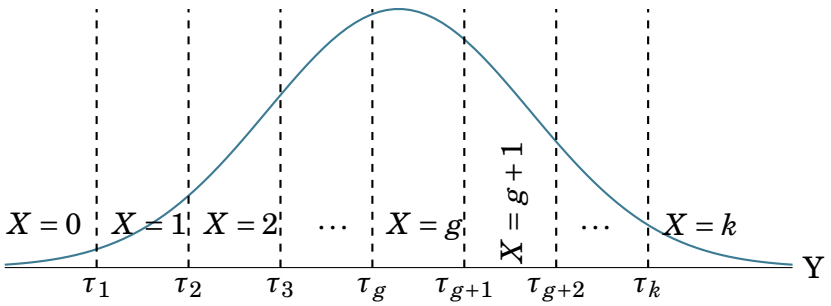

Figure 2. Pictorial representation of a Normal response variable underlying the observed polytomous response variable

Figure 2 implies that, the relationship between continuous variable, Y and the polytomous response variable, X is defined by

$$X_i = \begin{cases} g, & \text{if } \tau_g \le Y_i < \tau_{g+1} \\ 0, & \text{otherwise} \end{cases} \tag{40}$$

It follows that $\tau_0 = -\infty$ and $\tau_{k+1} = \infty$. To obtain a response in category g of the item, then

$$\begin{aligned} p(X_i = g|\theta) &= p(\tau_g \le Y_i < \tau_{g+1}) \\ &= \Phi\left(\frac{\tau_{g+1} - \lambda_i\theta_j}{\sqrt{1-\lambda_i^2}}\right) - \Phi\left(\frac{\tau_g - \lambda_i\theta_j}{\sqrt{1-\lambda_i^2}}\right) \\ &= \Phi\left[\frac{-\lambda_i}{\sqrt{1-\lambda_i^2}}\left(\theta_j - \frac{\tau_{g+1}}{\lambda_i}\right)\right] - \Phi\left[\frac{-\lambda_i}{\sqrt{1-\lambda_i^2}}\left(\theta_j - \frac{\tau_g}{\lambda_i}\right)\right] \\ &= \Phi\left[\frac{\lambda_i}{\sqrt{1-\lambda_i^2}}\left(\theta_j - \frac{\tau_g}{\lambda_i}\right)\right] - \Phi\left[\frac{\lambda_i}{\sqrt{1-\lambda_i^2}}\left(\theta_j - \frac{\tau_{g+1}}{\lambda_i}\right)\right]. \end{aligned} \tag{41}$$

In order to link factor analysis and item response theory models for polytomous data, the graded response (GR) model is considered. The form of the model makes it tractable. The GR model is stated as

$$p(X_i = g|\theta) = P(X_i \ge g|\theta) - P(X_i \ge g+1|\theta).$$

Thus,

$$\begin{aligned} p(X_i = g\,|\,\theta) &= \frac{1}{1+\exp\left[-\alpha_i(\theta - \delta_{ig})\right]} - \frac{1}{1+\exp\left[-\alpha_i(\theta - \delta_{i,g+1})\right]} \\ &= \Phi[\alpha_i(\theta - \delta_{ig})] - \Phi[\alpha_i(\theta - \delta_{i,g+1})] \end{aligned} \tag{42}$$

By comparing Equations (41) and (42), we observe that

$$\alpha_i = \frac{\lambda_i}{\sqrt{1-\lambda_i^2}}, \quad \delta_{ig} = \frac{\tau_{ig}}{\lambda_i}, \quad \text{and} \quad \delta_{i,g+1} = \frac{\tau_{i,g+1}}{\lambda_i}. \tag{43}$$

These establish the relationship among the parameters of factor analysis and IRT models for ordered polytomous items [7].

In many practical situations, responses to polytomous items are characterised by more than one dimension of person ability. Thus, an equivalent of Equation (41) for m-factor model is given by

$$p(X_i = g\,|\,\boldsymbol{\theta}) = \Phi\left(\frac{\boldsymbol{\lambda}_i'}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}}\boldsymbol{\theta} - \frac{\tau_g}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}}\right) - \Phi\left(\frac{\boldsymbol{\lambda}_i'}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}}\boldsymbol{\theta} - \frac{\tau_{g+1}}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}}\right). \tag{44}$$

This can be likened to a multidimensional IRT model, say MGR model, given by

$$p(X_i = g\,|\,\boldsymbol{\theta}_j) = \frac{1}{1+\exp\left[-(\boldsymbol{\alpha}_i'\boldsymbol{\theta} + d_{ig})\right]} - \frac{1}{1+\exp\left[-(\boldsymbol{\alpha}_i'\boldsymbol{\theta} + d_{i,g+1})\right]}.$$

That is,

$$p(X_i = g\,|\,\boldsymbol{\theta}) = \Phi[\boldsymbol{\alpha}_i'\boldsymbol{\theta} + d_{ig}] - \Phi[\boldsymbol{\alpha}_i'\boldsymbol{\theta} + d_{i,g+1}] \tag{45}$$

Equations (44) and (45) show that

$$\boldsymbol{\alpha}_i = \frac{\boldsymbol{\lambda}_i'}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}}, \quad d_{ig} = \frac{-\tau_g}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}} \quad \text{and} \quad d_{i,g+1} = \frac{-\tau_{g+1}}{\sqrt{1-\sum_{l=1}^{m}\lambda_{il}^2}}. \tag{46}$$

3. Data Simulation

In order to examine the effect of response scale on results of factor analysis, data is simulated using R package *mirt* [3]. Several datasets are generated under different conditions for a total of 20 items. Datasets are generated for varied response scales, namely dichotomous, three-point, five-point, and seven-point scales. For each response scale, different sample sizes are considered. These sample sizes include 30, 100, 150, 200, 500, 800, and 1000. The purpose is to investigate the effect of sample size on factor analysis results. In addition, for each response scale, different dimensions of underlying person-ability are considered, particularly unidimensional, two-dimensional and three-dimensional. We then determine how these underlying dimensions translate into factors. The datasets are generated using the command: $simdata(a, d, N, itemtype)$ [3], where argument a denotes a vector/matrix of discrimination parameter values, d vector/matrix of difficulty parameter values, N sample size and *itemtype* the underlying IRT model. These arguments are specified to generate the desired dataset.

Firstly, unidimensional dichotomous response dataset is simulated using 2PL model. The 2PL model is considered, because it assumes that items have different discrimination powers. In questionnaires, items differ in terms of content, and so are their discriminations. In this system, a 20×1 vector of discrimination values are specified for a, and another 20×1 vector of difficulty values for d for all sample sizes. Table 1 shows the discrimination and difficulty parameter values used in simulating unidimensional item response data for twenty items.

In the table, we have $0.4 < \alpha < 3.0$ and $-2.5 < \delta < 2.5$. High values of α means that the item is discriminating largely between low-ability and high-ability persons. High positive value of δ means the item is "difficult" and only high-ability persons can respond to it in higher response categories. Conversely, an item with high negative δ value is considered to be "easy" and persons with high ability levels tend to respond favourably to it. A δ value of zero or close to zero means that the item is averagely difficult and persons of average ability could respond to it in higher response categories.

To generate a two-dimensional dichotomous dataset, a argument is modified to a 20×2 matrix of discrimination values. Here, the same vector of discrimination values on the first dimension is repeated on the second dimension. Thus, the two dimensions have the same discrimination values. The intent is to determine how the information will manifest in factor model and facilitate easy identification of the factor. In this case, one of two possibilities is expected in the factor model. On one hand, a factor solution is expected to be dominated by two repeating factors since the same information is contained on the two dimensions that underlie

the data. On the other hand, a single dominant factor is expected with the other influenced by few items or none. To simulate a three-dimensional dichotomous dataset, 20×3 matrix of discrimination values is specified for a. The three dimensions have the same discrimination values, with similar rationale as for the two-dimensional case.

Table 1. Discrimination and Difficulty Levels for Each Item

Item	Discrimination (α)	Difficulty (δ)
1	0.5	0.00
2	0.7	0.12
3	0.8	-2.30
4	0.6	0.10
5	0.4	2.00
6	2.2	-2.50
7	1.5	-2.00
8	2.7	-1.50
9	1.8	-2.20
10	1.6	2.50
11	2.0	2.30
12	2.9	1.50
13	3.0	2.20
14	2.1	0.30
15	2.8	0.50
16	1.4	0.25
17	1.9	0.40
18	1.2	0.42
19	1.3	0.56
20	2.9	0.20

Next, a unidimensional three-point scale data is generated using a polytomous IRT model, specifically GPC model. Polytomous models incorporates category boundaries to cater for the multi-category nature of items. For a given scale, the number of response categories is specified using argument d. For three-point scale, d consists of a 20×2 matrix of difficulty values with 20×1 discrimination values depicting unidimensional fashion. The GPC model considers items of varying discriminations just as 2PL model.

Higher response scale datasets, five and seven-point, are also simulated in the same manner as three-point scale.

4. Analysis and Results

The simulated datasets are analysed using standard R 3.4.3 codes [20]. The IRT analyses are conducted using R package *mirt* [3]. The analyses of unidimensional dichotomous item response datasets are based on two-parameter logistic (2PL) IRT model, whereas multidimensional two-parameter logistic (M2PL) IRT model is employed in the analyses of multidimensional dichotomous datasets. For unidimensional polytomous item response datasets, we employ the generalised partial credit (GPC) model in the analyses. In terms of multidimensional polytomous datasets, the multidimensional generalised partial credit (MGPC) model is used to conduct the analyses.

Factor analysis is also performed on each simulated item response dataset using R package *psych* [26]. Factor analyses of dichotomous item response datasets are based on tetrachoric correlation matrices. On the other hand, polychoric correlation matrix is used as input in factor analysis of polytomous item response datasets.

We begin the analysis with an exploration of characteristics of items in the unidimensional dichotomous item response dataset. The estimates of discrimination and difficulty parameters of unidimensional 2PL model are presented in Table 2.

The results show that the estimated values of discrimination ($\hat{\alpha}$) and difficulty ($\hat{\delta}$) parameters generally fluctuates with increasing sample size. That is, the estimates of discrimination and difficulty parameters change depending on the sample. There is a marked difference between the specified and estimated item parameter values at lower samples ($n = 30$ and 100). However, the differences tend to reduce at sample sizes of 150 and beyond. In addition, the differences become negligible at larger samples ($n = 500, 800, 1000$). For example, from Table 1 the specified discrimination value (α) of Item 10 is 1.6, which is quite close to the estimated values $\hat{\alpha}$ at samples of sizes 150 through 1000.

Table 2. Discrimination and difficulty estimates of unidimensional 2PL model for various sample sizes

	Sample Size							
	30		100		150		200	
Item	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$
1	0.684	-0.153	0.567	0.719	0.744	0.331	0.490	0.148
2	1.287	0.147	1.539	-0.001	0.845	0.184	0.828	0.045
3	0.679	-1.514	0.267	-3.273	0.450	-1.944	0.344	-1.857
4	0.321	0.274	1.008	-0.005	0.696	0.059	0.654	0.131
5	-0.377	1.655	0.076	1.588	-0.044	2.537	0.492	2.237
6	0.501	-1.458	1.718	-1.715	1.737	-2.281	2.146	-2.204
7	0.432	-2.269	0.367	-1.491	2.250	-2.283	0.927	-1.723
8	1.866	-0.880	2.266	-0.743	3.201	-1.353	2.346	-0.927
9	1.298	-1.325	1.852	-2.412	2.175	-2.173	2.296	-2.504
10	0.967	2.187	9.008	12.668	1.522	2.352	1.118	1.942
11	4.018	3.418	2.143	2.214	2.092	2.015	2.084	2.271
12	3.056	2.370	2.537	1.760	3.153	1.735	3.964	1.750
13	8.339	6.370	5.092	3.717	3.521	2.889	2.694	1.687
14	1.469	0.522	2.730	0.522	2.221	0.157	1.963	0.506
15	4.389	1.254	3.263	1.062	3.648	0.374	3.289	0.411
16	0.863	0.622	0.732	-0.002	1.339	0.100	1.524	0.420
17	1.697	0.346	1.258	0.467	1.807	0.437	2.523	0.659
18	1.316	0.147	0.778	0.363	1.287	-0.041	1.091	0.346
19	1.874	-0.056	1.606	0.759	1.353	0.536	0.921	0.525
20	4.551	-0.557	3.292	0.402	2.440	0.406	4.249	0.041

Table 2 continued

	Sample Size					
	500		800		1000	
Item	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$
1	0.460	0.034	0.504	0.090	0.453	-0.029
2	0.710	0.187	0.081	-0.001	0.784	0.100
3	0.814	-2.227	0.895	-2.210	0.700	-2.396
4	0.690	-0.160	0.638	0.132	0.656	0.159
5	0.558	1.976	0.274	1.985	0.443	2.056
6	2.276	-2.385	2.196	-2.353	2.654	-2.520
7	1.641	-2.120	1.723	-1.864	1.886	-2.248
8	2.679	-1.435	2.835	-1.370	2.720	-1.362
9	2.275	-2.554	2.486	-2.866	1.739	-2.000
10	1.786	2.529	1.599	2.551	1.663	2.468
11	1.594	2.158	2.050	2.437	2.271	2.595
12	3.385	1.874	2.898	1.678	2.971	1.805
13	3.268	2.344	3.496	2.880	3.108	2.420
14	2.069	0.458	2.272	0.410	2.120	0.378
15	3.723	0.340	2.438	0.649	2.865	0.771
16	1.331	0.365	1.480	0.430	1.456	0.352
17	1.905	0.347	1.930	0.453	1.765	0.410
18	1.069	0.297	1.053	0.421	1.371	0.724
19	1.333	0.720	1.165	0.641	1.127	0.720
20	2.889	0.061	2.920	0.198	2.900	0.377

Table 3. P-values of Fitness of Items to Unidimensional 2PL Model for Various Sample Sizes

	Sample Size						
Item	30	100	150	200	500	800	1000
1	0.377	0.251	0.666	0.446	0.202	0.507	0.178
2	0.360	0.694	0.459	0.314	0.988	0.564	0.368
3	0.344	0.155	0.113	0.274	0.141	0.530	0.768
4	0.792	0.553	0.260	0.055	0.890	0.665	0.609
5	0.530	0.851	0.979	0.668	0.681	0.559	0.901
6	0.325	0.304	0.001	0.465	0.905	0.849	0.818
7	NaN	0.317	0.608	0.377	0.476	0.165	0.019
8	0.767	0.931	0.595	0.939	0.483	0.177	0.997
9	0.592	0.535	0.382	0.648	0.708	0.467	0.353
10	NaN	NaN	0.433	0.296	0.269	0.893	0.188
11	NA	0.877	0.682	0.486	0.490	0.318	0.079
12	NaN	0.234	0.008	0.033	0.783	0.429	0.182
13	NA	0.188	0.856	0.530	0.605	0.275	0.758
14	0.310	0.294	0.973	0.437	0.890	0.657	0.414
15	0.457	0.750	0.812	0.210	0.317	0.906	0.639
16	0.131	0.136	0.729	0.240	0.282	0.363	0.457
17	0.303	0.387	0.256	0.961	0.856	0.560	0.559
18	0.228	0.253	0.071	0.989	0.677	0.558	0.428
19	0.550	0.433	0.675	0.217	0.947	0.768	0.424
20	0.669	0.562	0.177	0.907	0.314	0.225	0.421
Model Fit	0.114	0.966	0.514	0.381	0.363	0.387	0.938

The significances of the fitness of items as well as the overall fitness of 2PL model to unidimensional dichotomous response dataset are in Table 3.

Table 3 shows that items fit the unidimensional 2PL model since the p-values are generally much higher than 0.05. Only at $n = 150$, 200 and 1000, it is detected that three items (6, 12, and 7, respectively) do not fit the model. The 2PL model significantly fits the unidimensional dichotomous item response data for all sample sizes.

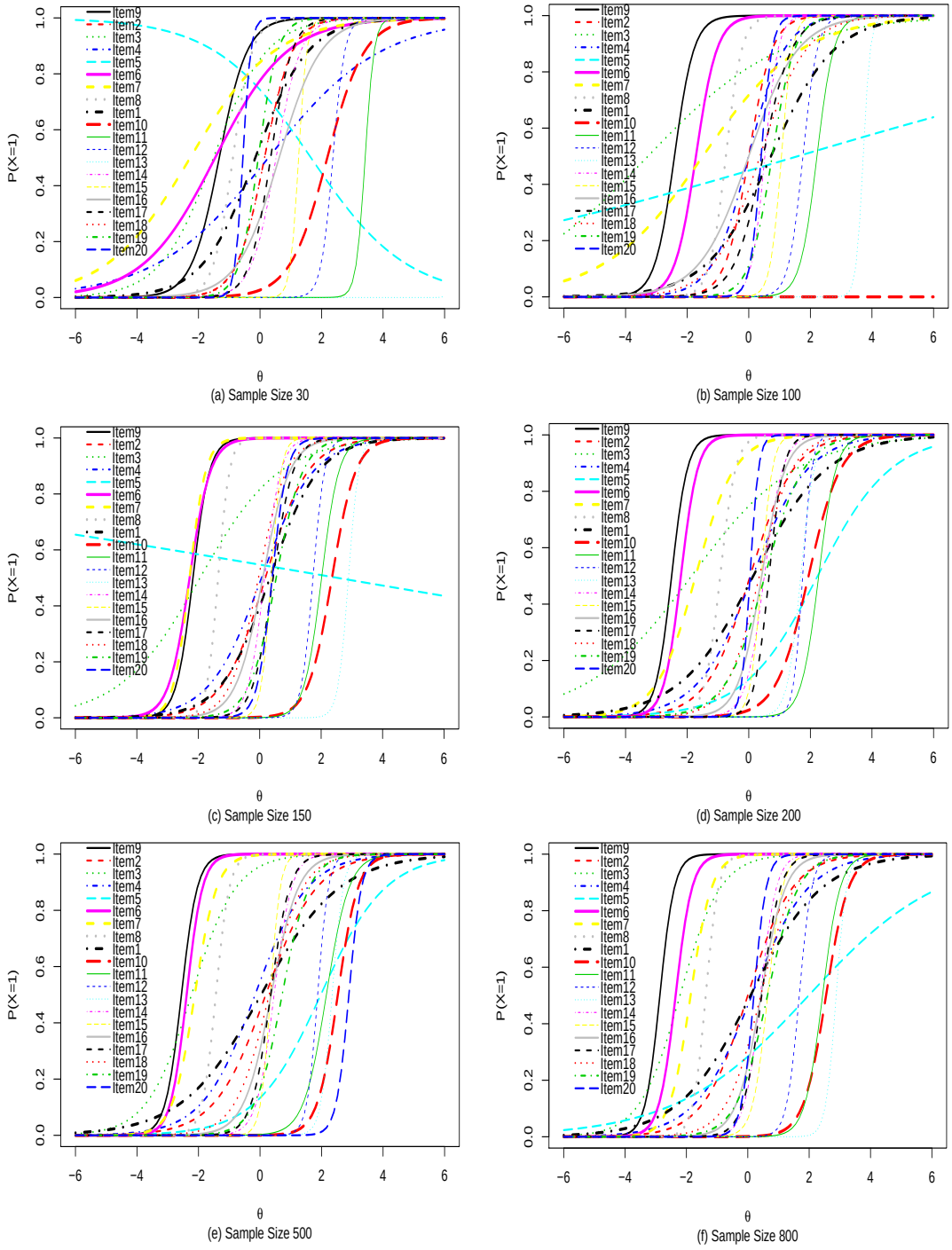

Figure 3. Item characteristic curves of unidimensional 2PL model for various sample sizes

Corresponding graphs of the items on unidimensional dichotomous response data is illustrated in Figure 3. By the graphs, we can assess whether or not items generally exhibit the desired or expected features in line with the item response model.

From Figure 3, we observe that the curve for Item 9 is extremely steep at a low ability level (at -2). This means that for individuals with extremely low ability, Item 9 sharply discriminates. The only item with similar nature is Item 6. However, Item 6 discriminates at a little higher ability. Items 2, 14, 16 through 20 discriminate maximally at average ability levels. Meanwhile, Item 15 discriminates among individuals at high ability levels for small samples but, tend to discriminate at average levels for large sample sizes. Item 10 discriminates the most at extremely high ability levels for $n = 30$, but conforms to discriminating at just high ability levels like other items for $n \geq 150$.

The rest of the results of IRT models are summarised and presented in the following segments.

4.1 IRT results across different scales on various dimensions

We assess IRT results under various conditions such as the number of points on response scales, number of dimensions, and sample size. Table 4 illustrates summary statistics for IRT results across different response scales with different number of dimensions. We examine the table for general features and those specific to the dimensionality.

Table 4 indicates that on unidimensionality across various scales, overall fitness of item response model deteriorates with increasing scale points for small samples, particularly $n = 30$ and 100. This suggests that on unidimensional higher response scales, samples of sizes 100 and below would not produce reliable results. Meanwhile, too large a sample, particularly $n = 1000$ may produce almost the same IRT results since the performance of the model does not change for all response scales. In some instances on polytomous response scales with one dimension, overall fitness of the model increases with increasing points on the scale.

Again, we observe from Table 4 that, in respect of two-dimensional response scales, the number of fit items is almost the same at sample sizes of 150 and beyond. For some of these samples, the overall fitness of the model increases with increasing points on the scale. For polytomous response scales with three dimensions, model fitness seems to be quite high at larger samples, particularly for three and five-point scales. However, there does not appear to be any relationship between the model fitness and the number of fit items.

We now compare the summary statistics of results of factor solutions under different dimensions of the underlying latent ability. For each dimension, a comparison of factor solutions is done at various response scales and sample sizes.

Table 4. Summary statistics for IRT results across different scales on various dimensions

Scale			Sample Size						
			30	100	150	200	500	800	1000
Two-Point	Unidim.	Fit Items	15	19	18	19	20	20	19
		Model Fit	0.114	0.966	0.514	0.381	0.363	0.387	0.938
	Two-dim.	Fit Items	4	18	19	19	20	20	18
		Model Fit	-	0.920	0.406	0.249	0.953	0.974	0.123
	Three-dim.	Fit Items	1	15	17	18	19	19	17
		Model Fit	-	0.882	0.462	0.885	0.556	0.783	0.371
Three-Point	Unidim.	Fit Items	15	20	19	19	20	18	19
		Model Fit	0.002	0.477	0.198	0.589	0.581	0.603	0.911
	Two-dim.	Fit Items	6	17	19	18	18	20	19
		Model Fit	0.006	0.469	0.969	0.482	0.872	0.924	0.779
	Three-dim.	Fit Items	1	14	15	20	16	18	18
		Model Fit	0.830	0.623	0.717	1.000	0.948	0.998	0.766
Five-Point	Unidim.	Fit Items	6	16	18	16	20	19	19
		Model Fit	0.007	0.253	0.368	0.809	0.435	0.150	0.990
	Two-dim.	Fit Items	3	17	18	17	19	18	18
		Model Fit	0.550	0.896	0.956	0.579	0.326	0.969	0.864
	Three-dim.	Fit Items	0	11	16	16	16	15	0
		Model Fit	0.728	0.977	0.997	1.000	0.887	0.997	1.000
Seven-Point	Unidim.	Fit Items	1	18	16	18	20	18	19
		Model Fit	0.000	0.038	0.703	0.144	0.533	0.476	0.963
	Two-dim.	Fit Items	0	10	14	19	19	17	19
		Model Fit	0.842	0.854	0.979	0.930	0.252	0.994	0.667
	Three-dim.	Fit Items	0	3	12	13	15	17	0
		Model Fit	0.834	0.594	0.437	0.489	0.981	0.997	1.000

4.2 One-factor solutions on unidimensional datasets for various scales

Table 5 presents summary statistics for one-factor solutions on unidimensional datasets at various response scales and sample sizes.

Table 5. Summary Statistics for One-Factor Solutions on Unidimensional Datasets

Scale	Measures	Sample Size						
		30	100	150	200	500	800	1000
Two-Point	No. of Ind.	11	14	15	13	15	15	15
	Cum. Var	0.339	0.424	0.450	0.429	0.455	0.461	0.446
	Fit	0.706	0.870	0.900	0.880	0.916	0.922	0.924
Three-Point	No. of Ind.	16	18	17	17	17	17	18
	Cum. Var	0.507	0.558	0.577	0.596	0.588	0.589	0.597
	Fit	0.900	0.953	0.962	0.968	0.968	0.969	0.971
Five-Point	No. of Ind.	17	19	20	20	20	19	20
	Cum. Var	0.580	0.678	0.697	0.714	0.717	0.720	0.729
	Fit	0.949	0.982	0.985	0.987	0.988	0.988	0.990
Seven-Point	No. of Ind.	18	19	20	20	20	20	20
	Cum. Var	0.616	0.736	0.759	0.776	0.784	0.787	0.794
	Fit	0.706	0.968	0.990	0.992	0.993	0.994	0.995

It is evident in Table 5 that the number of influential indicators on the factor increases as points on the scale increase across all sample sizes. In addition, the dominant number of indicators starts from $n = 150$ in each case. The cumulative variation (Cum. Var) accounted for by the factor peaks at $n = 150$ and fluctuates thereafter for two-point scales. In the others, the amount of cumulative variation is almost the same for $n \geq 150$ within rounding errors. Similarly, the significance of the fit of the model also increases with increasing scale points and sample size. This result is consistent with what is observed in the IRT analysis.

4.3 Two-factor solutions on unidimensional datasets for various scales

The summary statistics for unidimensional two-factor solutions at different scale-points are displayed in Table 6. The table also shows summary statistics for various sample sizes.

Table 6. Summary Statistics for Two-Factor Solutions on Unidimensional Datasets

			Sample Size						
Scale	Measures	Factors	30	100	150	200	500	800	1000
Two-Point	No. of Ind.	PA1	10	10	15	12	13	15	15
		PA2	8	8	1	6	6	0	4
	Cum. Var		0.339	0.424	0.450	0.429	0.455	0.461	0.446
	Fit		0.706	0.870	0.900	0.880	0.916	0.922	0.924
Three-Point	No. of Ind.	PA1	14	14	17	13	13	15	15
		PA2	5	9	1	13	14	4	13
	Cum. Var		0.596	0.599	0.620	0.618	0.605	0.603	0.606
	Fit		0.937	0.962	0.968	0.971	0.971	0.972	0.973
Five-Point	No. of Ind.	PA1	14	16	18	16	19	18	19
		PA2	9	14	6	16	1	10	2
	Cum. Var		0.645	0.704	0.732	0.739	0.739	0.731	0.740
	Fit		0.966	0.985	0.987	0.989	0.990	0.990	0.991
Seven-Point	No. of Ind.	PA1	14	19	16	17	19	18	19
		PA2	14	1	18	18	12	18	18
	Cum. Var		0.662	0.764	0.774	0.787	0.793	0.796	0.880
	Fit		0.976	0.993	0.993	0.994	0.994	0.995	0.995

It can be noticed that the desired factor structure is observed at $n = 150$ where there is the highest number of influential indicators on the first factor and only a few on the second. However, at higher scale-points (seven), the structure occurs at a small sample size of $n = 100$. It is notable that, in general, at the desired factor structure, the cumulative variation or the fitness peaks and deteriorates thereafter. This is true either for the cumulative variation or for the value of fitness of the model.

4.4 Three-factor solutions on unidimensional datasets for various scales

Table 7 presents the summary statistics for the unidimensional three-factor solutions for various scale-points. We consider the various statistics for sample sizes of 30, 100, 150 and 200. Higher sample sizes are ignored as their results do not show improvement beyond $n = 200$.

Table 7. Summary Statistics for Three-Factor Solutions on Unidimensional Datasets

Scale	Measures	Factors	Sample Size			
			30	100	150	200
Two-Point	No. of Ind.	PA1	6	10	10	10
		PA2	5	8	7	6
		PA3	5	1	0	1
	Cum. Var		0.544	0.530	0.572	0.563
	Fit		0.864	0.916	0.937	0.918
Three-Point	No. of Ind.	PA1	13	14	14	10
		PA2	3	7	4	5
		PA3	2	1	1	4
	Cum. Var		0.669	0.639	0.637	0.641
	Fit		0.956	0.971	0.973	0.975
Five-Point	No. of Ind.	PA1	14	16	18	14
		PA2	8	12	2	12
		PA3	1	1	1	3
	Cum. Var		0.694	0.732	0.737	0.745
	Fit		0.975	0.989	0.989	0.990
Seven-Point	No. of Ind.	PA1	14	17	18	17
		PA2	9	13	2	8
		PA3	1	1	2	1
	Cum. Var		0.696	0.779	0.794	0.802
	Fit		0.981	0.994	0.994	0.995

Generally, we find results for $n = 150$ to be consistent with underlying dimensionality of the data. It gives the first factor as the dominant one and the other two are just much fewer-indicator factor (or none) that contribute marginally to the cumulative proportion of variation explained. Again, at this sample size, cumulative variation (or the fitness) peaks and deteriorates thereafter.

4.5 Two-factor solutions on two-dimensional datasets for various scales

Two-dimensional datasets are generated by specifying the same vector of item discrimination parameter values on both dimensions of the underlying ability. In this system, we expect that a good factor solution should have two repeating factors since the same information is contained on the two underlying dimensions of the dataset. Alternatively, we could expect a single dominant first factor in the two-factor solution with similar reasoning as in the former instance. Here, we compare the results of two-factor solutions on two-dimensional datasets at various sample sizes and scale-points. The results are summarised in Table 8.

Table 8 shows that the cumulative variation and fitness of the model peak at $n = 100$ for all scale points, and deteriorates or fluctuates thereafter. The desired factor structure is thus obtained at $n = 100$. It is also observed that the amount of cumulative variation explained by the model increases with increasing scale-point. The fitness of the model as well as the number of significant indicators are also generally high at higher scale-points.

Table 8. Summary Statistics for Two-Factor Solutions on Two-Dimensional Datasets

Scale	Measures	Factors	Sample Size						
			30	100	150	200	500	800	1000
Two-Point	No. of Ind.	PA1	12	18	14	16	17	15	16
		PA2	4	1	13	8	1	6	1
	Cum. Var		0.522	0.677	0.639	0.651	0.486	0.492	0.478
	Fit		0.873	0.973	0.965	0.971	0.925	0.934	0.930
Three-Point	No. of Ind.	PA1	16	19	15	18	16	17	19
		PA2	13	3	14	2	15	14	1
	Cum. Var		0.714	0.763	0.749	0.742	0.723	0.700	0.747
	Fit		0.974	0.990	0.989	0.989	0.988	0.986	0.991
Five-Point	No. of Ind.	PA1	19	19	19	19	19	19	19
		PA2	1	1	17	15	17	17	5
	Cum. Var		0.781	0.862	0.839	0.832	0.822	0.808	0.840
	Fit		0.990	0.997	0.996	0.996	0.996	0.995	0.997
Seven-Point	No. of Ind.	PA1	19	19	19	19	19	19	19
		PA2	1	1	17	18	18	5	4
	Cum. Var		0.821	0.886	0.875	0.869	0.868	0.866	0.886
	Fit		0.994	0.998	0.998	0.998	0.998	0.998	0.998

4.6 Three-factor solutions on three-dimensional datasets for various scales

In this system, we expect that a plausible three-factor solution to possess either three repeating factors or a single dominant first factor since the same information is contained on three dimensions that underlie item responses. Table 9 displays the summary statistics for three-dimensional three-factor solutions at various samples and scale-points. Since the results do not show improvement for higher sample sizes, the results for $n = 500$, 800 and 1000 are excluded.

Table 9. Summary Statistics for Three-Factor Solutions on Three-Dimensional Datasets

Scale	Measures	Factors	Sample Size			
			30	100	150	200
Two-Point	No. of Ind.	PA1	12	13	17	16
		PA2	10	13	2	4
		PA3	2	3	1	1
	Cum. Var		0.746	0.753	0.789	0.702
	Fit		0.973	0.986	0.990	0.983
Three-Point	No. of Ind.	PA1	17	18	18	18
		PA2	8	12	1	2
		PA3	2	1	1	2
	Cum. Var		0.818	0.835	0.864	0.792
	Fit		0.995	0.995	0.997	0.994
Five-Point	No. of Ind.	PA1	17	18	18	18
		PA2	13	8	2	1
		PA3	1	6	1	0
	Cum. Var		0.880	0.887	0.900	0.880
	Fit		0.998	0.998	0.999	0.998
Seven-Point	No. of Ind.	PA1	18	16	18	18
		PA2	8	17	12	1
		PA3	5	8	5	1
	Cum. Var		0.878	0.907	0.916	0.907
	Fit		0.998	0.999	0.999	0.999

Generally, a sample size of 150 produces a more consistent factor solution based on the underlying dimensionality of the data. At this sample size, cumulative variation and/or model fitness peaks and deteriorates thereafter. The amount of cumulative variation explained increases with increasing scale-points. It follows that the number of influential indicators on factors increases with increasing scale-points, which is particularly true for the first factor. This means that factors are more well defined and could be more interpretable on larger scale-points. The results are, however, almost the same on higher scale-points of five and seven.

4.7 Factor solutions on lower point scale

The results so far shows that both IRT and Factor Analysis, for lower sample size, could be characterised with special features. In particular, $n = 150$ stands out clearly as an optimal size for obtaining the desired factor model. In addition, the results are generally good for higher scale points. In this section, we look at details of the models with respect to the dichotomous scale. We start with one-factor solution of unidimensional two-point item response data at various sample sizes. The results are shown in Table 10.

Table 10. Loadings of one-factor solutions of unidimensional item response dataset for various sample sizes on dichotomous scale

	Sample Size						
Item	30	100	150	200	500	800	1000
1	0.400	0.346	0.399	0.283	0.269	0.307	0.277
2	0.597	0.405	0.464	0.457	0.398	0.454	0.434
3	0.249	0.507	0.211	0.183	0.402	0.424	0.359
4	0.149	0.544	0.405	0.386	0.380	0.373	0.372
5	-0.220	0.000	0.000	0.251	0.275	0.143	0.229
6	0.228	0.639	0.658	0.774	0.769	0.745	0.810
7	0.156	0.211	0.773	0.434	0.660	0.697	0.701
8	0.608	0.755	0.831	0.787	0.828	0.853	0.831
9	0.422	0.640	0.734	0.753	0.784	0.789	0.676
10	0.474	0.796	0.626	0.520	0.700	0.637	0.678
11	0.889	0.774	0.775	0.760	0.661	0.749	0.781
12	0.744	0.865	0.857	0.888	0.874	0.864	0.855
13	0.931	0.894	0.902	0.826	0.878	0.891	0.870
14	0.529	0.839	0.773	0.738	0.765	0.805	0.774
15	0.865	0.865	0.855	0.858	0.894	0.814	0.849
16	0.402	0.429	0.618	0.663	0.626	0.665	0.654
17	0.677	0.627	0.721	0.823	0.736	0.750	0.725
18	0.603	0.430	0.575	0.555	0.540	0.539	0.636
19	0.640	0.685	0.627	0.473	0.615	0.573	0.565
20	0.809	0.861	0.806	0.907	0.848	0.863	0.852
Prop Var	0.339	0.424	0.450	0.429	0.455	0.461	0.4458
Fit	0.706	0.870	0.8964	0.881	0.916	0.922	0.924

By comparing Table 2 and Table 10 we see that there is direct relationship between parameters of IRT and those of factor models, particularly the discrimination parameter and the

factor loadings. We note that items with high discrimination values load highly on factors. The discrimination values of these items are greater than one. Thus, for an item to be influential in the formation of a factor, it should possess a discriminatory power with absolute value greater than one. Also, the results show that for $n = 30$, the loading for Item 5 is negative. We observe that this is as a result of negative estimate of discrimination parameter (see Table 2). In item response modelling, this implies that persons with high ability rather have a low probability of positive response (see graphs in Figure 3). However, this observation is associated with low sample size ($n = 30, 100, 150$). This is potentially problematic. The item interpretation changes accordingly for Item 5. The expected discrimination of the items is better achieved in larger sample sizes. The results show that it is difficult to obtain reasonable parameter estimates for smaller sample sizes. We observe from Table 10 that the number of influential indicators appear to converge (at 15) for higher sample size starting at $n = 150$. The indicators are the same at point of convergence. The result also shows that at low sample size, specific indicators have no influence or do influence in a different direction, for example, Variable 5. The proportion of variance accounted for by the single factor increases from 33.9% (for $n = 30$) to a highest of 46.1% (for $n = 800$).

Table 11 contains loadings of two-factor solutions of unidimensional item response dataset for various sample sizes on dichotomous scale.

Table 11. Loadings of two-factor solutions of unidimensional item response dataset for various sample sizes on dichotomous scale

	Sample Size 30		Sample Size 100		Sample Size 150		Sample Size 200	
Item	PA1	PA2	PA1	PA2	PA1	PA2	PA1	PA2
1	0.371	0.196	0.232	0.268	0.389	0.000	0.277	0.104
2	0.136	0.858	0.000	0.582	0.443	0.142	0.326	0.324
3	0.519	-0.255	0.608	0.000	0.126	0.381	0.271	0.000
4	0.338	-0.198	0.196	0.677	0.454	-0.138	0.217	0.350
5	0.152	-0.576	0.188	-0.170	0.141	-0.660	0.000	0.453
6	0.000	0.322	0.392	0.552	0.634	0.174	0.530	0.577
7	-0.132	0.437	0.000	0.339	0.757	0.155	0.000	0.673
8	0.495	0.345	0.761	0.242	0.821	0.139	0.603	0.504
9	0.515	0.000	0.405	0.536	0.734	0.000	0.371	0.768
10	0.215	0.506	0.594	0.533	0.683	-0.135	0.625	0.000
11	0.651	0.605	0.560	0.545	0.763	0.139	0.715	0.322
12	0.406	0.705	0.647	0.579	0.829	0.214	0.758	0.470
13	0.722	0.576	0.811	0.404	0.927	0.000	0.729	0.407
14	0.565	0.136	0.909	0.195	0.721	0.307	0.708	0.295
15	0.565	0.687	0.803	0.332	0.825	0.220	0.703	0.490
16	0.488	0.000	0.218	0.432	0.522	0.496	0.537	0.388
17	0.715	0.188	0.400	0.517	0.656	0.364	0.757	0.370
18	0.618	0.185	0.420	0.152	0.525	0.280	0.301	0.520
19	0.855	0.000	0.560	0.392	0.629	0.000	0.505	0.124
20	0.614	0.522	0.799	0.365	0.774	0.226	0.650	0.641
Prop Var	0.256	0.195	0.300	0.183	0.427	0.072	0.286	0.194
Cum Var	0.451		0.483		0.499		0.480	
Fit	0.801		0.896		0.917		0.904	

Table 11 continued

	Sample Size 500		Sample Size 800		Sample Size 1000	
Item	PA1	PA2	PA1	PA2	PA1	PA2
1	0.185	0.204	0.304	0.000	0.211	0.185
2	0.268	0.309	0.285	0.482	0.394	0.188
3	0.289	0.286	0.271	0.433	0.223	0.321
4	0.355	0.156	0.262	0.327	0.344	0.151
5	0.000	0.471	0.000	0.487	0.000	0.383
6	0.793	0.222	0.688	0.285	0.791	0.271
7	0.624	0.263	0.683	0.183	0.713	0.190
8	0.891	0.195	0.812	0.276	0.801	0.293
9	0.677	0.399	0.686	0.397	0.679	0.195
10	0.440	0.596	0.686	0.000	0.507	0.468
11	0.521	0.405	0.683	0.307	0.688	0.372
12	0.575	0.709	0.811	0.304	0.695	0.498
13	0.700	0.529	0.830	0.326	0.677	0.558
14	0.573	0.513	0.691	0.425	0.640	0.434
15	0.686	0.576	0.703	0.420	0.636	0.585
16	0.422	0.487	0.615	0.253	0.553	0.348
17	0.645	0.361	0.674	0.327	0.532	0.517
18	0.523	0.196	0.584	0.000	0.489	0.416
19	0.521	0.327	0.534	0.209	0.522	0.231
20	0.682	0.501	0.796	0.334	0.683	0.513
Prop Var	0.314	0.172	0.387	0.105	0.333	0.145
Cum Var	0.486		0.492		0.478	
Fit	0.925		0.934		0.930	

In the table, with exception of sample sizes $n = 150$ and $n = 800$, there is generally the incidence of repetition of high loadings on the same indicator variable of the two factors which can distract interpretation. However, for $n = 150$, the first factor loads highly on as many as 15 indicators and explains 42.7% of variance. The second factor loads highly on only one indicator (Variable 5) and is a contrast to its representation in IRT. In addition, amount of variance explained by the second factor appears to be negligible. These observations are consistent with the correlation matrix as Variable 5 has negative correlation with most of the other variables. The sample size of $n = 150$ thus gives a more plausible factor solution than all other samples. The $n = 150$ also explains the highest cumulative variation. For $n = 800$ the second factor is rather considered as redundant.

Table 12 are loadings of three-factor solutions of unidimensional item response dataset for various sample sizes on dichotomous scale.

From Table 12, the result becomes less meaningful and even unrealistic for sample sizes beyond 30. There is generally the incidence of repeating indicators on multiple factors. There is also the incidence of unrealistic loadings that are greater than 1 in higher factor numbers, particularly for Factor 3. Specifically, the loadings of Item 5 on Factor 3 is greater than 1 for $n = 150$ and 200. This incidence is as a result of an extraction of higher factor structure from a lesser dimensional dataset on a low scale point. It is interesting that it is at the optimal sample size that these weaknesses are revealed.

Table 12. Loadings of three-factor solutions of unidimensional item response dataset for various sample sizes on dichotomous scale

	Sample Size 30			Sample Size 100		
Item	PA1	PA2	PA3	PA1	PA2	PA3
1	0.785	0.000	-0.106	0.340	0.000	0.536
2	0.302	0.000	0.818	0.000	0.581	0.141
3	0.152	0.515	-0.232	0.635	0.000	0.000
4	0.127	0.312	-0.202	0.231	0.531	0.475
5	0.000	0.205	-0.544	0.125	0.000	-0.379
6	0.000	0.000	0.360	0.407	0.480	0.251
7	-0.119	0.000	0.509	0.000	0.270	0.199
8	0.495	0.265	0.245	0.773	0.233	0.000
9	0.000	0.724	0.155	0.437	0.429	0.343
10	0.610	0.176	0.334	0.519	0.662	0.000
11	0.742	0.289	0.453	0.558	0.509	0.202
12	0.377	0.280	0.672	0.621	0.594	0.114
13	0.593	0.494	0.499	0.730	0.607	-0.285
14	0.000	0.737	0.253	0.920	0.202	0.000
15	0.374	0.484	0.708	0.805	0.383	0.000
16	0.000	0.554	0.000	0.145	0.541	0.000
17	0.753	0.346	0.000	0.339	0.605	0.000
18	0.721	0.246	0.000	0.433	0.127	0.000
19	0.498	0.675	0.000	0.547	0.401	0.000
20	0.453	0.465	0.486	0.773	0.411	0.000
Prop Var	0.206	0.170	0.169	0.288	0.189	0.053
Cum Var			0.544			0.530
Fit			0.864			0.916

Table 12 continued

	Sample Size 150			Sample Size 200		
Item	PA1	PA2	PA3	PA1	PA2	PA3
1	0.435	0.000	0.000	0.206	0.189	-0.189
2	0.301	0.379	0.000	0.265	0.389	0.000
3	-0.120	0.617	-0.165	0.184	0.000	-0.168
4	0.398	0.000	0.212	0.215	0.341	0.000
5	0.000	-0.121	1.019	0.188	0.264	1.200
6	0.622	0.249	0.000	0.406	0.718	-0.125
7	0.572	0.513	0.158	0.000	0.709	0.124
8	0.809	0.285	0.000	0.564	0.546	0.000
9	0.741	0.204	0.000	0.385	0.717	0.000
10	0.401	0.473	0.429	0.519	0.194	-0.198
11	0.753	0.268	0.000	0.665	0.392	0.000
12	0.644	0.568	0.000	0.756	0.481	0.000
13	0.696	0.539	0.301	0.709	0.440	0.000
14	0.595	0.502	0.000	0.727	0.292	0.000
15	0.911	0.205	0.000	0.751	0.448	0.101
16	0.472	0.452	-0.172	0.499	0.431	0.000
17	0.421	0.689	0.000	0.682	0.465	-0.174
18	0.336	0.539	0.000	0.236	0.579	-0.107
19	0.486	0.385	0.114	0.636	0.000	0.130
20	0.790	0.279	0.000	0.633	0.652	0.000
Prop Var	0.328	0.170	0.074	0.266	0.214	0.084
Cum Var			0.572			0.563
Fit			0.937			0.918

5. Conclusions

The study shows that there is a direct relationship between parameters of IRT and those of factor models, particularly item discrimination and factor loadings. In this regard, items with high discrimination values load highly on factors. Such items possess a discriminatory power with absolute value greater than one.

The study reveals that for smaller sample size, particularly below 100, items/indicators may not generate the desired dataset. That is, the data generated may not follow the desired model. We may, therefore, not be able to obtain a reasonable factor solution. On the other hand, we could obtain unrealistic factor solution if we attempt to extract higher factor solution than the underlying dimensionality on few scale-points. This particularly shows that extracting more factors than necessary could run into difficulties, especially for low scale points.

The results also show that the amount of cumulative variation explained increases with increasing scale-points. It follows that, the number of influential indicators on factors increases with increasing scale-points. This means that, factors are more well defined and could be most interpretable on larger scale-points. The results are, however, almost the same on higher scale-points of five and seven.

Generally, a sample size of 150 produces a more consistent factor solution based on the underlying dimensionality of the data. However, in some cases of the factor structure (particularly, high dimensional datasets), a sample size of 100 gives a more consistent result.

In factor analysis, generally, results appear reasonable on higher scale points irrespective of sample, even though a sample size of 150 stands out. However, on IRT model, results are particularly not good for small sample size and at higher scale points. It will therefore be important to examine the IRT model, along with factor models on Likert scale data. This has the potential to help obtain the right interpretations of factors.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] N. Balakrishnan, *Handbook of the Logistic Distribution*, Taylor and Francis, New York (1991).

[2] I. H. Bernstein and G. Teng, Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization, *Psychological Bulletin* **105** (1989), 467 – 477.

[3] R. P. Chalmers, mirt: A multidimensional item response theory package for the R environment, *Journal of Statistical Software* **48**(6) (2012), 1 – 29, DOI: 10.18637/jss.v048.i06.

[4] A. L. Comrey and H. B. Lee, *A First Course in Factor Analysis*, 2nd ed., Lawrence Erlbaum Associates, New Jersey (1992).

[5] A. Cyr and A. Davies, Item response theory and latent variable modeling for surveys with complex sampling design: The case of the National Longitudinal Survey of Children and Youth in Canada, in *Conference of the Federal Committee on Statistical Methodology*, Office of Management and Budget, Arlington, VA (2005).

[6] R. J. de Ayala, *The Theory and Practice of Item Response Theory*, The Guilford Press, New York (2009).

[7] J. de Leeuw, Models and methods for the analysis of correlation coefficients. *Journal of Econometrics* **22** (1983), 113 – 137, DOI: 10.1016/0304-4076(83)90096-9.

[8] L. R. Fabrigar and D. T. Wegener, *Exploratory Factor Analysis*, Oxford University Press, New York (2012).

[9] P. J. Ferrando and U. Lorenzo-Seva, Unrestricted item factor analysis and some relations with item response theory (Tech. Rep.), Department of Psychology, Universitat Rovira i Virgili, Tarragona, retrieved from `http://psico.fcep.urv.es/utilitats/factor/` (2013).

[10] R. Gorsuch, *Factor Analysis*, W. B. Saunders Company, Philadelphia (1974).

[11] J. Jacoby and M. S. Matell, Three-point Likert scales are good enough, *Journal of Marketing Research* **8** (1971), 495 – 500, DOI: 10.1177/002224377100800414.

[12] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed., Pearson Education, New York (2007).

[13] C. R. Kothari, *Research Methodology: Methods and Techniques*, 3rd ed., New Age International, New Delhi, India (2004).

[14] R. C. MacCallum, M. W. Browne and H. M. Sugawara, Power analysis and determination of sample size for covariance structure modeling, *Psychological Modeling* **1**(2) (1996), 130 – 149, DOI: 10.1037/1082-989X.1.2.130.

[15] W. S. Martin, The effects of scaling on the correlation coefficient: A test of validity, *Journal of Marketing Research* **10**(3) (1973), 316 – 318.

[16] P. D. Mehta, M. C. Neale and B. R. Flay, Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes, *Psychological Methods* **9**(3) (2004), 301 – 333, DOI: 10.1037/1082-989X.9.3.301.

[17] E. Muraki, A generalised partial credit model: An application of an EM algorithm. *Applied Psychological Measurement* **16** (1992), 159 – 176.

[18] P. Osteen, An introduction to using multidimensional item response theory to assess latent factor structures, *Journal of the Society for Social Work and Research* **1**(2) (2010), 66 – 82, DOI: 10.4135/9781412985413.

[19] R. Ostini and M. L. Nering, *Polytomous Item Response Theory Models*, Sage Publications, California (2006).

[20] R Core Team, *R: A Language and Environment for Statistical Computing*, Computer software manual, Vienna, Austria, retrieved from `https://www.R-project.org/` (2017).

[21] M. D. Reckase, *Multidimensional Item Response Theory*, Springer, New York (2009).

[22] B. B. Reeve, *An Introduction to Modern Measurement Theory*, National Cancer Institute, 1 – 67 (2002).

[23] S. P. Reise, K. F. Cook and T. M. Moore, Evaluating the impact of multidimensionality on unidimensional item response theory model parameters, in: S. P. Reise and D. A. Revicki (eds.), *Handbook of Item Response Theory: Applications to Typical Performance Assessment*, pp. 13 – 40, Taylor and Francis, New York (2015).

[24] S. P. Reise and Y. Yu, Parameter recovery in the graded response model using MULTILOG, *Journal of Educational Measurement* **27**(2) (1990), 133 – 144, DOI: 10.1111/j.1745-3984.1990.tb00738.x.

[25] A. C. Rencher, *Methods of Multivariate Analysis*, 2nd ed., John Wiley & Sons, New York (2002).

[26] W. Revelle, *psych: Procedures for Psychological, Psychometric, and Personality Research*, Computer software manual, Evanston, Illinois, retrieved from `https://CRAN.R-project.org/package=psych` (R package version 1.7.8) (2017).

[27] C. A. Stone, Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG, *Applied Psychological Measurement* **16** (1992), 1 – 16, DOI: 10.1177/014662169201600101.

[28] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 6th ed., Pearson Education, New Jersey (2013).

[29] Y. Takane and J. de Leeuw, On the relationship between item response theory and factor analysis of discretized variables, *Psychometrika* **52**(3) (1987), 393 – 408, DOI: 10.1007/BF02294363.

[30] C. van der Eijk and J. Rose, *Risky business: Factor analysis of survey data - Assessing the probability of dimensionalisation*, *PLoS ONE* **10**(3) (2015), e0118900, DOI: 10.1371/journal.pone.0118900.