



Universidad Autónoma  
de Madrid

**Biblos-e Archivo**  
Repositorio Institucional UAM

Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Structural Equation Modeling: A Multidisciplinary Journal 27.5 (2020): 781–797

DOI: <https://doi.org/10.1080/10705511.2019.1668276>

**Copyright:** © Taylor & Francis Group, LLC

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

## **Factor analysis for nominal (first choice) data**

Javier Revuelta<sup>1</sup>

Alberto Maydeu-Olivares<sup>2,3</sup>

Carmen Ximénez<sup>1</sup>

<sup>1</sup> *Department of Psychology, Autonomía University of Madrid*

<sup>2</sup> *Department of Psychology, University of South Carolina*

<sup>3</sup> *Department of Psychology, University of Barcelona*

### **Author note**

Javier Revuelta.  <https://orcid.org/0000-0003-4705-6282>

Alberto Maydeu-Olivares.  <http://orcid.org/0000-0001-5790-392X>

Carmen Ximénez.  <https://orcid.org/0000-0003-1337-6309>

Correspondence concerning this article should be addressed to Javier Revuelta (e-mail: javier.revuelta@uam.es). This research was partially supported by grant PSI2015-66366-P from the Spanish Department of Economy and Competitiveness (Revuelta, PI) and by grant SES-1659936 from the National Science Foundation (Maydeu-Olivares, PI). The authors acknowledge the support of the Center for Scientific Computing at the Autonomía University of Madrid (CCC-UAM).

**Abstract**

We show how a factor analysis model can be fitted to nominal data using the computer program Mplus. The model is akin to multinomial logistic regression with unobserved predictors (the common factors) and was initially proposed by Bock (1972) in the one-factor case. Recently, extensions to multiple factors and alternative parameterizations to facilitate parameter interpretation have been proposed. We present four examples in which several versions of the model are estimated using Mplus: a) a one-factor model applied to situational items measuring assertiveness, b) an exploratory factor analysis applied to attitudinal data, c) a confirmatory factor analysis applied to educational data with testlets, and d) the newest parameterization of the model applied to an emotional stability scale. All data files and computer codes are provided as supplementary materials.

**Keywords:** nominal factor analysis, multidimensional nominal categories model, situational judgment tests, item response theory, IRT.

### **Factor analysis for nominal (first choice) data**

To this date, factor analysis remains one of the statistical workhorses in the social sciences (Brown, 2015; Cudeck & MacCallum, 2007). The common factor model is suitable for continuous outcomes such as test scores. A model suitable for ordinal outcomes such as Likert-type items can be obtained by adding a threshold process to a common factor model. The resulting model is often referred to as ordinal factor analysis model in the literature (Jöreskog & Moustaki, 2001, 2006). Is there a factor analysis model for nominal data? Yes, there is, and interestingly the model has existed for quite some time now as it was introduced by Bock (1972). However, the model was introduced for an item response theory (IRT) audience (and referred to as nominal categories model –NCM) and as a result, many users of factor analysis may not be aware of its existence.

We need to distinguish between two kinds of nominal data. Most often, when we think of nominal data we think of purely unordered data such as country of residence (1 = US, 34 = Spain, etc.). In these cases, factor analysis is not applicable. However, in some cases, we can conceive the existence of an underlying order among the response alternatives, and the observed outcome the result of a decision making process. For instance, if we ask “in what country would you like to retire?”, we could use the same numeric codes as before, but the obtained data is different; it is first choice data. In first choice data, we assume that the respondent orders the alternatives in her mind but only provides her top choice to follow the instructions received (Bock, 1997; Maydeu-Olivares & Böckenholt, 2005; Maydeu-Olivares & Böckenholt, 2009). First choice data is still nominal data, as the response alternatives are unordered. However, because of the assumed underlying order, one of the major tasks in applications is to ‘uncover’ the ordering of the alternatives presented. As such, first choice data is widely used in Marketing and Consumer Psychology applications, in which the objective is to scale the response alternatives (Maydeu-Olivares & Böckenholt, 2009).

However, when common factors are assumed to underlie the ordering of response alternatives (Maydeu-Olivares & Böckenholt, 2005; Maydeu-Olivares & Böckenholt, 2009; Maydeu-Olivares & Brown, 2010), first choice data can be used to obtain individuals’ scores on those underlying traits. In other words, first choice data, a type of unordered data, can be factor analyzed, provided a suitable model, such as that proposed by Bock (1972), is used.

Many empirical applications involve the use of use nominal variables that consist of a number of response alternatives in which it is reasonable to assume an underlying ordering. One example are multiple-choice test data from educational measurement (Thissen, Steinberg, & Fitzpatrick, 1989). This has been arguably the most popular area of application of the factor analysis model and the area in which most of the model developments originated. To see how the model can be applied in this case, consider a set of mathematical problems (items) each with four response alternatives. The items can be scored binarily and ordinal factor analysis can be applied. Alternatively, a set of experts may order the four response alternatives in terms of degree of accuracy. In this case, an ordinal factor analysis model may be applied again. Lastly, arbitrary scores may be assigned to the response alternatives (so that the data is nominal), and factor analysis for nominal data be applied. This amounts to consider that the respondent has ordered the alternatives in terms of degree of accuracy and has provided her 'preferred response' (the one she believes is most likely to be correct).

However, there are many other types of items that originate nominal variables. Situational judgment tests (Campion & Ployhart, 2012; Ployhart & Ward, 2013; St-Sauveur, Girouard, & Goyette, 2014; Whetzel & McDaniel, 2016), widely used in personnel assessment are another area of application of this model. In these tests, individuals may be presented a number of hypothetical scenarios and be asked to identify the action that they most likely take; the set of actions have no explicit ordering thus generating a nominal response. Survey research is another area of application of this model. Here, individuals may be asked to select options reflecting opinions or preferences from an unordered set. Another area of application of these techniques is marketing research, where consumers are asked to provide their preferences about predefined lists of products. One final area of application is personality and competence assessment. In this case, individuals can be presented with a set of situations and for each one be presented with a set of alternative courses of action. Individuals are then asked to choose the alternative that they are most likely to perform in each situation. A factor analysis model for nominal data is appropriate because the alternatives are unordered, but may reflect different degrees of the unobserved construct of interest, for instance, assertiveness. Finally, the NCM has been applied to ordinal Likert-type responses in the area of attitude and personality measurement to estimate the ordering of the alternatives and to separate substantive from response style factors (Thissen, Cai & Bock, 2010).

When a factor model is applied to nominal data, the researcher will obtain a set of factor loadings and factor correlations, and, if requested, a set of factor scores reflecting the common factors (aka latent traits) of interest, just as in the common factor model. However, whereas in the common factor model one obtains one loading per factor per item (e.g., the loading of assertiveness on item 1), in nominal factor analysis one obtains one loading per factor per item alternative (e.g., if each assertiveness item provides three alternatives to choose from, three loadings per item, one per alternative, will be estimated). Therefore, the interpretation of model parameters is more cumbersome than in factor analysis for continuous outcomes, or in ordinal factor analysis.

The aim of this article is to help bridge the gap between the factor analysis and IRT traditions by introducing the nominal categories model to a factor analysis audience. Until recently, this model was only available through IRT software (e.g., flexMIRT: Cai, 2015; IRTPro: Cai, du Toit, & Thissen, 2011; Multilog: Thissen, Chen, & Bock, 2003; Latent GOLD, Vermunt & Magidson, 2016; and the R package mirt, Chalmers, 2012). In addition, until the advent of the most recent IRT programs (flexMIRT, IRTPro, Latent GOLD, mirt) only unidimensional models (i.e., models with a single common factor) could be estimated. However, the nominal categories model has been available (scarcely documented) for some time now in one of the most widely used structural equation modeling programs, Mplus (Muthén & Muthén, 2016), and in this article we will show how to use Mplus to fit factor analysis models to nominal (first choice) data. Mplus has great flexibility to impose parameter constraints, enabling the estimation of different parameterizations of the model, and in this paper, we show how to do so. In the supplementary materials to this article we also provide flexMIRT code to reproduce some the examples provided, as this program has clear advantages over the current implementation of Mplus to fit models when some specific parameterizations of the NCM are used.

The remaining of the paper is organized into three sections of increasing technical intricacy. Each section begins by introducing the methods, followed by an application. The data and computer code are provided as supplemental material for the user to adapt these codes to her specific purposes. We begin with the basic one factor model introduced by Bock and apply it to a situational test to measure assertiveness. In the second section, we describe the multidimensional version of the model and provide nominal exploratory and confirmatory factor analysis examples. The third section contains the recent parameterization of the model by Thissen, Cai

and Bock (2010) and its application to evaluating response styles. We conclude with some remarks and recommendations to applied researchers.

### **Relationships between the common factor model, the ordinal factor model, and the nominal factor model**

In the common factor model, a set of unobserved predictors (the common factors) is used to account for the observations in a set of outcome variables. The common factor model is a linear model and both the outcome and predictor variables are assumed to be continuous. As a result, the common factor analysis model can be described as multivariate regression with unobserved continuous predictors.

Ordinal responses can be accommodated within a factor analysis framework by adding a threshold process to a common factor model. The model assumes the existence of a continuous response tendency underlying each item. It also assumes that a common factor model holds for these response tendencies. The response tendencies are unobserved because individuals are forced to respond using a discrete scale. The model further assumes that individuals will choose a) the highest response category if their response tendency is above the largest threshold, b) the lowest category if their tendency is below the smallest threshold, and c) an intermediate category if their continuous response tendency is between two cutoff points (thresholds). Assuming that these response tendencies are normally distributed, the resulting model is the ordinal factor analysis model. Because the outcome variables are discrete ordinal, the predictors are continuous, and an underlying normal distribution is assumed, the ordinal factor analysis model can be simply described as a multivariate probit regression model with unobserved continuous predictors (the common factors).

The nominal factor analysis model is a model for first choice data, not for nominal data proper. That is, the model is suitable to model tasks in which a) individuals are presented a set of more than two alternatives and are asked to choose their preferred option, and b) the different options are indicators of a set of common factors (latent traits). Mathematically, the model takes the form of a multivariate multinomial logistic regression (Agresti, 2002; Smithson & Merkle, 2013) with unobserved predictors. The model originated within the item response theory (IRT) literature where it is referred to as (multidimensional) nominal categories model. Throughout this

article we use the terms “nominal factor analysis”, “Bock’s (1972) nominal model”, and “nominal categories model” as interchangeable.

Factor analysis and IRT are two closely related fields that in many cases yield equivalent models. The graded response model described in the IRT literature (McDonald, 1997; Samejima, 1969) is mathematically equivalent to the ordinal factor analysis model (Takane & de Leeuw, 1987). One difference between the factor analysis and IRT traditions when it comes to modeling discrete ordinal data is the choice of the response function. In the factor analysis tradition, the normal distribution function is the function of choice. In contrast, in the IRT tradition, the logistic function is vastly more popular. Hence, when a reference is made to the graded response model in the IRT literature, one is to assume that a logistic distribution function is being used. Samejima’s graded logistic model can be described as multivariate ordinal logistic regression with unobserved predictors (the latent traits/common factors).

Another difference between the factor analysis and IRT traditions when it comes to modeling discrete ordinal data is the choice of estimation method. In the factor analysis tradition, two-stage estimation procedures in which polychoric correlations are estimated in a first stage and model parameters are estimated from the polychoric correlations in a second stage are the most commonly used estimation procedures (Muthén, 1984). In the IRT tradition, the method of choice is (marginal) maximum likelihood estimation from the observed ordinal responses under multinomial data assumptions (Cai & Thissen, 2015).

### **The nominal categories model**

The classical article by Bock (1972) “Estimating item parameters and latent ability when responses are scored in two or more nominal categories” introduced a model that may be used to perform factor analysis on nominal data, although he only considered models with a single factor. Generalizations of Bock’s model involving multiple common factors (latent traits) were proposed by Takane and de Leeuw (1987), Thissen, Cai and Bock (2015) and Revuelta (2014). In this section we describe the unidimensional (one factor) case.

Consider an item,  $Y$ , consisting of  $K$  unordered response alternatives (categories),  $Y = 1, 2, \dots, K$ . When a nominal factor analysis model with a single factor,  $\eta$ , is assumed to underlie the responses to the item, there is an intercept,  $\nu_k$ , and a factor loading (slope),  $\lambda_k$ , associated to



each of the response alternatives. The probability of endorsing a particular alternative, say  $k$ , for a given value of  $\eta$  is

$$\Pr(Y = k | \eta) = \frac{\exp(v_k + \lambda_k \eta)}{\sum_{m=1}^K \exp(v_m + \lambda_m \eta)} \quad (1)$$

Except for the choice of notation, and for the fact that the predictor (the common factor) is unobserved, this is just the formula for multinomial logistic regression. In fact, letting  $p$  denote the number of items,  $p$  multinomial logistic regressions are estimated simultaneously in which the predictor is the common factor. A parametric distribution is usually assumed for the latent factor. Here, we assume that the common factor is normally distributed. Estimation of the model proceeds by maximum likelihood using the EM algorithm (Bock & Aitkin, 1981).

The one factor nominal model we have described is not identified. To understand why, we compute log-odds. The log-odds of category  $k$  against category  $j$  is

$$\log \frac{\Pr(Y = k | \eta)}{\Pr(Y = j | \eta)} = (v_k - v_j) + (\lambda_k - \lambda_j) \eta. \quad (2)$$

The log-odds depend on differences between parameters and do not change if a constant value is added to all the parameters. This indeterminacy is the source of identification problems in this model. To identify the model, exclusion or deviation constraints on the parameters can be used.

### Exclusion and deviation constraints

The simplest constraint that can be used to identify the model consists of fixing to zero the parameters of one category (say category  $K$ ). In this fashion, the log-odds become a function of a single intercept and slope

$$\log \frac{\Pr(Y = k | \eta)}{\Pr(Y = K | \eta)} = v_k + \lambda_k \eta \quad (3)$$

With this constraint, referred to as *exclusion constraint*, the category intercept and slope,  $v_k$  and  $\lambda_k$ , represent the change in the log-odds in relation to  $\eta$ . Note that the probability function in Eq. (1) depends on all the parameters simultaneously, and thus the parameters of one category cannot be related to the probability of that category without taking into account the parameters of the other categories. However, the log-odds in Eq. (3) depend only on the parameters of the category.

Exclusion parameter constraints are the standard way to identify multinomial logistic models. They are a convenient way to identify these models when there is a meaningful reference category. Consider multiple-choice items, the substantive area for which the nominal factor model was developed. In these items, there is a correct answer and the remaining categories are distractors. In this case, there is a meaningful reference category, the correct answer, and negative slopes indicate that the probability of endorsing the distractors decreases relative to the probability of passing the item as the ability level,  $\eta$ , increases. In Likert-type items with an odd number of response alternatives, the middle category, often labeled as “Indifferent”, “I Don’t Know”, etc. may serve as a meaningful reference category.

When the items do not have a category that can be interpreted naturally as a baseline, other types of identification constraints may be preferable. When first introducing the model, Bock (1972) used *deviation constraints* to identify the model. This amounts to setting the sum of the intercepts to be zero, and the sum of the factor loadings to be zero as well:

$$\sum_{k=1}^K \tilde{v}_k = 0, \quad \sum_{k=1}^K \tilde{\lambda}_k = 0 \quad (4)$$

In this equation we use the tilde  $\sim$  to distinguish deviation constrained parameters from the exclusion constrained parameters (without tildes). Deviation constrained parameters can be different from zero for all the categories and are interpreted by comparing pairs of categories using (2) or by comparing parameters to the zero value (the mean of the parameters for that item).

One can transform the parameters obtained using exclusion constraints to deviation-constrained parameters by subtracting the mean of the exclusion constrained item parameters

$$\begin{aligned} \tilde{v}_k &= v_k - \bar{v} \\ \tilde{\lambda}_k &= \lambda_k - \bar{\lambda} \end{aligned} \quad (5)$$

for  $k = 1, \dots, K-1$ , where the means are

$$\bar{v} = \frac{1}{K-1} \sum_{k=1}^{K-1} v_k, \quad \bar{\lambda} = \frac{1}{K-1} \sum_{k=1}^{K-1} \lambda_k. \quad (6)$$

For instance, in a model with three response categories ( $K = 3$ ), there are six deviation constrained parameters per item (three intercepts and three slopes). These can be estimated from the four exclusion constrained parameters as follows:

$$\begin{aligned}
 \tilde{v}_1 &= v_1 - \frac{v_1 + v_2}{3} & \tilde{\lambda}_1 &= \lambda_1 - \frac{\lambda_1 + \lambda_2}{3} \\
 \tilde{v}_2 &= v_2 - \frac{v_1 + v_2}{3}, & \tilde{\lambda}_2 &= \lambda_2 - \frac{\lambda_1 + \lambda_2}{3}, \\
 \tilde{v}_3 &= -\frac{v_1 + v_2}{3} & \tilde{\lambda}_3 &= -\frac{\lambda_1 + \lambda_2}{3}
 \end{aligned} \tag{7}$$

where we have assumed that the last category is used as reference category when imposing exclusion constraints; that is,  $v_3 = \lambda_3 = 0$ .

### Empirical application: The assertiveness situational test

An eight-item questionnaire was designed to measure the assertiveness of Chinese young adults. The questionnaire is provided in the supplementary materials to this article. The first item of the questionnaire is

1. You have spent ~~Buy 100 tickets~~ to attend a music concert. Security at the concert's entrance refuses to let you in, claiming that the ticket you hold is a fake. It is a genuine ticket. You...
- A. ask to talk to the supervisor
  - B. call your parents for suggestions
  - C. after arguing with the concert's security, you give up and leave.

Individuals are asked to choose the alternative that they are most likely to perform in this situation. As we discussed in the introduction, one way to score items such as this is to rely on a group of experts to order the alternatives and analyze the data using ordinal factor analysis. An alternative is to use a nominal factor analysis model and let the data (i.e., the respondents) inform us of the ordering of the response alternatives, which can be determined by the item's intercepts.

This example illustrates the application of the NCM in the original, one-dimensional, version. The questionnaire was applied to a sample of 797 college students. Of these, 224 were male and 573 female; age ranged from 17 to 23 years. As we see in the example item, each alternative indicates possibly a different degree of assertiveness, but the alternatives are not ordered. The questionnaire is intended to measure a single factor, assertiveness.

We provide in Table 1 the parameter estimates obtained using Mplus along with their standard errors estimated using the observed information matrix. These are obtained under exclusion contrast as this is the default parameterization for this model in Mplus.

-----

Insert Table 1 about here

-----

The last response alternative for item 1 is "quit". We see in table 1 that for item 1, the estimated intercept for alternative A ("ask to talk to the supervisor") is positive and statistically significant at the 5% level. We interpret this as follows: When confronted to the situation described in item 1, individuals with an average level of assertiveness ( $\eta = 0$ ) are more likely to ask to talk to the supervisor than to simply quit. Similarly, the estimated intercept for alternative B ("call parents") is not statistically significant. We interpret this to mean that individuals with an average level of assertiveness are equally likely to call their parents for help than to quit. We do not know whether an individual with an average level of assertiveness is more likely to ask to talk to a supervisor or to call her parents. We can find out in two ways:

1. Using a likelihood ratio test: We fit a model where both intercepts are equal and compute -2 times the difference of log-likelihoods between this model and the model unconstrained intercepts. The resulting statistic is asymptotically distributed as a chi-square with degrees of freedom equal to the difference in degrees of freedom (Maydeu-Olivares & Cai, 2006).
2. Using a Wald test: We estimate the difference between the intercepts for the first and the second category,  $d_{i,12} = v_{i,1} - v_{i,2}$  using the MODEL CONSTRAINT command in Mplus. The program will output the estimated difference  $d_{i,12}$ , its standard error, and the  $p$ -value for the hypothesis that the population value of  $d_{i,12}$  is zero.

Turning to the interpretation of the loadings, we see that both loadings for item 1 are statistically significant and positive. We interpret this as follows: When confronted to the situation described in item 1, the relationship between the trait measured (assertiveness) and alternative A ("talk to supervisor") is larger than the relationship between the trait and the reference category (C, "quit"). Similarly, the relationship between assertiveness and alternative B ("call parents") is statistically larger than the relationship between assertiveness and alternative C.

Turning now to the results for item 2 presented in Table 1, we see that none of the slopes is statistically significant at the 5% level. Strictly speaking, this means that neither the loading for alternative A nor for alternative B are larger than the loading for alternative C. However, this may be an indication that this item does not provide information on the assertiveness level of the respondents and therefore it could be removed from future administrations of the questionnaire.

To verify this, we re-estimated the model setting both slopes to zero and computed -2 times the difference in log-likelihoods:  $G_{dif}^2 = -2 \times (-4,163.204 + 4,162.578) = 1.25$  on 2 df,  $p = .54$ . Because a model with this item's slopes set to zero cannot be statistically distinguished from a model in which the slopes are estimated, we conclude that this item does not provide information on the level of assertiveness of the respondents.

We note in Table 1 that the estimated SEs for the parameter estimates are rather large, despite the sample size being 797. This is because rather large samples are needed to accurately estimate the parameters of the nominal model (Wollack, Bolt, Cohen, & Lee, 2002). Bayesian estimation approaches overcome this problem by introducing a priori distributions for the parameters that incorporate additional information and stabilize estimates. Bayesian inference can be applied to the NCM (Revuelta & Ximénez, 2017) although the Bayesian methods implemented in Mplus are not applicable to nominal variables at the time of this writing.

Finally, in applications, the sign of the loadings must be examined carefully, as it is possible that the program estimates 'lack of assertiveness' instead of 'assertiveness'. This was indeed the case here; with default starting values Mplus estimates 'lack of assertiveness'. To reverse the sign of the latent trait being estimated, the model needs to be re-estimated providing starting values of the desired sign.

For this example, we also obtained the deviation constraints parameters; see the supplementary materials to this article for the Mplus and flexMIRT code, and Table 2 for the parameter estimates and SEs. We see in Table 2 that we now obtain intercepts and factor loadings for all response alternatives. However, the intercepts add up to zero within an item, and so do the factor loadings. In item 1 ("You're denied entry to a concert"), the estimated intercepts with SEs in parentheses are 1.70 (.08), -.87 (.11) and -.83 (.13), respectively. We interpret this as implying that individuals with an average assertiveness are more likely to endorse alternative A ("ask to talk to a supervisor") than alternatives B or C (as the intercept for A is statistically different from the intercepts for B and C). However, alternatives B and C are equally attractive to individuals with an average assertiveness (as the intercepts are not statistically different from each other). Notice that whether an intercept is statistically different from zero is not particularly relevant, as zero is simply the average value of the intercepts.

Turning now to the factor loadings, we see that for item 1 the estimates for response alternatives A and B are positive, but close to zero, while the loading for alternative C is negative

and statistically significant. We interpret the difference between the loadings of alternatives A and C, and the difference between B and C. This implies that assertive individuals are less likely to endorse alternative C than the other two.

-----  
 Insert Table 2 about here  
 -----

### **Comparing parameter estimates across items**

Comparing parameter estimates across items requires some thought in this model. As (2) shows, strictly speaking only differences of parameters can be estimated. If the model is applied to items in which a “Don’t Know” or similar response is used as reference category for every item, across items comparisons are meaningful. However, in the assertiveness example the reference category has a different meaning in each item. Therefore, comparing parameter estimates across items is not very meaningful in this example.

### **Scoring individuals**

Mplus estimates scores on the latent traits for each individual along with its standard error of measurement (SEs). In the common factor model for continuous outcomes, there is a common SE for all individuals. In the present model, as in ordinal factor analysis, the SE of the factor score varies as a function of the latent trait. Should new observations be available, they can be scored with Mplus using the estimated parameters. Details are provided in the supplementary materials to this article.

The three most widely used factor score estimator in factor analysis models for categorical data are the Bayesian expected a posteriori estimator (EAP), the Bayesian maximum a posteriori estimator (MAP), and the maximum likelihood estimator (ML) (Thissen & Wainer, 2001). Unlike the ML estimator, which is computed without reference to a population distribution for  $\eta$ , the MAP and EAP estimators use the assumed distribution of the factors to score individuals (Ferrando & Lorenzo-Seva, 2016). An important advantage of the EAP and MAP estimators over ML is that they yield estimates for all response patterns.

For nominal variables, Mplus computes EAP estimates of the factor scores (i.e., the mean of the posterior distribution of  $\eta$ ). The EAP estimator have good statistical properties, it has minimum mean square error, so it cannot be improved upon in terms of average accuracy (Bock & Mislevy, 1982). The EAP is regressed toward the mean because of the effect of the prior

distribution, although as the number of items increase the likelihood and the posterior become virtually indistinguishable (Ferrando & Lorenzo-Seva, 2016), which implies that the EAP and ML estimates become more similar.

### The multidimensional case

The multidimensional nominal categories model is a generalization of the NCM to multiple latent factors. We have seen in the previous section the interpretation of the parameters when a single factor is involved. When multiple factors are used the interpretation is similar but extended to a multidimensional latent space. Due to the heavy parameterization of the model, constrained versions of the model have been recently proposed that aim at facilitating interpreting the model results. In this section we first describe the model in its full generality (we refer to it as the full rank case). As we will illustrate, this model is well suited for exploratory factor analysis of nominal data. Constrained versions of this model, best suited for specific applications, are described in later sections.

#### The multidimensional nominal categories model

In this case, the probability function of  $Y$  conditional on  $\boldsymbol{\eta}$  is given by (Johnson & Bolt, 2010)

$$\Pr(Y = k | \boldsymbol{\eta}) = \frac{\exp(v_k + \boldsymbol{\lambda}_k' \boldsymbol{\eta})}{\sum_{m=1}^K \exp(v_m + \boldsymbol{\lambda}_m' \boldsymbol{\eta})}, \quad k = 1, \dots, K, \quad (8)$$

where, as before,  $v_k$ , is the intercept for category  $k$ . However, now  $\boldsymbol{\lambda}_k$  is a vector of slopes (factor loadings) and  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_D)'$  is a vector of  $D$  common factors (latent traits). Similar to the one-dimensional case, the parameters of the log-odds,

$$\ln \frac{\Pr(Y = k | \boldsymbol{\eta})}{\Pr(Y = j | \boldsymbol{\eta})} = v_k^* + \boldsymbol{\lambda}_k^{*'} \boldsymbol{\eta}, \quad v_k^* = v_k - v_j, \quad \boldsymbol{\lambda}_k^* = \boldsymbol{\lambda}_k - \boldsymbol{\lambda}_j, \quad (9)$$

are differences between the category parameters. Equations (8) and (9) remain invariant if a constant value is added to the parameters  $\mathbf{v}$  or  $\boldsymbol{\lambda}$  because  $\mathbf{v}_k^*$  and  $\boldsymbol{\lambda}_k^*$  are differences between parameters. In consequence, only  $K-1$  intercepts and  $D(K-1)$  slopes per item can be identified (estimated). The total number of parameters per item that can be estimated is therefore  $(K-1) + D(K-1) = (K-1)(D+1)$ . Exclusion or deviation constraints are imposed to identify

parameters. Thissen, Steinberg and Fitzpatrick (1989) and Thissen, Cai and Bock (2010) provide a useful insight of the parameters of the multidimensional NCM: The probability of choosing alternative  $k$  given that the response is either alternative  $k$  or  $j$  follows a multidimensional two-parameter logistic model with parameters  $v_k^*$  and  $\lambda_k^*$ . Preston et. al (2011) refer to  $\lambda_k^*$  as category boundary discrimination parameters because it enables to discriminate between alternatives  $k$  and  $k-1$  assuming that the response is in the set  $\{k, k-1\}$ .

### **Exploratory nominal factor analysis**

At the time of this writing, Mplus has no special routine to perform exploratory nominal factor analysis. In the common factor model (or the ordinal factor model), an exploratory model can be estimated using software that only performs confirmatory factor models as follows (McDonald, 1985): Let  $\Lambda$  be the  $p \times D$  matrix of factor loadings. An exploratory factor model can be obtained by setting the factors to be uncorrelated with variance one, and letting  $\Lambda$  to be a lower triangular matrix. These constraints enable obtaining an unrotated solution. More specifically, the first  $d-1$  loadings are set to zero for factors  $d = 2, \dots, D$ , so that each factor has one less estimated loadings. After this solution has been obtained, a more interpretable solution can be obtained by rotating it. Depending of the rotation used, the correlation matrix between the factors can also be estimated (Browne, 2001).

An unrotated solution for the nominal factor analysis model can be obtained similarly using Mplus. For this model, if  $p$  items consist of  $K$  categories,  $\Lambda$  is of dimensions  $pK \times D$ , and an unrotated solution is obtain by setting loadings to zero in  $\Lambda$  so that it is a lower triangular matrix. It can be shown that rotation preserves the identifiability constraints (Revuelta, 2014). That is, if the item parameters follow exclusion or deviation constraints, the same type of constraints will hold after rotation.

### **Exploratory nominal factor analysis of a scale of attitudes towards immigration**

A scale of attitudes towards immigration was applied to a sample of 1,358 individuals as part of a sample survey about social values conducted in 2014 by the Spanish Center for Social Research (Centro de Investigaciones Sociológicas, CIS). The scale contains six items scored in nominal categories, and the number of categories per item varies from 4 to 6. An exploratory factor analysis was conducted to determine the meaning and interpretation of the dimensions



underlying the scale. The English translations of the items and the Mplus code are included in the supplementary materials to this article.

Mplus provides the following goodness of fit indices: the log-likelihood function, Akaike's Information Criterion (AIC: Akaike, 1974), the Bayesian Information Criterion (BIC: Schwarz, 1978), and the sample-size adjusted BIC (BIC<sub>2</sub>, say). Model fit for models including between 1 and 5 factors were compared using the AIC, BIC, BIC<sub>2</sub>, and the likelihood-ratio statistic for pairs of nested models ( $G^2_{dif}$ ; Haberman, 1977; Maydeu-Olivares & Cai, 2006).  $G^2_{dif}$  is computed as -2 times the difference in log-likelihoods between a factor model with  $D$  factors and a model with  $D - 1$  factors. The results are presented in Table 3. The AIC and  $G^2_{dif}$  criteria suggest that five or more factors are present in these data. However, the more conservative BIC statistics suggests two factors, and the BIC<sub>2</sub> (the sample-size corrected BIC) suggests three factors. All in all, we have selected the model with three factors to interpret the data to achieve a balance between fit and parsimony.

-----  
 Insert Table 3 about here  
 -----

The estimated parameters for the three-factor model are shown in Table 4. The column labeled Exclusion Constraints contains the lower triangular matrix  $\Lambda$  rendered by Mplus. The fixed slopes in the upper right corner of  $\Lambda$  have been set to zero in the Mplus code to overcome the rotation indeterminacy. This unrotated exploratory factor analysis solution was transformed to deviation constraints because these items do not have a natural reference category against which the others can be compared, and rotated using Varimax (Browne, 2001). The results appear in the column labeled *Deviation constraints and Varimax* in Table 4.

-----  
 Insert Table 4 about here  
 -----

Factor 1 appears to be a general factor of positive attitudes towards immigration. High factor scores reflect a positive and liberal attitude towards immigration. Factors 2 and 3 are related to some specific categories of items 4 and 6 that have similar wording. Factor 2 loads on the third response alternatives in these items. This alternative involves kind relations between the

local population and immigrants. Factor 3 loads on response alternative 5 of the same items. This alternative makes a reference to the relations between the two populations being ‘normal’.

### **Confirmatory nominal factor analysis. Modeling responses to testlets**

This example illustrates the application of a confirmatory factor model. The data consist of the responses to eight items from a data analysis exam. Each item has three response categories, one of which is correct. The items are clustered in two groups (testlets): items 1 to 5 share a common stem, and items 6 to 8 share another stem. The sample contains the responses from 500 students.

Three models were applied, a one-dimensional model, a bi-factor model with one general factor and two uncorrelated group factors (one per testlet), and a bifactor model with correlated group factors. The correct answer to each item was assigned the highest category so that it is used by Mplus as the reference category. Complete materials for this example are provided in the supplementary materials to this article. Table 5 shows the goodness of fit results for the three models provided. The bi-factor model improves the fit of the one-factor model, thus a single factor does not suffice to capture the associations between the items, but estimating the correlation between the specific (testlet) factors does not improve fit.

-----  
 Insert Table 5 about here  
 -----

Table 6 contains the parameter estimates for the bi-factor model with uncorrelated factors. The negative loadings on the general factor mean that the tendency to select distractors decreases as the factor score (i.e., knowledge of data analysis) increases. The group factors represent knowledge specific for each cluster of items. Items 1 to 5 involve questions about the chi-square statistical test; individuals with a low score in Factor 2 (the first group factor) are more likely to err in those questions. Items 6 to 8 involve questions about the independent samples *t* test; individuals with a low score in Factor 3 (the second group factor) are more likely to err in those questions.

-----  
 Insert Table 6 about here  
 -----

### Alternative parameterizations of the multidimensional NCM in item response theory

Item response theory pays special attention to modeling the probabilistic relation between responses to test items and the underlying latent variables and obtaining meaningful item parameters (van der Linden, 2019). In this vein, Thissen, Cai and Bock (2010) and Falk and Cai (2016) present sophisticated parameterizations of the NCM (henceforth referred to as the TCB parameterization) aimed at obtaining more interpretable parameters. In particular the TCB parametrization provides factor loadings associated with items instead of slopes for the categories within the item, and category scores to uncover the ordering of the categories.

#### The TCB parameterization

The core idea of the TCB parameterization is to decompose the loading of the category into an item loading and a category score. This is achieved by computing the loading of category  $k$  in factor  $d$  as the product  $\lambda_{kd} = a_d s_{kd}$ , where  $a_d$  is an item loading independent of the category and  $s_{kd}$  is the score of category. The item loading  $a_d$  represents the strength of the association between the item and the factor, as in the linear and ordinal factor models. The score of the category indicates the strength of the association between the category and the factor. Constraints are imposed on the scores to avoid redundancies and identify parameters.

The item parameters are a vector of  $D$  item loadings  $\mathbf{a} = (a_1, \dots, a_D)$  and a matrix  $\mathbf{S} = (s_{kd})$  of order  $K \times D$  that contains the scores. The linear term in the probability function of Eq. (8) is reparametrized as follows in the TCB parameterization:

$$\begin{aligned} \mathbf{v}_k + \boldsymbol{\lambda}_k' \boldsymbol{\eta} &= \mathbf{v}_k + \lambda_{k1} \eta_1 + \dots + \lambda_{kD} \eta_D \\ &= \mathbf{v}_k + a_1 s_{k1} \eta_1 + \dots + a_D s_{kD} \eta_D \\ &= \mathbf{v}_k + (\mathbf{a} \circ \mathbf{s}_{k*}) \boldsymbol{\eta} \end{aligned} \quad (10)$$

where  $\mathbf{s}_{k*}$  is the  $k$ -th row vector of  $\mathbf{S}$  (that is,  $\mathbf{s}_{k*}$  contains the scores for category  $k$  in the  $D$  factors) and  $\circ$  is the Hadamard product (a term by term product).

The NCM includes  $(K-1)D$  non-redundant slopes whereas the TCB has  $K$  item slopes and  $K \times D$  scores. Identifiability constraints are necessary to reduce the number of scores to  $(K-2)D$  for the total number of parameters to be the same as in the original parameterization of the NCM. The constraints amount to setting the scores of the first category to 0 for all factors and setting the scores of the last category to  $K-1$  (i.e.,  $s_{1d} = 0$  and  $s_{Kd} = K-1$  for all  $d$ ). The

choice of the first and the last category is arbitrary since they are unordered except in those items in which categories have some structure. For example, in Likert-type items the categories have a given ordering, or in a multiple-choice item the correct option is category  $K$ .

The comparison of item slopes across items requires that all the items have the same format and number of categories because the item slope is inversely proportional to the number of categories. In particular, the relation between the slope of the upper category and the item slope is  $\lambda_{Kd} = a_d(K - 1)$ , and the item slope depends on the number of categories. In those items where the selection of category  $K$  is completely arbitrary, as would be the case for the questionnaire of attitudes towards immigration, the item slope would be arbitrary as well and hardly comparable from one item to another.

### TCB parameterization using a contrast matrix

The TCB parameterization uses a matrix of coefficients,  $\mathbf{T}$ , to compute  $\mathbf{v}$  and  $\mathbf{S}$ . Apart from its mathematical elegance, the use of matrix  $\mathbf{T}$  provides a means to impose constraints in the parameters and to estimate models with varying levels of flexibility between the ordinal and the full rank nominal model. Let  $\mathbf{s}_{*d} = (s_{1d}, \dots, s_{Kd})'$  be the  $d$ -th column of  $\mathbf{S}$  (the elements of  $\mathbf{s}_{*d}$  are the scores of the categories in factor  $d$ ). In a purely ordinal model the scores are equally spaced integer numbers, i.e. for a four categories item,  $\mathbf{s}_{*d} = (0, 1, 2, 3)'$ . In the most general version of the TCB two of the categories have fixed scores,  $\mathbf{s}_{*d} = (0, s_{2d}, s_{3d}, 3)'$ . Intermediate models between these two extremes are generated in the TCB parameterization by using a linear equation to impose constraints:

$$\mathbf{s}_{*d} = \mathbf{T}_d \begin{pmatrix} 1 \\ \boldsymbol{\alpha}_d \end{pmatrix} \quad (11)$$

Where  $\mathbf{T}_d$  is a matrix of fixed coefficients that can be different from one vector  $\mathbf{s}_{*d}$  to another and the symbol  $\boldsymbol{\alpha}_d$  represents a vector of  $K - 2$  parameters.

Thissen, Cai and Bock (2010) and Thissen and Cai (2019) suggested using either a Fourier- or an identity-based  $\mathbf{T}$  matrix. These matrices are given by

$$\mathbf{T}_{Fourier} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & t_{22} & \cdots & t_{2(K-1)} \\ 2 & t_{32} & \cdots & t_{3(K-1)} \\ \vdots & \vdots & & \vdots \\ K-1 & 0 & \cdots & 0 \end{pmatrix} \quad \mathbf{T}_{Identity} = \begin{pmatrix} 0 & \mathbf{0}_{K-2} \\ \mathbf{0}_{K-2} & \mathbf{I}_{K-2} \\ K-1 & 0 \end{pmatrix} \quad (12)$$

The elements of  $\mathbf{T}_{Fourier}$  are given by  $t_{pq} = \sin[\pi(q-1)(p-1)/(K-1)]$ ,  $\mathbf{0}$  is a vector of zeros and  $\mathbf{I}$  is an identity matrix. The contrast matrices in (12) impose the identification constraints  $s_{1d} = 0$  and  $s_{Kd} = K-1$ . The columns of  $\mathbf{T}_{Fourier}$  represent linear, quadratic and higher-order terms. By removing columns from  $\mathbf{T}_{Fourier}$ , models with different levels of flexibility can be estimated and compared. The identity-based  $\mathbf{T}$  matrix imposes no structure on the scores.

For example, the Fourier-based contrast matrix for an item with four categories is (Houts & Cai, 2016, p. 191)

$$\mathbf{T}_{Fourier} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & .866 & .866 \\ 2 & .866 & -.866 \\ 3 & 0 & 0 \end{pmatrix} \quad (13)$$

Note that the columns of  $\mathbf{T}_{Fourier}$  define a linear, quadratic and cubic tendency. Inserting (13) in (11), the scoring of the categories is

$$\mathbf{s}_{*d} = (0 \quad 1 + 0.866\alpha_1 + 0.866\alpha_2 \quad 2 + 0.866\alpha_1 - 0.866\alpha_2 \quad 3)' \quad (14)$$

Several interesting particular cases arise from Eq. (14). The model is equivalent to the generalized partial credit model for ordinal responses when  $\alpha_1 = \alpha_2 = 0$  (Muraki, 1992). If both  $\alpha_1$  and  $\alpha_2$  are free parameters the model is equivalent to Bock's nominal model. When  $\alpha_1$  is free and  $\alpha_2 = 0$  the scores combine a linear and a quadratic term,  $\mathbf{s}_{*d} = (0, 1 + 0.866\alpha_1, 2 + 0.866\alpha_1, 3)'$ , and the model is more flexible than Muraki's model and more constrained than the full rank model.

Similarly to the slopes, the item intercepts in this parameterization are given by  $\mathbf{v} = \mathbf{T}\boldsymbol{\gamma}$ , where  $\boldsymbol{\gamma}$  is a vector of at most  $K-1$  parameters and  $\mathbf{T}$  is a matrix of constant coefficients. Imposing constraints in the intercepts is usually of interest mainly for certain particular cases of the Rasch family of models (Fischer & Molenaar, 1995).

In our experience, the application of the full-rank nominal factor analysis model to ordinal data is seldom justified from a statistical point of view (see also Maydeu-Olivares, 2005). When model parsimony is taken into account, by computing a Root Mean Squared Error of Approximation (RMSEA) for categorical data (Maydeu-Olivares & Joe, 2014) the ordinal factor analysis model most often outperforms the nominal factor analysis model. For this reason, the development of intermediate cases between the nominal and the ordinal case is particularly interesting.

### Estimating the ordering of the categories

This empirical example illustrates how the TCB parameterization of the NCM can be used to estimate the ordering of the categories. The items, data, flexMIRT and Mplus codes are provided in the supplementary materials. An Emotional Stability scale was applied as part of a personnel selection process for traffic air controllers in a high stakes testing situation. The test includes questions related to cognitive, emotional and physiological symptoms of instability. We have analyzed the responses of 1,000 participants to 10 items of emotional instability scored in four ordinal categories (Completely Disagree, Disagree, Agree, and Completely Agree) using a one-factor model. Responses were coded in reversed order for Mplus to set the parameters of the Completely Disagree category to zero. The TCB is estimated in Mplus by programming the following equations in the MODEL CONSTRAINT section of the code to obtain  $a_1$ ,  $\alpha_1$  and  $\alpha_2$  from the category slopes:

$$\begin{aligned}\lambda_1 &= a_1 3 \\ \lambda_2 &= a_1 (2 + 0.866\alpha_1 - 0.866\alpha_2) \\ \lambda_3 &= a_1 (1 + 0.866\alpha_1 + 0.866\alpha_2) \\ \lambda_4 &= 0\end{aligned}\tag{15}$$

Goodness of fit of the models is given in the first three rows of Table 7. The likelihood ratio chi-square and the AIC statistics support the full rank model, whereas the BIC indicates that the second model is the most appropriate one.

-----  
Insert Table 7 about here  
-----

The estimated parameters for the full rank model appear in Table 8. Results in this table are in natural order and the fourth category is Completely Agree. The category scores are far

from being ordered because the two intermediate categories have a too small score. This suggests that responses tend to concentrate in the two extreme categories, an effect which is analyzed in the next section.

-----  
 Insert Table 8 about here  
 -----

### Using the multidimensional NCM to separate substantive and response style factors

A response style factor has no substantive interpretation in relation to the construct measured by the test but determines the strategies followed by the individual to select categories. For examples, individuals that follow the extreme response style (ERS; Bolt & Newton, 2011) exhibit a tendency to select extreme categories (e.g. Strongly Disagree and Strongly Agree) on an attitude scale irrespective of the question posed. Two factors will emerge from data contaminated with ERS, a substantive factor and a factor representing response style. Apart from defining intermediate models between the ordinal and nominal cases, Falk and Cai (2016) used matrix  $\mathbf{T}$  to implement scoring rules that represent several response styles commonly found in Likert-type scales.

A score vector that represents ERS is  $\mathbf{s}_{*d} = (0 \ -1 \ -1 \ 0)'$ . Because the middle categories have negative scores, the responses will concentrate on the extreme category as the level in the ERS factor increases. The following matrix of scores represents a two-dimensional model with a substantive and a ERS factor. Data are coded in reverse order for category Completely Disagree to have a 0 score:

$$\mathbf{S} = \begin{pmatrix} 3 & 0 \\ 2 + 0.866\alpha_1 + 0.866\alpha_2 & -1 \\ 1 + 0.866\alpha_1 & -1 \\ 0 & 0 \end{pmatrix} \quad (16)$$

The columns of  $\mathbf{S}$  are the scores in the substantive and the ERS factors. From Eq. (10), the relation between the category slopes and the TCB parameters is

$$\begin{aligned} \lambda_{11} &= 3a_1 & \lambda_{12} &= 0 \\ \lambda_{21} &= a_1(2 + 0.866\alpha_1 - 0.866\alpha_2) & \lambda_{22} &= -a_2 \\ \lambda_{31} &= a_1(1 + 0.866\alpha_1 + 0.866\alpha_2) & \lambda_{32} &= -a_2 \\ \lambda_{41} &= 0 & \lambda_{42} &= 0 \end{aligned} \quad (17)$$

Three models were estimated with different constraints in the scores of the substantive factor: parameters  $\alpha_1$  and  $\alpha_2$  are free in Model 1, Model 2 imposes the constraint  $\alpha_2 = 0$ ; in Model 2 both  $\alpha_1$  and  $\alpha_2$  are set to 0. The factors were correlated in all three models. Table 7 summarizes the goodness of fit results for this example when an ERS factor is added to the model. The BIC statistic supports Model 2, whereas the other statistics support Model 1.

Estimated intercepts and slopes ( $\nu$  and  $\lambda$ ) for Model 1 are shown in Table 9. The slopes in the substantive factor keep the ordering of the categories for most of the items, although the spacing of the slopes does not follow a smooth sequence. The slopes for the ERS factor are significantly different from zero; as a result, one can conclude that there is ERS in these data. Because the ERS slopes are negative; this means that individuals with a high score on the ERS factor are more likely to select extreme categories.

-----  
Insert Tables 9 about here  
-----

Table 10 contains the parameters  $\alpha_1$  and  $\alpha_2$ , and the item slopes. The high standard errors for the alpha parameters suggests that the model over fits the data of this example. On the other hand, all the factor loadings for ERS reach statistical significance are well above 1, supporting the need for an ERS factor.

-----  
Insert Table 10 about here  
-----

Figure 1 contains the item response probabilities for Item 8 as a function of the emotional stability factor. We chose to display this item because of its large slope. The upper, middle and lower panels correspond to ERS values of -1.5, 0, and 1.5, respectively. The figure illustrates the transition between the preferences for middle categories when the factor score is low, to the preference for extreme categories for high factor scores.

-----  
Insert Figure 1 about here  
-----



### **Discussion and concluding remarks**

In this article, we have shown how factor analysis can be performed on first choice data using Mplus. The model implemented in Mplus is well known in the IRT literature (Bock's nominal categories model, NCM) but may be unknown to many applied factor analysis users. Just as factor analysis can be described as multivariate regression with unobserved predictors, nominal factor analysis can be described as multivariate multinomial logistic regression with unobserved predictors. Interestingly, the parameterization of this model implemented in Mplus corresponds to this description.

Because of the prolix parameterization of the NCM, several constrained versions have been proposed to facilitate interpretability. The MODEL CONSTRAINT command of Mplus can be used to estimate the NCM under linear (Revuelta, 2008, 2010) and nonlinear constraints (Falk & Cai, 2016). Any constrained version of the NCM can be estimated in Mplus. Examples include two-parameter logistic model (Birnbaum, 1968), the generalized partial credit model (Huggins-Manley & Algina, 2015; Muraki, 1992), and Masters' (1982) partial credit model, which are described in the Mplus documentation. In this manuscript, we describe how to estimate in Mplus the full model originally proposed by Bock (1972), including confirmatory and exploratory multidimensional versions, as well as the parameterization described in Thissen, Cai and Bock (Thissen et al., 2010; see also Falk & Cai, 2016; Thissen & Cai, 2016).

The term factor analysis for nominal data is misleading. It is a model for first choice data. In other words, the model assumes that there exists an ordering among response alternatives. However, this ordering is not reported by the respondent, as in raking data, but it is estimated from the data. In the nominal factor analysis model, factors have ordinal properties because a monotonic relation between the factors and the log-odds is assumed, albeit the strength of this relation (the category slope) is estimated instead of being given beforehand. If the data were purely nominal neither an estimated ordering of the categories nor a monotonic relation between category responses and factors could be postulated. Nominal factor scores have the same properties as factor ordinal factor analysis scores except for the origin of the scale. Factor analysis of nominal data assumes that respondents make comparative judgments among the response alternatives within an item. Comparative judgements provide information about whether a response alternative is preferred over another, but the information about how much each alternative is preferred is lost. As a result, the origin of the scale, unlike in ordinal factor

analysis, is not determined. Several methods have been devised to set the origin of the scale when comparative judgements are used (Böckenholt, 2004), the simplest one consisting in modeling simultaneously one rating scale item.

The nominal factor analysis model is rooted in Thurstone's seminal ideas on choice modeling (Bock, 1997; Maydeu-Olivares & Böckenholt, 2009). As a result, it is closely related to the factor analysis model for ranking and partial ranking data and its reparameterization, the Thurstonian IRT model (Brown & Maydeu-Olivares, 2011, 2012; Maydeu-Olivares & Böckenholt, 2005). Consider the assertiveness questionnaire described earlier. In this application, individuals were asked to indicate their most likely response (their first choice) in each of the situations. Bock's model was designed for this kind of data. However, individuals could have been asked instead to provide a full ranking of the alternatives for each situation. Or they could have been asked to provide both their most likely and least likely response (a partial ranking). The suitable model in this case is the factor analysis model for (possibly partial) ranking data (Brown & Maydeu-Olivares, 2011, 2012; Maydeu-Olivares & Böckenholt, 2005). First choice data is a special case of ranking data. Because first choice data contains less information than rankings, to reach the same precision of measurement of individuals' scores, one should expect that more items are needed when applying the nominal factor model than when applying a factor model for rankings.

Unlike the ordinal factor analysis model, which can be estimated from polychoric correlations, the nominal factor model is estimated by maximum likelihood assuming the data is multinomial. This requires integrating over the latent traits for each observed response pattern (Bock & Aitkin, 1981; Muthén & Asparouhov, 2012). As a result, estimating a factor model for nominal data with many factors becomes computationally very demanding because the number of quadrature points used to approximate the integral has an exponential growth rate in relation to the number of factors. The number of quadrature points can be reduced using the GAUSSHERMITE option of the ANALYSIS Mplus command. Schilling and Bock (2005) suggested using as few as two points for each factor for models up to seven factors. For models with more than three factors, we feel that alternative algorithms and computer programs shall be used. The most advanced current algorithm for high dimensional models is the Metropolis-Hastings Robbins-Monro algorithm (Cai, 2010b, 2010a; Cai & Thissen, 2015) implemented in the flexMIRT computer program (Cai, 2015).

In closing, the most popular application of nominal factor analysis to date has been to fit multiple-choice items in educational settings. In these applications not only the correct response is modeled, but also the incorrect responses. Not all incorrect responses are equally incorrect; some provide more information about the competency of the respondent than others. As a result, the use of nominal factor analysis over a binary (correct/incorrect) scoring leads to higher precision in estimating the respondents' competency of interest and provides information about the relation of incorrect categories with the latent trait. Alternatively, for a fixed precision of measurement fewer items need to be administered if scored polytomously using nominal factor analysis than if scored dichotomously. However, we feel that a major application of the model is to fit data obtained using situational questionnaires such as the assertiveness example we have provided. Situational tests hold the promise to obtain performance-based assessments of non-cognitive attributes. We feel they would be a welcome departure from current across-situations across-time self-reported assessments of personality, attitudes, etc. Yet, situational items are likely to be weakly related to the attributes being measured. This means that many items will be needed to accurately assess the attributes of interest. When in doubt as to how many items to use, the standard error of the estimated factor scores provides the most intuitive measure of the precision of measurement. We expect and look forward to applications of the nominal factor analysis to situational tests.

### References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29–51. <https://doi.org/10.1007/BF02291411>
- Bock, R. D. (1997). The nominal categories model. In W.J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 33–49). New York: Springer.

- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443–459.  
<https://doi.org/10.1007/BF02293801>
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431–444.  
<https://doi.org/10.1177/014662168200600405>
- Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, 71(5), 814–833.  
<https://doi.org/10.1177/0013164410388411>
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71(3), 460–502.  
<https://doi.org/10.1177/0013164410375112>
- Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, 44(4), 1135–1147.  
<https://doi.org/10.3758/s13428-012-0217-x>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1), 111–150.  
[https://doi.org/10.1207/S15327906MBR3601\\_05](https://doi.org/10.1207/S15327906MBR3601_05)
- Cai, L. (2010a). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75(4), 581–612. <https://doi.org/10.1007/s11336-010-9178-0>
- Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35(3), 307–335.  
<https://doi.org/10.3102/1076998609353115>
- Cai, L. (2015). flexMIRT: A numerical engine for multilevel item factor analysis and test scoring [Computer program]. Chapel Hill, NC: Vector Psychometric Group.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). IRTPRO: Flexible, multidimensional, multiple categorical IRT modelling [Computer software]. Chicago, IL: Scientific Software International.
- Cai, L., & Thissen, D. (2015). Modern approaches to parameter estimation in item response

- theory. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to typical performance assessment* (pp. 41–59). New York, NY: Routledge.
- Campion, M. C., & Ployhart, R. E. (2012). Assessing personality with situational judgment measures: Interactionist Psychology Operationalized.  
<https://doi.org/10.1017/CBO9781107415324.004>
- Cudeck, R., & MacCallum, R. C. (2007). *Factor analysis at 100: Historical developments and future directions. Factor analysis at 100 years*. Mahwah, NJ: Lawrence Erlbaum Associates. [https://doi.org/10.1207/S15327906MBR3803\\_06](https://doi.org/10.1207/S15327906MBR3803_06)
- De Ayala, R. J., & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*, 23(1), 3–19.  
<https://doi.org/10.1177/01466219922031130>
- Falk, C. F., & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods*, 21(3), 328–347. <https://doi.org/10.1037/met0000059>
- Ferrando, P. J., & Lorenzo-Seva, U. (2016). A note on improving EAP trait estimation in oblique factor-analytic and item response theory models. *Psicológica*, 37, 235–247. Retrieved from <https://www.uv.es/revispsi/articulos2.16/7Ferrando.pdf>
- Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments and applications*. New York: Springer.
- Haberman, S. J. (1977). Log-linear models and frequency tables with small expected cell counts. *Annals of Statistics*, 5, 1148–1169.
- Houts, C. R., & Cai, L. (2016). *flexMIRT: Flexible multilevel item factor analysis and test scoring. User's manual*. Seattle, WA: Vector Psychometric Group.
- Huggins-Manley, A. C., & Algina, J. (2015). The partial credit model and generalized partial credit model as constrained nominal response models, with applications in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(January), 1–11.  
<https://doi.org/10.1080/10705511.2014.937374>
- Johnson, T. R., & Bolt, D. M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style. *Journal of Educational and Behavioral Statistics*, 35(1), 92–114.  
<https://doi.org/10.3102/1076998609340529>

- Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36(3), 347–387. Retrieved from <http://www.tandfonline.com/doi/abs/10.1207/S15327906347-387>
- Jöreskog, K. G., & Moustaki, I. (2006). Factor analysis of ordinal variables with full information maximum likelihood. Retrieved from <http://www.ssicentral.com/lisrel/techdocs/orfiml.pdf>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Maydeu-Olivares, A. (2005). Further empirical results on parametric versus non-parametric IRT modeling of Likert-type personality data. *Multivariate Behavioral Research*, 40(2), 261–279. [https://doi.org/10.1207/s15327906mbr4002\\_5](https://doi.org/10.1207/s15327906mbr4002_5)
- Maydeu-Olivares, A., & Böckenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods*, 10(3), 285–304. <https://doi.org/10.1037/1082-989X.10.3.285>
- Maydeu-Olivares, A., & Böckenholt, U. (2009). Modeling preference data. In A. Millsap, R. Maydeu-Olivares (Ed.), *Handbook of Quantitative Methods in Psychology* (pp. 264–282). London: Sage.
- Maydeu-Olivares, A., & Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioral Research*, 45(6), 935–974. <https://doi.org/10.1080/00273171.2010.531231>
- Maydeu-Olivares, A., & Cai, L. (2006). A cautionary note on using G2 (dif) to assess relative model fit in categorical data analysis. *Multivariate Behavioral Research*, 41(1), 55–64. [https://doi.org/10.1207/s15327906mbr4101\\_4](https://doi.org/10.1207/s15327906mbr4101_4)
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4), 305–328. <https://doi.org/10.1080/00273171.2014.911075>
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- McDonald, R. P. (1997). Normal ogive multidimensional model. In Wim J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257–269). New York: Springer.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176. <https://doi.org/10.1177/014662169201600206>

- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335.  
<https://doi.org/10.1037/a0026802>
- Muthén, L. K., & Muthén, B. O. (2016). Mplus. Version 7.2 [Computer program]. Los Angeles, CA: Muthén & Muthén.
- Ployhart, R. E., & Ward, A. (2013). Situational judgment measures. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. (pp. 551–564). Washington: American Psychological Association. <https://doi.org/10.1037/14047-030>
- Preston, K., Reise, S., Cai, L., & Hays, R. D. (2011). Using the nominal response model to evaluate response category discrimination in the PROMIS emotional distress item pools. *Educational and Psychological Measurement*, 71(3), 523–550.  
<https://doi.org/10.1177/0013164410382250>
- Revuelta, J. (2008). The generalized logit-linear item response model for binary-designed items. *Psychometrika*, 73(3), 385–405. <https://doi.org/10.1007/s11336-007-9048-6>
- Revuelta, J. (2010). Estimating difficulty from polytomous categorical data. *Psychometrika*, 75(2), 331–350. <https://doi.org/10.1007/s11336-009-9145-9>
- Revuelta, J. (2014). Multidimensional item response model for nominal variables. *Applied Psychological Measurement*, 38(7), 549–562. <https://doi.org/10.1177/0146621614536272>
- Revuelta, J., & Ximénez, C. (2017). Bayesian dimensionality assessment for the multidimensional nominal response model. *Frontiers in Psychology*, 8, 961.  
<https://doi.org/10.3389/fpsyg.2017.00961>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No. 17*.
- Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70(3), 533–555.  
<https://doi.org/10.1007/s11336-003-1141-x>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>

- Smithson, M., & Merkle, E. C. (2013). *Generalized linear models for categorical and continuous limited dependent variables*. New York: Chapman and Hall.
- St-Sauveur, C., Girouard, S., & Goyette, V. (2014). Use of situational judgment tests in personnel selection: Are the different methods for scoring the response options equivalent? *International Journal of Selection and Assessment*, 22(3), 225–239.  
<https://doi.org/10.1111/ijsa.12072>
- 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., & Cai, L. (2016). Nominal categories model. In Wim J. van der Linden (Ed.), *Handbook of Item Response Theory Modeling, Volume One: Models* (pp. 54–73). Boca Raton, FL: CRC Press.
- Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories item response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of Polytomous Item Response Theory Models* (pp. 43–75). New York: Routledge.
- Thissen, D., Chen, W. H., & Bock, R. D. (2003). Multilog (version 7) [Computer program]. Lincolnwood, IL: Scientific Software International.
- Thissen, D., Steinberg, L., & Fitzpatrick, A. R. (1989). Multiple-choice models: The distractors are also part of the item. *Journal of Educational Measurement*, 26(2), 161–176.  
<https://doi.org/10.1111/j.1745-3984.1989.tb00326.x>
- Whetzel, D. L., & McDaniel, M. A. (2016). Are situational judgment tests the best way to measure personality in high stakes testing? *The Wiley Handbook of Personality Assessment*, 205–214.
- Wollack, J. A., Bolt, D. M., Cohen, A. S., & Lee, Y.-S. (2002). Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, 26(3), 339–352. <https://doi.org/10.1177/0146621602026003007>



Table 1

*Parameter estimates for the situational assertiveness questionnaire*

<i>item</i>	<i>Stem / Response option</i>	<i>Intercept</i>	<i>Loading</i>
1	You're denied entry to a concert		
	A-C Ask to talk to supervisor vs. quit	<b>2.53</b> (.19)	<b>.75</b> (.24)
	B-C Call parents for suggestions vs. quit	-.04 (.23)	<b>.82</b> (.35)
2	Your transportation home fails after an event		
	A-C Call parents for a ride vs. walk home	<b>-.52</b> (.08)	.12 (.16)
	B-C Ride with acquaintances vs. walk home	<b>-1.07</b> (.10)	.18 (.18)
3	Boss asks you for overtime, had dinner plans with friends		
	A-C Don't go to dinner vs. decline overtime	<b>-.41</b> (.14)	-.30 (.22)
	B-C Arrive late to dinner vs. decline overtime	<b>.90</b> (.09)	<b>.39</b> (.19)
4	Parents ask you to be home early		
	A-C Obey parents vs. ignore parents' orders	.67 (.46)	.34 (.45)
	B-C Ty to convince parents vs. ignore parents' orders	<b>3.82</b> (.40)	<b>1.14</b> (.38)
5	Parents insist on a choice of major		
	A-C Go along with parents' suggestion vs. seek for relatives' support	<b>-1.08</b> (.50)	<b>-1.11</b> (.50)
	B-C Persuade parents major is inappropriate vs. seek for relatives' support	<b>2.94</b> (.17)	.20 (.33)
6	You and your friends are choosing a restaurant		
	A-C Push for your choice vs. suggest but not push	<b>-3.53</b> (.24)	.04 (.49)
	B-C Listen to others' proposals vs. suggest but not push	<b>-2.26</b> (.20)	<b>-1.03</b> (.26)
7	Received a lower grade than expected at school		
	A-C Complain to the instructor vs. resolve to work harder	<b>-2.35</b> (.15)	-.52 (.28)
	B-C Do nothing vs. resolve to work harder	<b>-3.20</b> (.33)	<b>-1.23</b> (.34)
8	Friends suggest excessive money for a present		
	A-C Go along with the group vs. set amount you contribute	.20 (.11)	<b>-.61</b> (.24)
	B-C Argue amount is excessive vs. set amount you contribute	<b>.41</b> (.10)	.34 (.23)

*Note:* This is the standard Mplus output; it uses exclusion constraints. Standard errors in parentheses.

Table 2

*Parameter estimates for the situational assertiveness questionnaire using deviation constraints*

<i>item</i>	<i>Response option</i>	<i>Intercept</i>	<i>Loading</i>
1	You're denied entry to a concert		
	A Ask to talk to supervisor	<b>1.70</b> (.08)	.23 (.13)
	B Call parents for suggestions	<b>-.87</b> (.11)	.29 (.20)
	C Quit	<b>-.83</b> (.13)	<b>-.52</b> (.18)
2	Your transportation home fails after an event		
	A Call parents for a ride	.01 (.05)	.02 (.11)
	B Ride with acquaintances	<b>-.54</b> (.06)	.08 (.12)
	C Walk home	<b>.53</b> (.05)	-.10 (.09)
3	Boss asks you for overtime, had dinner plans with friends		
	A Don't go to dinner	<b>-.57</b> (.08)	<b>-.33</b> (.13)
	B Arrive late to dinner	<b>.74</b> (.05)	<b>.36</b> (.11)
	C Decline to do overtime this time	<b>-.16</b> (.07)	-.03 (.12)
4	Parents ask you to be home early		
	A Obey parents	<b>-.83</b> (.21)	-.16 (.22)
	B Try to convince parents	<b>2.32</b> (.16)	<b>.65</b> (.18)
	C Ignore parents' orders	<b>-1.50</b> (.27)	-.49 (.26)
5	Parents insist on a choice of major		
	A Go along with parents' suggestion	<b>-1.71</b> (.33)	<b>-.81</b> (.29)
	B Persuade parents that major is inappropriate	<b>2.33</b> (.17)	<b>.51</b> (.19)
	C Seek relatives' support to persuade parents	<b>-.62</b> (.20)	.30 (.25)
6	You and your friends are choosing a restaurant		
	A Push for your choice	<b>-1.60</b> (.17)	.37 (.34)
	B Listen to others' proposals	<b>-.32</b> (.15)	<b>-.70</b> (.24)
	C Suggest but not push	<b>1.92</b> (.11)	<b>.33</b> (.18)
7	Received a lower grade than expected at school		
	A Complain to the instructor	<b>-.50</b> (.15)	.07 (.19)
	B Do nothing	<b>-1.35</b> (.22)	<b>-.65</b> (.22)
	C Resolve to work harder	<b>1.85</b> (.13)	<b>.58</b> (.16)
8	Friends suggest excessive money for a present		
	A Go along with the group	.00 (.06)	<b>-.52</b> (.14)
	B Argue amount is excessive	<b>.21</b> (.06)	<b>.43</b> (.14)
	C Set amount you contribute	<b>-.21</b> (.06)	.09 (.13)

*Note:* This is the result of using Bock's original parameterization. Standard errors in parentheses

Table 3

*Goodness of fit statistics for the exploratory factor analysis of the scale of attitude towards immigration*

Factors	<i>Pars</i>	<i>Log likelihood</i>	<i>AIC</i>	<i>BIC</i>	<i>BIC<sub>2</sub></i>	$G^2_{dif}$	$df_{dif}$	<i>p-value</i>
1	52	-11495.84	23,095.67	23,366.79	23,201.61			
2	77	-11400.05	22,954.10	23,355.56	23,110.96	191.58	25	<.001
3	101	-11346.09	22,894.17	23,420.76	23,099.93	107.93	24	<.001
4	124	-11320.62	22,889.23	23,535.74	23,141.84	50.94	23	<.001
5	146	-11291.18	22,874.36	23,635.57	23,171.79	58.87	22	<.001

*Note:* pars = number of parameters;  $G^2_{dif}$  is the likelihood ratio statistic to test each model against the model in the row above with degrees of freedom  $df_{dif}$ .  $BIC_2$  is the sample-size adjusted BIC given by Mplus.

Table 4

*Estimated parameters for the scale of attitudes towards immigration. Three-factor model*

			Exclusion constraints			Deviation constraints rotated using Varimax		
Item	Category	Intercept	Slope 1	Slope 2	Slope 3	Slope 1	Slope 2	Slope 3
1	1	0.46	5.36	<b>0</b>	<b>0</b>	2.49	-0.44	-0.02
	2	2.05	3.73	0.27	<b>0</b>	0.91	0.03	0.07
	3	2.78	3.35	0.10	0.11	-0.48	0.02	0.13
	4	-0.16	2.73	0.15	-0.25	-0.09	0.12	-0.20
	5	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	-2.82	0.27	0.01
2	1	0.29	3.97	0.39	0.37	2.11	-0.22	0.00
	2	2.45	2.29	0.51	0.51	0.44	0.07	0.19
	3	2.00	1.20	0.43	0.60	-0.66	0.11	0.25
	4	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	-1.88	0.04	-0.44
3	1	3.73	-4.48	-0.78	0.98	-2.43	0.06	0.25
	2	4.61	-2.89	-0.79	1.07	-0.86	-0.19	0.32
	3	3.59	-1.94	-0.93	0.87	0.08	-0.38	0.09
	4	3.02	-1.02	-0.51	0.49	1.06	0.01	-0.17
	5	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	2.15	0.49	-0.49
4	1	3.55	-0.23	-0.29	1.59	0.09	0.10	-0.46
	2	3.19	0.12	-0.04	1.03	0.48	0.46	-0.92
	3	3.64	-0.76	-1.56	3.01	-0.64	-1.45	0.53
	4	4.50	-0.11	0.30	1.17	0.30	0.77	-0.69
	5	2.64	-0.86	0.01	5.57	-0.64	-0.68	3.44
	6	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	0.41	0.81	-1.90
5	1	1.70	-0.11	-0.39	0.21	-0.16	-0.17	-0.05
	2	1.60	-0.11	-0.34	0.44	-0.17	-0.20	0.19
	3	1.61	0.39	-0.22	0.14	0.35	-0.06	-0.07
	4	1.00	0.23	-0.34	0.31	0.17	-0.20	0.06
	5	0.32	-0.19	0.16	0.11	-0.17	0.39	0.01
	6	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	-0.00	0.24	-0.14
6	1	2.14	1.06	-0.67	0.40	-0.03	-0.03	-0.17
	2	1.63	0.41	-0.79	0.19	-0.69	0.01	-0.40
	3	1.74	2.21	-2.18	0.49	0.91	-1.63	-0.53
	4	3.25	1.27	-0.34	0.46	0.21	0.25	-0.01
	5	2.01	1.60	0.42	1.77	0.59	0.54	1.46
	6	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	-0.99	0.87	-0.35

*Note:* fixed parameters are boldfaced. Mplus outputs the exclusion constraints parameters. The estimated slopes were transformed to deviation constraints and then rotated using Varimax.

Table 5

*Goodness of fit statistics for the data analysis examination*

Model	<i>pars</i>	<i>Log likelihood</i>	<i>AIC</i>	<i>BIC</i>	<i>BIC<sub>2</sub></i>	$G^2_{dif}$	$df_{dif}$	<i>p-value</i>
1 Factor	32	-3,133.69	6,331.38	6,466.25	6,364.68			
Bi-factor Uncorrelated factors	48	-3,016.10	6,128.21	6,330.51	6,178.15	235.18	16	<0.01
Bi-factor Correlated factors	51	-3,015.74	6,133.48	6,348.43	6,186.555	0.72	3	0.87

*Note:*  $G^2_{dif}$  is the likelihood ratio statistic to test each model against the model in the row above with degrees of freedom  $df_{dif}$ .  $BIC_2$  is the sample-size adjusted BIC.

Table 6

*Parameter estimates for the data analysis examination example, bi-factor model*

<i>item</i>	<i>Intercept</i>	<i>Slopes</i>		
1	-2.23 (.67)	-1.89 (.62)	2.00 (.89)	<b>0</b>
	-2.82 (.36)	-2.14 (.45)	-0.50 (.46)	<b>0</b>
2	-2.44 (.32)	-2.57 (.41)	-0.99 (.32)	<b>0</b>
	-2.28 (.23)	-1.39 (.35)	0.19 (.26)	<b>0</b>
3	-2.67 (.48)	-1.90 (.59)	-2.95 (.78)	<b>0</b>
	-2.36 (.44)	-2.45 (.56)	-2.59 (.78)	<b>0</b>
4	-0.58 (.14)	1.12 (.23)	-0.17 (.18)	<b>0</b>
	-2.09 (.27)	-1.25 (.27)	-0.43 (.21)	<b>0</b>
5	-0.32 (.13)	-0.43 (.18)	-0.36 (.21)	<b>0</b>
	-2.72 (.66)	-2.58 (.62)	-2.78 (.67)	<b>0</b>
6	-2.88 (.74)	-2.97 (.99)	<b>0</b>	-3.34 (1.25)
	-2.62 (.69)	-2.90 (.96)	<b>0</b>	-2.88 (1.18)
7	-1.92 (.19)	-0.90 (.27)	<b>0</b>	-0.41 (0.29)
	-2.52 (.56)	-2.28 (.50)	<b>0</b>	-2.48 (0.67)
8	-0.58 (.83)	-1.97 (.52)	<b>0</b>	-3.02 (0.93)
	3.03 (.70)	0.20 (.35)	<b>0</b>	-2.68 (0.93)

*Note:* fixed parameters are boldfaced. Standard errors in parentheses

Table 7

*Goodness of fit statistics for the emotional stability test*

ERS	Constraints	<i>Pars.</i>	<i>Log likelihood</i>	<i>AIC</i>	<i>BIC</i>	$G^2_{dif}$	<i>df</i>	<i>p-value</i>
No	$\alpha_I$ and $\alpha_2$ free	60	-8,231.60	16,583.19	16,877.65			
	$\alpha_I$ free, $\alpha_2 = 0$	50	-8,245.74	16,591.48	16,836.87	28.28	10	.002
	$\alpha_I = 0, \alpha_2 = 0$	40	-8,389.50	16,859.00	17,055.31	287.52	10	<.001
Yes	$\alpha_I$ and $\alpha_2$ free	71	-8,032.04	16,206.08	16,554.53			
	$\alpha_I$ free, $\alpha_2 = 0$	61	-8,059.54	16,241.09	16,540.46	55.00	10	<.001
	$\alpha_I = 0, \alpha_2 = 0$	51	-8,078.88	16,259.76	16,510.06	38.68	10	<.001

*Note:* ERS indicates whether the model includes and Extreme Response Style factor or not  $G^2_{dif}$  is the likelihood ratio statistic to test each model against the model in the row above with degrees of freedom *df*.  $BIC_2$  is the sample-size adjusted BIC.

Table 8

*Estimated TCB parameterization for the emotional stability test*

	Intercepts					Scores			
Item	Slope	$v_1$	$v_2$	$v_3$	$v_4$	$s_{11}$	$s_{21}$	$s_{31}$	$s_{14}$
1	0.02	0	-0.11	3.78	4.25	0	-94.79	-66.29	3
2	0.12	0	1.69	4.32	3.37	0	-15.24	-9.45	3
3	0.61	0	2.11	4.95	3.48	0	-1.02	0.30	3
4	0.13	0	1.34	2.82	3.43	0	-9.41	-8.56	3
5	0.51	0	1.28	4.00	3.03	0	-1.95	-0.62	3
6	0.44	0	1.33	3.57	2.40	0	-2.89	-1.77	3
7	0.43	0	1.41	1.83	-0.13	0	-0.60	0.12	3
8	0.88	0	2.55	3.94	1.95	0	0.17	0.92	3
9	0.64	0	1.57	3.39	1.80	0	-0.16	0.61	3
10	0.01	0	0.17	4.21	5.00	0	-165.76	-157.56	3

*Note:* The intercept and the slope for the first category structural zeros. The score for category four is fixed to 3.



Table 9

*Intercepts and category slopes for the emotional stability test*

Item	Category	Slopes		
		Intercept	Factor ES	Factor ERS
1	CA	4.40 (0.42)	0.62 (0.48)	<b>0</b>
	A	3.86 (0.43)	0.00 (2.06)	-1.76 (0.36)
	D	-0.67 (0.74)	-1.50 (2.10)	-1.76 (0.36)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
2	CA	4.48 (.70)	1.81 (0.55)	<b>0</b>
	A	5.38 (.70)	1.11 (1.84)	-1.53 (0.31)
	D	2.63 (.73)	-0.25 (1.85)	-1.53 (0.31)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
3	CA	5.25 (1.11)	2.09 (0.78)	<b>0</b>
	A	6.65 (1.12)	1.49 (2.05)	-1.63 (0.33)
	D	3.20 (1.14)	-0.13 (2.05)	-1.63 (0.33)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
4	CA	4.23 (0.46)	1.47 (0.39)	<b>0</b>
	A	3.58 (0.47)	0.80 (1.76)	-1.49 (0.30)
	D	1.72 (0.49)	-0.16 (1.76)	-1.49 (0.30)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
5	CA	4.60 (0.74)	2.10 (0.58)	<b>0</b>
	A	5.48 (0.74)	1.24 (2.03)	-1.69 (0.34)
	D	1.74 (0.81)	-0.66 (2.04)	-1.69 (0.34)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
6	CA	4.10 (0.61)	2.12 (0.49)	<b>0</b>
	A	5.19 (0.62)	1.30 (2.37)	-2.02 (0.41)
	D	2.50 (0.64)	-0.16 (2.37)	-2.02 (0.41)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
7	CA	0.10 (0.25)	1.74 (0.29)	<b>0</b>
	A	2.72 (0.25)	1.38 (1.29)	-1.11 (0.24)
	D	2.33 (0.25)	0.76 (1.29)	-1.11 (0.24)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
8	CA	6.10 (1.35)	3.73 (0.83)	<b>0</b>
	A	7.93 (1.37)	3.36 (2.34)	-1.88 (0.39)
	D	6.37 (1.37)	2.84 (2.32)	-1.88 (0.39)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
9	CA	4.04 (0.68)	2.84 (0.54)	<b>0</b>
	A	5.48 (0.69)	2.31 (1.70)	-1.40 (0.29)
	D	2.90 (0.68)	0.47 (1.68)	-1.40 (0.29)
	CD	<b>0</b>	<b>0</b>	<b>0</b>
10	CA	5.55 (0.81)	1.28 (0.67)	<b>0</b>
	A	4.65 (0.81)	0.16 (2.58)	-1.91 (0.39)
	D	-1.05 (1.30)	-1.56 (2.41)	-1.91 (0.39)
	CD	<b>0</b>	<b>0</b>	<b>0</b>

*Note:* fixed parameters are boldfaced. The scores of categories 2 and 3 in factor ERS are set equal. Items are coded as 1: Completely Agree, 2: Agree, 3: Disagree, 4: Completely Disagree. The estimated correlation between the factors is is -0.16 (1.11).

Table 10

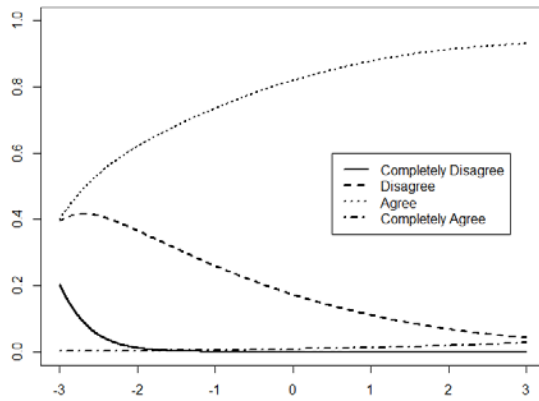
*Item parameters estimates for the emotional stability test*

Item	Parameters alpha		Item slopes	
	$\alpha_1$	$\alpha_2$	$a_1$	$a_2$
1	-5.92 (12.50)	-3.60 (3.32)	0.21 (0.16)	1.76 (0.36)
2	-0.91 (3.45)	-0.72 (0.42)	0.60 (0.18)	1.53 (0.36)
3	-0.60 (3.24)	-0.77 (0.51)	0.70 (0.26)	1.63 (0.33)
4	-0.98 (4.11)	-0.55 (0.35)	0.49 (0.13)	1.49 (0.30)
5	-1.26 (3.31)	-0.99 (0.46)	0.70 (0.19)	1.69 (0.34)
6	-0.94 (3.84)	-0.48 (0.28)	0.71 (0.16)	2.02 (0.41)
7	0.39 (2.53)	-0.04 (0.15)	0.58 (0.10)	1.11 (0.24)
8	0.78 (2.02)	-0.03 (0.14)	1.24 (0.28)	1.88 (0.39)
9	-0.04 (1.99)	-0.54 (0.21)	0.95 (0.18)	1.40 (0.29)
10	-3.63 (6.81)	-1.76 (1.46)	0.43 (0.23)	1.91 (0.39)

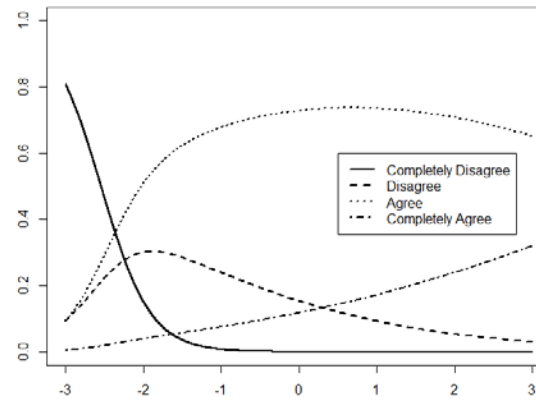
Figure 1.

Option response functions for Item 8 of the emotional stability test. The horizontal axis is the Emotional Stability factor and the vertical axis is the response probability.

*Extreme Response Style = -1.5*



*Extreme Response Style = 0*



*Extreme Response Style = 1.5*

