A Constrained Confirmatory Mixture IRT Model: Extensions and Estimation of the Saltus model using Mplus

Minjeong Jeon [a], ✉

[a]University of California, Los Angeles

Abstract ∎ In this paper, I will discuss applications, extensions, and estimation of the Saltus model, a specialized confirmatory mixture IRT model. The Saltus model is confirmatory because the number/nature of latent classes is pre-specified and prior item information is utilized for class differentiation. Such a confirmatory model has not been fully utilized in applied research due to two main misconceptions: (1) the model is designed for specific purposes only and (2) a specialized software package is required to estimate the model. In this study, I will discuss that (1) such a confirmatory approach is applicable to various applications of mixture IRT modeling, (2) the model can actually be parameterized as a constrained mixture IRT model and (3) the model can be readily extended and estimated with regular, off-the-shelf mixture IRT software packages that allow for linear constraints on model parameters. An application and estimation of the constrained confirmatory IRT model is illustrated with an empirical example.

Keywords ∎ Mixture IRT models; Confirmatory approach; Saltus model; Linear Constraints. **Tools** ∎ Mplus.

✉ mjjeon@ucla.edu

MJ: 0000-0002-5880-4146

10.20982/tqmp.14.2.p120

Introduction

Item response theory (IRT) is a modeling framework widely used to analyze discrete data in various disciplines such as psychology and education. IRT models explain an observed response to a test/survey item as a function of a respondent's latent trait (such as ability or proficiency) and the item's properties (such as difficulty). IRT models typically assume that subjects are drawn from a homogeneous population. However, this assumption is often violated in practice due to known or unknown clustering of subjects. If the subjects' population membership is known, multiple-group IRT models (e. g., B. Muthén & Lehman, 1985) can be used. If the population membership is unknown, finite-mixture, or simply mixture (or mixed) IRT models (e. g., Rost, 1990; von Davier & Yamamoto, 2007) are often utilized.

Mixture IRT models postulate that subjects are drawn from two or more unknown (or latent) populations that present systematic differences in their item response behavior. Latent class IRT models (e. g., Vermunt, 1999; von Davier, 2008) have a similar feature of subject heterogeneity. The key difference lies in the fact that latent class IRT models posit homogeneity (or no variability) of the person latent traits within a latent class; that is, all subjects with the same latent class are assumed to have the same latent trait level. In contrast, mixture IRT models allow for within-class heterogeneity (or variation of person latent traits within a class).

Mixture IRT models have been extended in various ways, e. g., with item covariates (Mislevy & Verhelst, 1990), person covariates (Vermunt & Magidson, 2005), polytomous data (Bolt, Cohen, & Wollack, 2001), multilevel person structures (Vermunt, 2008), and multidimensional test structures (De Boeck, Cho, & Wilson, 2011). Various mixture IRT models have been utilized to examine subjects'

heterogeneity in depression/anxiety (Wanders et al., 2016), extreme response styles (H.-Y. Huang, 2016), at-risk unhealthy behaviors (Finch & Pierson, 2011), and item solving strategies (Mislevy & Verhelst, 1990). Other applications of mixture IRT models also exist, e. g., for studying test speededness (Bolt, Cohen, & Wollack, 2002), models' goodness of fit (Rost, 1990), and differential item functioning (Cho & Cohen, 2010; Cohen & Bolt, 2005). For more information on mixture IRT model and its applications, I refer readers to e. g., Cho, Suh, and Lee (2016) and De Boeck et al. (2011).

Exploratory vs. Confirmatory Mixture IRT Analysis

Mixture IRT models are often used in an exploratory manner, as the number of latent classes is empirically determined based on the data. To choose the optimal number of latent classes, a model selection method is usually utilized; in this case, researchers set up several competing models with varying numbers of latent classes and compare those models' goodness-of-fit statistics (such as Akaike information criterion, AIC; Bayesian information criterion, BIC; and Deviance information criterion, DIC) (e. g., Nylund, Asparouhov, & Muthén, 2007). In addition, a priori hypotheses on the item parameters are usually not assumed across latent classes; hence, the item parameters are freely estimated for each latent class.

Although relatively less common compared to an exploratory approach, mixture IRT models have also been used in a confirmatory fashion. When the number and nature of latent classes are known prior to data analysis, the use of a mixture IRT model is confirmatory. For example, Mislevy and Verhelst (1990) applied a mixture IRT model to identify test takers who adopted a guessing or ability-based strategy to solve test items. Their model includes two latent classes that represent ability-based and guessing strategies, respectively. The ability-based strategy class is based on a regular IRT model, while the guessing strategy class is based on a modified IRT model without a person (ability) parameter (since guessing rather than ability is utilized for item solutions). The item easiness parameter of the guessing strategy class model was set at .25 (which is the reciprocal of the number of response options) so that the probability of giving correct responses for the guessing strategy class could be fixed across test items. A similar line of research has been conducted to investigate examinees' rapid guessing strategy during speeded tests (e. g., Schnipke & Scrams, 1997). Recently, Molenaar, Oberski, Vermunt, and De Boeck (2016) utilized a confirmatory mixture IRT modeling approach, to investigate how examinees utilize different types of intelligence (either slow or fast one) depending on test items. Their model includes two latent classes that represent slow and fast responses, respectively. A two-class mixture model was applied to re-

sponse time data to determine whether a response to each item belongs to a slow or a fast response class. A regular IRT model was then applied to analyze response accuracy data (correct or incorrect) for each of the two (fast and slow response) latent classes. Tijmstra, Bolsinova, and Jeon (2018) adopted a confirmatory mixture modeling approach with two pre-specified latent classes (associated with different response styles) to test whether/how subjects engage in item solutions with different response styles. The authors specified two latent classes where one class follows a general partial credit model (Muraki, 1992) for polytomous item responses, while the other class follows a more complex model where choosing a mid-point response category is associated with an additional latent trait (following an item response tree model (e. g., Boeckenholt, 2012; De Boeck & Partchev, 2012; Jeon & De Boeck, 2016). Jin, Chen, and Wang (2018) utilized a confirmatory mixture IRT approach for differentiating inattentive responses from normal responses in Likert-scale data. They also assumed two latent classes where one class is based on a regular partial credit model, while the other class follows a modified polytomous IRT model that does not include a person ability parameter (so that the class can model 'inattentive' item responses).

A Constrained Confirmatory Mixture IRT Model

Another example of a confirmatory mixture IRT approach is the Saltus model, which was presented in the context of examining children's developmental stages (Mislevy & Wilson, 1996; Wilson, 1989). The modeling idea is to utilize specific sets of test items, which are specially designed to differentiate children's performance (based on their developmental stages) for class differentiation. This approach is confirmatory in two senses: (1) the number/nature of latent classes (i.e., developmental stages) is pre-specified in advance prior to data analysis; and (2) the prior information on test items is utilized for differentiating latent classes of subjects. Some may argue that pre-specifying the number/character of latent classes and using prior item information seem somewhat restrictive. However, similar ideas have often been employed in psychometrics. An important example is confirmatory (item) factor analysis. Confirmatory factor analysis assumes based on prior research: (1) pre-determined number/nature of factors and (2) the membership of items to the factors (or item-factor relationship). This is different from exploratory factor analysis, which assumes no prior knowledge on factor structures before data analysis. Exploratory factor analysis aims to identify an underlying factor structure (or relationship between items and factors), while confirmatory analysis is used to verify a known factor structure and examine the relationship between factors. Both confirmatory

and exploratory factor analyses are extensively utilized in applied research.

Another example is the linear logistic test model (Fischer, 1973), which utilizes item property information as predictors to explain item difficulties in an IRT model. This constrained model imposes a stronger assumption than a regular IRT model, including a smaller number of parameters (regression coefficients for the item characteristics). This seemingly restrictive IRT model has been broadly applied in various areas of research (e. g., Daniel & Embretson, 2010; Kubinger, 2008). Hence, it may be reasonable to utilize prior knowledge on classes and item properties in mixture IRT modeling.

Motivation

The original Saltus model has been extended for polytomous data with a partial credit model (Draney, 2007) and with a linear logistic test model (Draney & Wilson, 2008). However, these models do not seem to have been fully utilized in applied research. One potential reason may stem from the impression that the model is designed solely for examining developmental stages. The Saltus modeling idea, however, can successfully be applied in practice to a variety of scenarios where a researcher would like to test a clearly developed hypothesis on structural differences between latent classes. For instance, suppose there is an English test that contains a set of items that are designed to identify students who would need extra support to improve their English speaking skills. In this case, the Saltus model may be utilized to differentiate such students who need special needs (from students without needing further support) and to evaluate whether the designed test items function properly.

Another misconception that people may have is that a specialized software package is required to estimate the Saltus model. In fact, Draney's (2007) Fortran code appears to be the only available software package for fitting the Saltus model at present. Some researchers have applied a Bayesian approach to estimate the Saltus model (e. g., Draney & Jeon, 2011). In this article, I will show that the Saltus model is actually estimable with mixture IRT models' existing software packages, such as Mplus (L. Muthén & Muthén, 2017).

Aims and Structure

In this paper, I will introduce the Saltus model as a constrained confirmatory mixture IRT model, and describe its application, extensions, and estimation. The specific aims of this paper are threefold: (1) to show that the Saltus model can be parameterized as a confirmatory mixture IRT model with linear constraints on model parameters, (2) to present several useful extensions of the Saltus model, and (3) to provide an empirical illustration of an extended Saltus model with Mplus.

The remainder of this paper is structured as follows: In Section "Models", I will describe the original parameterization of the Saltus model and present its new parameterization as a constrained mixture IRT model. Subsequently, I will present several important extensions of the Saltus model. In Section "Empirical Illustrations", I will present an empirical illustration of the extended Saltus models. The paper ends in the "Discussion" section with a summary and discussion on future developments.

Models

Saltus Model: Original Parameterization

In its original formulation, the Saltus model was specified as an extension of the Rasch model in the context of modeling subjects' cognitive development. Subjects are assumed to be classified into one stage at the time of testing, while stage membership is not directly observed (Wilson, 1989). The number of latent stages is pre-specified and different latent stages are presumably differentiated by the subjects' performance on specific item groups. The differences in the performance between latent classes are quantified with so-called "Saltus parameters".

To specify the model, it is assumed that there are H latent classes and H item groups. Typically, the number of item groups is set equivalent to the number of latent classes because each item group is supposed to represent the performance of a corresponding latent class (Wilson, 1989). The Saltus model can then be formulated as follows:

$$\Pr\left(y_{ijk} = 1 \mid \theta_{jh}\right) = \frac{\exp\left(\theta_{jh} - \beta_i + \sum_h \phi_{jh}\tau_{hk}\right)}{1 + \exp\left(\theta_{jh} - \beta_i + \sum_h \phi_{jh}\tau_{hk}\right)}, \quad (1)$$

where y_{ijk} is a binary response for subject j to item i that belongs to item group k, θ_{jh} is the latent trait of person j in Class h with $\theta_{jh} \sim N\left(\mu_h, \sigma_h^2\right)$, and β_i is the difficulty parameter of item i.[1] Note that the parameter ϕ_{jh} indicates person j's latent group membership, with $\phi_{jh} = 1$ if person j is in group h and 0 otherwise ($h = 1, \ldots, H$). The latent classes are exclusive and exhaustive so that a person is classified into only one of the H latent classes. For model identification, it was assumed that $\sum \beta_i = 0$ (so that μ_h can be freely estimated); in addition, Wilson (1989) applied $\tau_{1k} = 0$ (for all k) and $\tau_{h1} = 0$ (for all h), meaning that the first latent class ($h = 1$) and the first item group

---

[1]The original specification of the Saltus model used θ_j, implicitly assuming that the model is a conditional model given a latent class. For clarification, I replaced θ_j with θ_{jh} in Equations (1) and (2). In addition, I replaced y_{ij} in the original specification with y_{ijk} to clarify that the model specifies the response for person j to item i that belongs to item group k.

$(k = 1)$ are used as the reference latent class and the reference item group, respectively.

The Saltus parameter τ_{hk} represents the effect of item group k on the item responses given by subjects in latent class h. This parameter also indicates the difference in the item difficulty parameters for item group k between a reference latent class and focal latent classes (latent classes that are not the reference latent class). For instance, suppose $H = 3$. Then the linear predictor of Equation (1) is $\theta_{j1} - \beta_i + \phi_{j1}\tau_{11}$ (when $h = 1$), $\theta_{j2} - \beta_i + \phi_{j2}\tau_{22}$ (when $h = 2$), $\theta_{j3} - \beta_i + \phi_{j3}\tau_{33}$ (when $h = 3$) (since $h = k$ is implied in the original Saltus model specification). Because $\tau_{11} = 0$ is assumed for identification, the Saltus parameters to be estimated are τ_{22} and τ_{33} and they indicate the amounts of advantage (or disadvantage) that subjects in Class 2 ($h = 2$) and subjects in Class 3 ($h = 3$) have in solving the second item group ($k = 2$) and the third item group ($k = 3$), respectively, compared to subjects in Class 1 (reference latent class). Note that the Saltus parameter is conceptually similar to the differential feature functioning (DIF; e. g., Ackerman, 1992; Holland & Wainer, 1993) parameter for the item groups between different latent classes. In a standard DIF analysis, a DIF parameter represents a difference in an item parameter (e. g., difficulty parameter) between two manifest groups, such as male and female students (after controlling for respondents' ability levels). A Saltus parameter represents a difference in the difficulty levels of an item group between two latent classes (e. g., subjects in different developmental stages). In that sense, the concept of the Saltus and DIF parameters may be seen to be similar (although technical differences do exist).

A more general formulation of the Saltus model was provided by Mislevy and Wilson (1996):

$$\Pr(y_{ijk} = 1 \mid \theta_{jh}) =$$
$$\prod_h \prod_k \left[\frac{\exp(\theta_{jh} - \beta_i + \sum_h \phi_{jh}\tau_{hk})}{1 + \exp(\theta_{jh} - \beta_i + \sum_h \phi_{jh}\tau_{hk})} \right]^{\phi_{jh}b_{ik}}, \quad (2)$$

where the new variable b_{ik} is an indicator variable that denotes whether or not item i belongs to item group k; that is, $b_{ik} = 1$ if item i belongs to item group k and $b_{ik} = 0$ if otherwise. The main difference of the new formulation (2) from the original formulation (1) is that the Saltus parameter τ_{hk} is now represented with a full matrix (that is, $h = k$ is no longer assumed). For instance, when $H = 3$, the 3×3 full Saltus parameter matrix is specified as

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{bmatrix}. \quad (3)$$

The original Saltus model, Equation (1), includes only the diagonal elements of this matrix as Saltus parameters. For

example, when $h = 2$, three Saltus parameter terms are included in the general model (i.e., $\theta_{j2} - \beta_i + \phi_{j2}\tau_{21} + \phi_{j2}\tau_{22} + \phi_{j2}\tau_{23}$), while the original model includes only one Saltus parameter term (i.e., $\theta_{j2} - \beta_i + \phi_{j2}\tau_{22}$).

For identification of the general Saltus model, the same set of constraints on the Saltus parameters is imposed as the original Saltus model (Equation 1). That is, the first latent class (or one of other latent classes) and the first item group (or one of other item groups) are used as the reference groups. That is, $\tau_{1k} = \tau_{h1} = 0$ for all k and all h ($k = 1, \ldots, H$ and $h = 1, \ldots, H$). As discussed earlier, the Saltus parameter matrix (3) for the general model includes two off-diagonal elements (τ_{23} and τ_{32}), whereas the original model (Equation 1) contains only the diagonal elements of the Saltus parameter matrix. This means that the general Saltus model allows for Class 2 and Class 3 subjects to present a different performance level on item group 3 and item group 2, respectively. Hereinafter, all discussions will be based on the general Saltus model.

Re-parameterization

There has been little discussion in the literature on the relationship between the Saltus model and regular mixture IRT models. In addition, based on the formulations of the original and general Saltus model discussed above, it is difficult to understand how the Saltus model is related to the regular mixture IRT model. Here, it will be shown that the Saltus model can actually be re-parameterized as a constrained regular mixture IRT model with a pre-specified number of latent classes.

To this end, let us first specify a regular, one-parameter logistic (1PL) mixture IRT model with H latent classes for a binary response as follows:
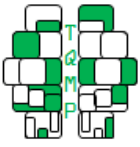
$$\text{logit } (\Pr(y_{ij} = 1 \mid \theta_{jh}, C_j = h)) = \theta_{jh} - \beta_{ih}, \quad (4)$$

where C_j is a categorical latent variable ($C_j = 1, ..., H$) that indicates person j's class membership and θ_{jh} is a continuous latent variable that indicates person j's latent trait in Class h with $\theta_{jh} \sim N(\mu_h, \sigma_h^2)$. To remove the indeterminacy of the scale and identify the model, the 1PL mixture IRT model usually assumes that the means of the latent trait distribution are fixed at 0 (i.e., $\mu_h = 0$ for all h), so that the class-specific item difficulty parameters can be freely estimated. The item parameter β_{ih} represents the difficulty of item i in latent class h.

Equation (4) is a conditional model given a person's latent class. A marginal probability model can be formulated across all possible class membership as follows:

$$\Pr(y_{ij} = 1 \mid \theta_j) = \sum_{h=1}^{H} \pi_h \Pr(y_{ij} = 1 \mid \theta_{jh}, C_j = h),$$

where π_h indicates the probability of belonging to Class h

or a proportion of latent class h. The parameter π_h is also referred to as a mixing proportion (with $\sum_{h=1}^{H} \pi_h = 1$).

Now let us re-formulate the 1PL general Saltus model specified in Equation (2) by using similar notation to Equation (4) as

$$\text{logit}\left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh}, C_j = h\right)\right) = \theta_{jh} - \beta_i + \tau_{hk}. \quad (5)$$

In this formulation, the third term τ_{hk} implies $\phi_{jh}\tau_{hk}$ (that is, this term appears only if person j belongs to Class h). In addition, an additional subscript k appears in Equation (5). Since item i's group membership is known a priori, $\beta_{ikh}^* = \beta_{ih}^* = \beta_i - \tau_{hk}$. Thus, the following can be obtained

$$\text{logit}\ \left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh},\ C_j = h\right)\right) = \ \theta_{jh} - \ \beta_{ih}^*. \quad (6)$$

Note that the reformulation of the general Saltus model with Equation (6) appears equivalent to the formulation of the 1PL mixture IRT model in Equation (4) (with the replacement of β_{ih} with β_{ih}^*). Note, however, that there is an important difference in the latent trait distribution assumption: the Saltus model assumes that the class-specific latent trait θ_{jh} follows a respective normal distribution, $\theta_{jh} \sim N\left(\mu_h,\ \sigma_h^2\right)$, where the mean ($\mu_h$) of the distribution is fixed at zero only for one reference latent class (e. g., $h = 1$), while the latent trait means for other classes are freely estimated. In comparison, the regular 1PL mixture model specified in Equation (4) requires that the latent trait means are fixed at 0 for all latent classes.

As discussed earlier, the Saltus parameter indicates the difference in the difficulty level of a particular item group between latent classes. To see this more clearly, suppose $H = 2$ (that is, there are two latent classes and two item groups). From Equation (5), the Saltus model for latent class 1 ($h = 1$) can be written as follows:

$$\text{logit}\left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh},\ C_j = 1\right)\right) = \ \theta_{j1} - \underbrace{\beta_i + \ \tau_{11} + \ \tau_{12}}_{=\beta_{i1}^*}. \quad (7)$$

Here $\beta_{i1}^* = \beta_i - \ \tau_{11} - \ \tau_{12}$ (therefore, $-\beta_{i1}^* = -\beta_i + \ \tau_{11} + \ \tau_{12}$) is defined as in Equation (6). Note that two Saltus parameter terms for both item groups 1 and 2 ($k = 1, 2$) appear in Equation (7), following the general Saltus model specification (see Saltus parameter matrix of Equation 3). Similarly, the model for latent class 2 ($h = 2$) can be specified as follows:

$$\text{logit}\ \left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh},\ C_j = 2\right)\right) = \ \theta_{j2} - \underbrace{\beta_i + \ \tau_{21} + \tau_{22}}_{=\beta_{i2}^*} \quad (8)$$

where $\beta_{i2}^* = \beta_i - \tau_{21} - \tau_{22}$ (therefore, $-\beta_{i2}^* = -\beta_i + \tau_{21} + \tau_{22}$) is defined.

Since $\tau_{11} = \ \tau_{12} = 0$ is set for identification, $\beta_{i1}^* = \beta_i$ when $h = 1$; similarly, since $\tau_{21} = 0$ for identification, $\beta_{i2}^* = \beta_i - \ \tau_{22}$ with $h = 2$. By replacing β_i with β_{i1}^* in the latter, $\tau_{22} = \beta_{i1}^* - \beta_{i2}^*$ is obtained. This means that the Saltus parameter (τ_{22}) can be defined as the difference between the item difficulty parameters for latent class 1 (β_{i1}^*) and for latent class 2 (β_{i1}^*) when item i belongs to item group 2 ($k = 2$).

Now let us consider a more complex scenario with three latent classes ($H = 3$; that is, there are three latent classes and three item groups). In this case, three class-specific models can be specified for latent classes 1, 2 and 3. For latent class 1 ($h = 1$), the model can be written as

$$\text{logit}\ \left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh},\ C_j = 1\right)\right) = \ \theta_{j1} - \underbrace{\beta_i + \ \tau_{11} + \tau_{12} + \ \tau_{13}}_{\beta_{i1}^*} \quad (9)$$

As in Equation (7) and (8), the Saltus parameter terms for all three item groups ($k = 1, 2, 3$) are specified in Equation (9). Similarly, the models for latent class 2 ($h = 2$) and ($h = 3$) can be specified as follows:

$$\text{logit}\ \left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh},\ C_j = 2\right)\right) = \ \theta_{j2} - \underbrace{\beta_i + \tau_{21} + \tau_{22} + \tau_{23}}_{\beta_{i2}^*}, \quad (10)$$

$$\text{logit}\ \left(\Pr\left(y_{ijk} = 1 \mid \theta_{jh},\ C_j = 3\right)\right) = \ \theta_{j3} - \underbrace{\beta_i + \tau_{31} + \tau_{32} + \tau_{33}}_{\beta_{i3}^*}. \quad (11)$$

For Equations (9) to (11), it is defined that $\beta_{i1}^* = \beta_i - \tau_{11} - \tau_{12} - \ \tau_{13}$, $\beta_{i2}^* = \beta_i - \ \tau_{21} - \tau_{22} - \tau_{23}$, and $\beta_{i3}^* = \beta_i - \ \tau_{31} - \tau_{32} - \tau_{33}$. Since $\tau_{11} = \tau_{12} = \ \tau_{13} = 0$ and $\tau_{21} = \ \tau_{31} = 0$ are set for model identification, $\beta_{i1}^* = \beta_i$ when $h = 1$, $\beta_{i2}^* = \beta_i - \ \tau_{22} - \tau_{23}$ when $h = 2$, and $\beta_{i3}^* = \beta_i - \tau_{32} - \tau_{33}$ when $h = 3$. Accordingly, $\tau_{22} = \beta_{i[2]1}^* - \beta_{i[2]2}^*$, when item i belongs to item group 2 and $\tau_{23} = \beta_{i[3]1}^* - \beta_{i[3]2}^*$, when item i belongs to item group 3.[2] Similarly, $\tau_{33} = \beta_{i[3]1}^* - \beta_{i[3]3}^*$, when item i belongs to item group 3, and $\tau_{32} = \beta_{i[2]1}^* - \beta_{i[2]3}^*$, when item i belongs to item group 2.

From this exercise, it is clear that the general Saltus model (2) can be formulated as a confirmatory mixture IRT model with a set of linear constraints on the item parameters. That is, instead of freely estimating item difficulty parameters β_{ih} for latent class h as in a regular mixture

---

[2] Here subscript $[k]$ in $\beta_{i[k]h}^*$ is used to differentiate the i-th item's group membership ($k = 1, \ldots, H$).

IRT model (see Equation 4), the Saltus model estimates β_{ih}^* that implies a set of constraints, which are defined as, for instance, $\beta_{i1}^* = \beta_i$ when $h = 1$ and $\beta_{i2}^* = \beta_{i1}^* - \tau_{22}$ when $h = 2$ for a two-class model. This means that in the two-class Saltus model, the item difficulty is set to equal between the two classes ($\beta_{i1}^* = \beta_i$) if item i belong to item group 1; however, if item i belongs to item group 2, its difficulty is as large (or small) as τ_{22} in latent class 2 compared to latent class 1. This means that the Saltus model can be estimated with mixture IRT model software packages that allow for constraints on item parameters.

Important Extensions

In the previous section it has been shown that the Saltus model can be formulated as a constrained confirmatory mixture IRT model. This discovery is crucial because it implies that the Saltus model can be extended in a variety of ways and that these extended Saltus models can be estimated without having to develop separate estimation routines or software packages.

In this section, three important extensions of the Saltus model will be discussed with (1) item discrimination parameters, (2) person predictors, and (3) ordinal response data. Note that these extensions are not exhaustive and other extensions are possible, for instance with additional item guessing parameters.

Item Discrimination Parameters. The original Saltus model, in its both simpler and general forms, is based on the Rasch model. However, it is possible to extend the Saltus model with the item discrimination (or loading, slope) parameters. Such a two-parameter model extension is useful because applying a one-parameter model to two-parameter data can lead to identifying spurious latent classes (Alexeev, Templin, & Cohen, 2011). In addition, a researcher may want to test his/her hypothesis that relationships between items and the latent trait of interest (or items' discriminating power) may differ between latent classes.

A two-parameter logistic (2PL) Saltus model can be specified as follows

$$\text{logit} \left(\Pr \left(y_{ijk} = 1 \, | \theta_{jh}, \, C_j = h \right) \right)$$
$$= \alpha_{ih} \theta_{jh} - \beta_i + \tau_{hk}^{(\beta)}, \quad (12)$$

where α_{ih} is the discrimination/loading/slope parameter for item i in Class h. The Saltus parameter $\tau_{hk}^{(\beta)}$ represents a systematic difference in the item difficulty parameters for item group k between subjects in latent class 1 and latent class h.[3] With the inclusion of the class-specific item discrimination parameters (α_{ih}), it is assumed that

$\theta_{jh} \sim N \left(\mu_h, \, \sigma_h^2 \right)$, where μ_h = 0 for one reference latent class (e. g., $h = 1$) and σ_h = 1 for all h (so that all α_{ih} can be freely estimated in all latent classes). To further simplify the model, the item discrimination parameters can be set to be equal across latent classes by replacing α_{ih} with α_i.

Alternatively, if a structured difference can be postulated in the items' discriminating power for a particular item group between latent classes, the Saltus parameters can be introduced for the item discrimination parameters as well as the item intercept parameters. With two types of Saltus parameters, Equation (12) can be modified as

$$\text{logit} \left(\Pr \left(y_{ijk} = 1 \, | \theta_{jh}, \, C_j = h \right) \right)$$
$$= \left(\alpha_i + \tau_{hk}^{(\alpha)} \right) \theta_{jh} - \beta_i + \tau_{hk}^{(\beta)}, \quad (13)$$

where $\tau_{hk}^{(\alpha)}$ represents the Saltus parameter for α_i, which indicates a systematic difference in the discrimination power of item group k between subjects in latent class h and latent class 1 (reference latent class).

Similar to $\tau_{hk}^{(\beta)}$, new Saltus parameter $\tau_{hk}^{(\alpha)}$ can be specified by imposing a set of constraints on the class-specific item discrimination parameters. By defining $\alpha_{ih}^* = \alpha_i + \tau_{hk}^{(\alpha)}$) (with $\tau_{11}^{(\alpha)} = \tau_{12}^{(\alpha)} = \tau_{21}^{(\alpha)} = 0$ for identification), it is obtained that $\tau_{22}^{(\alpha)} = \alpha_{i2}^* - \alpha_{i1}^*$ (when $h = 2$). This means that the item discrimination parameters for latent class 2 (α_{i2}^*) is assumed to be greater or smaller by $\tau_{22}^{(\alpha)}$ than the item discrimination parameters for latent class 1 (α_{i1}^*) when item i belongs to item group 2.

Person Predictors. For ordinary mixture IRT analysis, it is often recommended to incorporate person covariates in the model to predict subjects' class membership. It has been reported that including person predictors helps to prevent the possible misclassification of latent classes (G.-H. Huang & Bandeen-Roche, 2004) and to obtain more precise model parameter estimates and standard errors (Smit, Kelderman, & van der Flier, 2000). For similar reasons, it can be useful to consider person predictors in the Saltus model.

Suppose Q person covariates are utilized to predict the probability of subject j' class membership C_j. The class membership is a latent variable that takes discrete values ($C_j = 1, \ldots, H$); hence, a multinomial logit model can be utilized as follows:

$$\Pr \left(C_j = h \right) = \frac{\exp \left(\gamma_{h0} + \sum_q^Q \gamma_{hq} W_{jq} \right)}{\sum_{u=1}^H \exp \left(\gamma_{h0} + \sum_q^Q \gamma_{hq} W_{jq} \right)}, \quad (14)$$

where γ_{h0} is the intercept and γ_{hq} is the regression coefficient of the q-th covariate $W_{jq}(q = 1, \ldots, Q)$ for person j

---

[3] Here $\tau_{hk}^{(\beta)}$ is equivalent to τ_{hk} in the 1PL Saltus model. For the 2PL Saltus model, I additionally used superscript (β) to distinguish two types of Saltus parameters for the item difficulty and discrimination parameters.

in latent class h.

Ordinal Response Data. Next, an extension of the Saltus model for ordinal response data is introduced based on the graded response model (Samejima, 1969).[4] For items with $M + 1$ response categories, a 2PL graded respond Saltus model can be specified as follows:

$$\text{logit } (\Pr (y_{ijk} \geq m \,|\theta_{jh}, \, C_j = h))$$
$$= \left(\alpha_i + \tau_{hk}^{(\alpha)}\right)\theta_{jh} - \delta_{im} + \tau_{hk}^{(\delta)}, \quad (15)$$

where m indicates a response category ($m = 1, \ldots, M$) and δ_{im} represents the threshold parameters for item i's category m (with category 0 as the reference category). The Saltus parameter $\tau_{hk}^{(\alpha)}$ for the item discrimination parameters are defined similar to Equation (13). The Saltus parameter for the threshold parameter, $\tau_{hk}^{(\delta)}$ can be defined as $\tau_{hk}^{(\delta)} = \delta_{im1}^* - \delta_{imh}^*$, where δ_{imh}^* is the threshold parameter for the m-th category of item i (that belongs to item group k) for the subjects in latent class h (when $h = 1$ is the reference latent class). For the sake of simplicity, $\tau_{hk}^{(\delta)}$ is assumed to be equal across all response categories m (except for the reference category) although this assumption can be relaxed by replacing $\tau_{hk}^{(\delta)}$ with $\tau_{hkm}^{(\delta)}$. In the latter case, M Saltus parameters are estimated for the item thresholds (except for the reference category).

Empirical Illustrations

Data

To illustrate the estimation of the Saltus model and its extensions, a verbal aggression dataset (De Boeck & Wilson, 2004; Vansteelandt, 2000) is utilized. My intention for using this well-known dataset is to show that the confirmatory models discussed in this paper can be used for standard psychological research. In addition, this dataset has been frequently utilized to illustrate a variety of newly developed IRT models or procedures (e. g., Braeken, Tuerlinckx, & De Boeck, 2007; De Boeck & Wilson, 2004; Jeon & Rijmen, 2016), including various versions of mixture IRT models (e. g., Cho et al., 2016; Choi & Wilson, 2015). Hence, researchers who are interested in new methodology developments in IRT may already be familiar with this dataset and can easily follow the illustrations provided in this paper.

In brief, the data contain responses on 24 verbal aggression items obtained from 243 female and 73 male first-year psychology students. The test items are designed to measure the source of verbal aggression and its inhibition, where each item presents a scenario that reflects a frus-

trating situation ('Bus', 'Train', 'Store', and 'Operator'), a situation type ('Other-to-blame' and 'Self-to-blame'), a type of verbally aggressive behaviors ('Curse', 'Scold', and 'Shout'), and a behavioral mode ('Want' and 'Do'). For instance, "A bus fails to stop for me. I would want to curse." is a 'Want' item, whereas and "A bus fails to stop for me. I would curse" is a 'Do' item. Respondents are asked to decide whether they agree to give an aggressive verbal response in each scenario and choose one of the three response options ('No', 'Perhaps', and 'Yes') and they are coded 0, 1, and 2, respectively.

The Saltus model is a confirmatory model that is designed to verify a researcher's hypothesis about classes and class differentiation. A specialized Saltus model is constructed for this dataset based on the following rationale and hypotheses: First, the levels of the subjects' angry behaviors elicited by frustrating situations may differ across subjects (Vansteelandt & Van Mechelen, 2004). This means that the subjects may come from different sub-populations (or latent classes) that are characterized by a different order of the situations (in terms of the amount of anger they evoke). Second, there are likely substantial individual differences in the way people express their anger (Vansteelandt & Van Mechelen, 2004). Behavior types, especially, 'Do' vs 'Want' behavior items are likely to create large individual differences in the item thresholds (and other item features) because the 'Do' behavior implies a higher risk of causing actual damage to the given situation compared to the 'Want' behavior. Hence, the 'Do' items are utilized to differentiate the levels and types of verbal aggression that subjects manifest. Specifically, it is assumed that two sub-populations (or latent classes) of subjects are differentiable based on the thresholds and discrimination power of the 'Do' (verses 'Want') behavior items. Third, researchers have consistently reported gender differences in verbal aggression in the literature (e. g., Archer, 2004). Thus, it is hypothesized that there may be gender differences in subjects' verbal aggression level; especially in the way the subjects react to the 'Do' behavior items.

Based on this reasoning and hypotheses, a 2PL ordinal Saltus model (Equation 15) is constructed for the verbal aggression data with male as a covariate to predict class memberships (Equation 14). Further, this model enables us to demonstrate how a combination of the three extensions described in the "Important Extensions" Section can be readily specified and estimated with a standard software package.

---

[4]In this article, I focus on the ordinal response Saltus model, but the specification of a partial credit version for polytomous item responses may also be possible with the latest Mplus version 8 that allows for an adjacent logit link function.
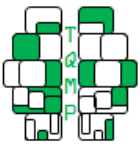
Estimation

As discussed earlier, the Saltus model and its various extensions can be estimated with software packages that permit the estimation of mixture IRT models and linear constraints on model parameters. Here, I use Mplus version 7.4 (L. Muthén & Muthén, 2017) for the full information maximum likelihood (ML) estimation of the discussed, extended Saltus model. Below are the details of Mplus specification for the 2PL ordinal Saltus model with a covariate for the verbal aggression data. The full Mplus syntax for this model is presented in the Appendix.

The initial part of the Mplus syntax includes four commands: (1) `TITLE:`, (2) `DATA:`, (3) `VARIABLE:`, and (4) `ANALYSIS:`. First, the optional command `TITLE` simply specifies a title that will appear on the Mplus output file. The `DATA` command specifies the name of the dataset to be analyzed (`FILE = verbal_pol.dat`). The data are in a space delimited format (with a space in between each column) where a row indicates each subject and a column indicates each variable (the dataset `verbal_pol.dat` includes 324 rows (for 324 subjects) and 25 columns (for 24 items and one male indicator variable). The `VARIABLE` command specifies the names and characteristics of the variables in the dataset. The `NAME` option specifies that the example dataset includes a person covariate (male) and 24 item dummy variables (u1-u24). The `CATEGORICAL` option indicates that u1-u24 are categorical response variables (scored as 0, 1, and 2). The `MISSING` option indicates that missing responses are coded as 99 in the example dataset. The `CLASSES` option specifies the label and number of latent classes. Hence, `c(2)` denotes that `c` is the label for two latent classes. Lastly, the `ANALYSIS` command is used to specify options for estimation. The `TYPE` option indicates the type of analysis to be performed. To estimate models that include latent classes, `TYPE = MIXTURE` is needed. The `ALGORITHM` option specifies the estimation algorithm to be used. `ALGORITHM = INTEGRATION` denotes that full information maximum likelihood estimation with numerical integration is adopted (with the default number of integration points, 15). The `STARTS` option indicates the number of random starts to be used for the analysis. `STARTS = 50 10` tells Mplus to use 50 sets of starting values in the first step of the optimization and to pick 10 starting values that show the largest log likelihood values (in the first optimization step) for the second step of the optimization process. Multiple random starting values are utilized to ensure the ML solutions are not obtained at local maxima of the log-likelihood function.

The next part of the Mplus syntax is the `MODEL` command that specifies the specific model that is estimated (see Listing 2). The heading `%OVERALL%` indicates that the text that follows defines the general model that applies both latent classes. `f1 BY u1-u24*` indicates that 24 item responses (`u1-u24`) are loaded on the latent variable labeled as `f1`. The asterisk at the end of `u1-u24` denotes that all item loading (slope) parameters are freely estimated (instead of the loading of one item is fixed at 1). `f1@1s` indicates that the variance of the latent variable `f1` is fixed at 1. The last term `c ON male` indicates that the male covariate is regressed on the latent class c (that is, male is used as a covariate for the multinomial regression to predict class memberships). Next, headings `%c#1%` and `%c#2%` indicate that the text that follows defines the model specifically for Class 1 and Class 2, respectively. Under the `%c#1%` heading , `f1 BY u1-u24* (x1-x24)` indicates that 24 item responses (`u1-u24`) are loaded on the `f1` latent variable where `x1-x24` in the parenthesis denote the labels for the 24 item loading parameters. The next two terms, `[u1$1-u24$1](a1-a24)` and `[u1$2-u24$2](ax1-ax24)` are used to specify the first and second threshold parameters for the 24 items with labels `a1-a24` and `ax1-ax24`, respectively. `[f1@0]` is used to indicate that the mean of the latent trait `f1` is fixed at 0 for Class 1 (which means that Class 1 is the reference latent class). Under the `%c#2%` heading, the model for Class 2 is specified to be similar to the model for Class 1. The differences are found in the labels for the item loading and threshold parameters. `f1 BY u1-u24* (x1-x12 y13-y24)` means that the loading parameters for the first 12 items (`x1-x12`) are set to be equal to the loading parameters for Class 1 (since the same labels `x1-x12` are used in both latent classes), while the loading parameters for the next 12 items (`y13-y24`) are allowed to be different in Class 2. This means that two item groups are used in this example (first 12 items ('Want') and last 12 items ('Do') and the second set of items ('Do' items) are used to differentiate performance between Class 1 and Class 2 subjects. The last term `[f1]` indicates that the mean of the latent trait distribution for Class 2 is freely estimated.

The next part of the Mplus syntax is the `MODEL CONSTRAINT` command, which is necessary for specifying the Saltus model as a constrained mixture IRT model (see Listing 3). The first line `NEW(tau_d tau_a)` is used to create two new variables labeled for the Saltus parameters (for item threshold parameters and item loading parameters). The next 24 lines (`tau_d = a13-b13;` to `tau_d = a24-b24;` and `tau_d = ax13-bx13;` to `tau_d = ax24-bx24;`) are used to tell Mplus that the difference in the item threshold parameters (for Categories 1 and 2) between Class 1 and Class 2 (`tau_d`) is constant across items 13 to 24 ('Do' items). That is, `tau_d` refers to $\tau_{22}^{(\delta)}$ (that is, it is assumed that all response categories (except the reference category) have the same threshold Saltus parame-
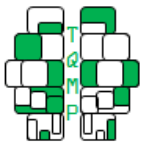
Listing 1 ■ Initial part of the MPlus syntax

```
----------------------------------------------------------------------
!! Header of input file
TITLE: 2PL ordinal Saltus model for verbal aggression data

!! Data file specification
DATA: FILE = verbal_pol.dat;

!! Define variables and specify number of latent classes
VARIABLE:
NAMES = male u1-u24; ! male is dummy variable for male subjects
CATEGORICAL = u1-u24; ! ordinal item responses (0,1,2)
MISSING = ALL(99); ! missing data are coded as 99
CLASSES = c (2); ! number of latent classes

!! Estimation settings
ANALYSIS: TYPE = MIXTURE; ! estimate finite mixture model
ALGORITHM = INTEGRATION; ! use integration method with 15 default quadrature points
STARTS = 50 10 ; ! use multiple random starts (can be increased if needed)
----------------------------------------------------------------------
```

ter). The last 12 lines (`tau_a = y13-x13;` to `tau_a = y24-x24;`) are used to instruct Mplus that the difference in the item loading parameters between Class 1 and Class 2 (`tau_a`) is constant across items 13 to 24 ('Do' items). That is, `tau_a` refers to $\tau_{22}^{(\alpha)}$.

The final part of the Mplus syntax includes the `SAVEDATA` command (see Listing 4). The SAVE option tells Mplus to save, as an external file, the subjects' posterior probability (`cprob`) in each latent class. The FILE option indicates the name of the external file to be saved (`prob_pol.txt`). The external file will be saved to the folder where the syntax file is located.

Once the full syntax input file has been constructed, the input file (`saltus_pol.inp`) can be submitted to Mplus by using the RUN button found on the top of the user interface. Once Mplus terminates successfully, the output will appear in a text file in the folder where the input file is located, with the same name as the input file, ending with the extension .out (`saltus_pol.out`).

Results

Here the estimated results are described for the 2PL ordinal Saltus model obtained from Mplus. Recall that the model was specified for the following purposes: (1) to confirm that there are two groups of subjects who substantially differ in the way that they express their anger; (2) to evaluate the hypothesis that items on the behavior types ('Do' vs. 'Want' items) can differentiate the two groups of subjects; specifically, to check whether there are differ-

ences in the two item parameter (loading and threshold) values for the 'Do' items between the two latent classes; and (3) to test the hypothesis that there is a gender difference in subjects' latent class membership (e. g., to see whether male and female students have different probabilities of belonging to the subject groups who show a different anger level to the 'Do' type items). Note that these questions, which can be investigated with the proposed Saltus model, may be unanswerable with regular mixture IRT models or with other existing approaches.

Mplus output contains voluminous information, but here I focus on the output sections that contain the most important information to interpret key Saltus model results. First, the Mplus output section for the final class counts and estimated proportions reports the number of subjects in each latent class. There are three pieces of information provided (based on the estimated model, based on estimated posterior probabilities, and based on the most likely class). The information based on the estimated model was used (that provides the final class counts and estimated proportions based on the most likely model parameter estimates). The result shows that approximately 31.3% of the subjects were classified into Class 1, while 68.7% of the subjects into Class 2. This result suggests that two groups of subjects were well differentiated as hypothesized, based on the estimated model. However, to understand the nature of the two latent classes, it is necessary to further examine the estimates of other model parameters.

The two Saltus parameters are useful for this purpose.

Listing 2 ■ Second part of the MPlus syntax

```
--------------------------------------------------------------------------------
!! Model specification
MODEL:
! Overall model
%OVERALL%
f1 BY u1-u24*; ! free loading parameters
f1@1; ! factor variance is fixed at 1
! Multinomial regression of class membership with male variable
c ON male; ! male variable is used to predict subjects' class membership

! Model for Class 1
%c#1%
f1 BY u1-u24* (x1-x24); ! loading parameters
[u1$1-u24$1](a1-a24); ! threshold parameters for response category 1
[u1$2-u24$2](ax1-ax24); ! threshold parameters for response category 2
f1@1; ! factor variance is fixed at 1
[f1@0] ; ! factor mean fixed at 0 (reference latent class)

! Model for Class 2
%c#2%
! Use different loading parameter labels for Do items (u13-u24)
f1 BY u1-u24* (x1-x12 y13-y24);
! Use different threshold parameter labels for 'Do' items (u13-u24)
[u1$1-u24$1](a1-a12 b13-b24); ! for response category 1
[u1$2-u24$2](ax1-ax12 bx13-bx24); ! for response category 2
f1@1; ! factor variance is fixed at 1
[f1]; ! factor mean freely estimated
--------------------------------------------------------------------------------
```
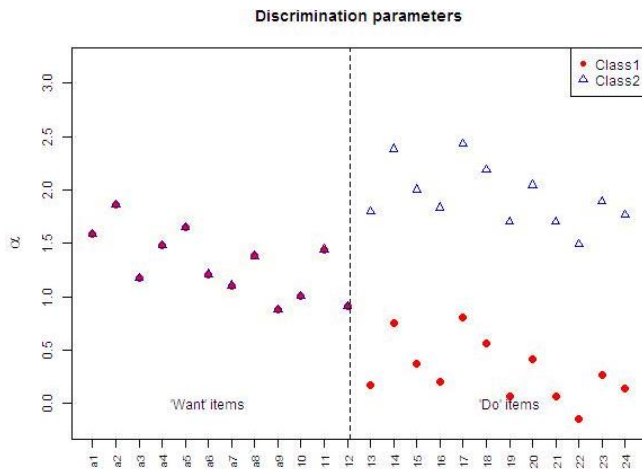
Figure 1 ■ Class-specific estimated discrimination (α_i) parameter values (`a1` to `a24`) and threshold (δ_{im}) parameter values (`d11` to `d242`) for the 2PL ordinal Saltus model for the verbal aggression data.

(a)   **(b)**
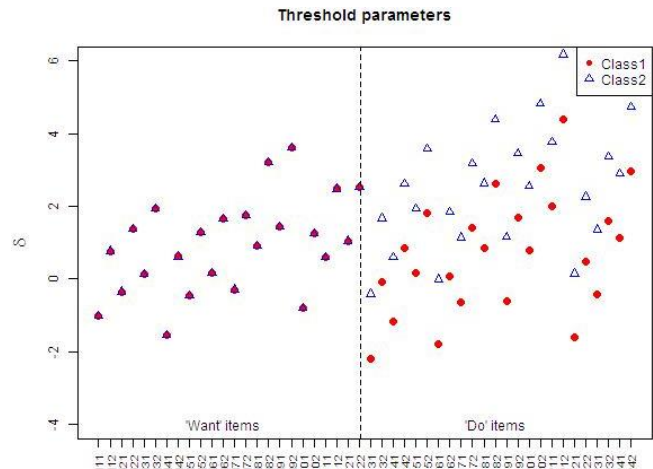
Listing 3 ■ Third part of the MPlus syntax

```
---------------------------------------------------------------------------
! Set model constraints
MODEL CONSTRAINT:
NEW(tau_d tau_a); ! define Saltus parameters
!! Define Saltus parameter for threshold parameters as difference
!! in threshold parameters for 'Do' items between Class 1 and Class 2
! Category 1
tau_d = a13-b13;
tau_d = a14-b14;
tau_d = a15-b15;
tau_d = a16-b16;
tau_d = a17-b17;
tau_d = a18-b18;
tau_d = a19-b19;
tau_d = a20-b20;
tau_d = a21-b21;
tau_d = a22-b22;
tau_d = a23-b23;
tau_d = a24-b24;

! For Category 2
tau_d = ax13-bx13;
tau_d = ax14-bx14;
tau_d = ax15-bx15;
tau_d = ax16-bx16;
tau_d = ax17-bx17;
tau_d = ax18-bx18;
tau_d = ax19-bx19;
tau_d = ax20-bx20;
tau_d = ax21-bx21;
tau_d = ax22-bx22;
tau_d = ax23-bx23;
tau_d = ax24-bx24;

!! Define Saltus parameter for discrimination parameters as difference
!!in discrimination parameters for Do items between Class 2 and Class1
tau_a = y13-x13;
tau_a = y14-x14;
tau_a = y15-x15;
tau_a = y16-x16;
tau_a = y17-x17;
tau_a = y18-x18;
tau_a = y19-x19;
tau_a = y20-x20;
tau_a = y21-x21;
tau_a = y22-x22;
tau_a = y23-x23;
tau_a = y24-x24;
---------------------------------------------------------------------------
```

Listing 4 ■ Last part of the MPlus syntax

```
--------------------------------------------------------------------------
!! Save posterior probability in each class
SAVEDATA:
SAVE is cprob;
FILE is prob_pol.txt ;
--------------------------------------------------------------------------
```

The Saltus parameter estimates are given at the Mplus output section with New/Additional parameters. TAU_D and TAU_A are the labels created for two Saltus parameters ($\tau_{22}^{(\delta)}$ and $\tau_{22}^{(\alpha)}$). Specifically, the Saltus parameter for the item threshold parameters ($\tau_{22}^{(\delta)}$) was negative and significant at the 5% level with an estimate of -1.770 (SE=0.387, p=0.000). Recall that $\tau_{22}^{(\delta)} = \delta_{im1}^* - \delta_{im2}^*$ (for all m, except the reference category 0). Hence, the negative Saltus parameter estimate indicates that the 'Do' items' thresholds were higher for subjects in Class 2 compared to subjects in Class 1. In other words, the subjects in Class 2 were less likely to endorse 'Yes' or 'Perhaps' (rather than 'No') to the 'Do' items than the subjects in Class 1. This result also suggests that Class 2 included a group of subjects who showed a lower level of verbal aggression to the 'Do' behavior items compared to Class 1 subjects. On the other hand, the Saltus parameter for the item loading parameters ($\tau_{22}^{(\alpha)}$) was positive and significant at the 5% level with an estimate of 1.632 (SE=0.465, p=0.000). Since $\tau_{22}^{(\alpha)} = \alpha_{i2}^* - \alpha_{i1}^*$, this means that the 'Do' items were more discriminating for subjects in Class 2 compared to subjects in Class 1. In other words, the 'Do' behavior items can better differentiate between people who display similar levels of verbal aggression in Class 2 than in Class 1.

Next, it is also important to examine the mean of the class-specific latent trait distribution (μ_h) to understand the characteristics of the specified latent classes. This information is given at the output section for Means each under Latent Class 1 and Latent Class 2 headings. F1 is the label for the latent trait specified in the model. The mean of the Class 1 latent trait was fixed at zero (hence, both the estimate and the standard error are 0.000 in the output), while the mean of the Class 2 latent trait was 0.228 ($SE = 0.190, p = 0.229$). This means that the two latent class means were not significantly different from each other at the 5% significance level. In other words, little evidence was found that there was a difference in the overall verbal aggression level between the two classes.

Note that the between-class difference was found only in the Saltus parameters for the threshold and loading parameters ($\tau_{22}^{(\delta)}$, $\tau_{22}^{(\alpha)}$), that is, in terms of the way the subjects react to the 'Do' items. The fact that the two

Saltus parameter estimates were significantly different from zero means that there were systematic differences in the estimated item parameter values between the two latent classes. To see this more clearly, class-specific item parameter values can be examined at the F1 BY and Thresholds output sections each under Latent Class 1 and Latent Class 2 headings. At the F1 BY section, the estimates for labels U1 to U24 indicate the item loading parameters for items 1 to 24, respectively. At the Thresholds section, labels U1$1 and U1$2 (to U24$1 and U24$2) denote the item threshold parameters for Category 1 and Category 2, respectively (Category 0 is the reference response category). Figure 1 displays the estimated item parameter values (α_i and δ_{im}) for Class 1 and Class 2.

It can be observed in Figure 1 that the α_i and δ_{im} parameter estimates for the two latent classes differ only for the 'Do' items (items 13 to 24). The differences in the item parameters amount to $\tau_{22}^{(\alpha)}$ and $\tau_{22}^{(\delta)}$ for the α_i and δ_{im} parameters, respectively.

Lastly, to test whether there was a gender difference in subjects' latent class membership, the regression coefficients of the multinomial logit model (Equation 14) need to be examined. The Mplus output section *Categorical Latent Variables* can be used to this purpose. Label MALE is found under C#1 ON at this output section, indicating that the corresponding result reports the estimate of the regression coefficient of the male variable for predicting subjects' Class 1 membership (meaning that Class 2 was treated as the reference latent class). Since Class 1 is used as the reference latent class in the current analysis, the result reported under ALTERNATIVE PARAMETERIZATIONS FOR THE CATEGORICAL LATENT VARIABLE REGRESSION should be utilized, which contains the section Parameterization using Reference Class 1. Under C#2 ON of this section, label MALE indicates the regression coefficient estimate of the male variable for predicting subjects' Class 2 membership. The estimate (γ_{h1}, $h = 2$) was negative and significant with -1.567 ($SE = 0.523, p = 0.003$). This means that there was sufficient evidence at the 5% significance level that male subjects had a lower probability of being classified into Class 2. Or equivalently, male students had a higher probability of being classified into Class

1 (who displayed a higher verbal aggression level to 'Do' behavior items compared to Class 2 subjects). This result confirms that there was indeed a gender difference in the subjects' class membership as hypothesized (that is, more male students were found in Class 1 than in Class 2). An additional piece of information that can be taken from this result is that male subjects had a higher tendency to express a higher level of verbal aggression to the 'Do' behavior items compared to female students, which is in line with the literature that reports a higher verbal aggression level for males. The fitted Saltus model adds to the literature by finding that the difference might have been driven by the 'Do' behavior type items.

Discussion

This paper discussed a specialized, constrained confirmatory mixture IRT model, referred to as the Saltus model (Mislevy & Wilson, 1996; Wilson, 1989). The model is confirmatory in that the number/nature of latent classes is known in advance and prior knowledge on item characteristics is used to hypothesize the item-class relationship.

Researchers who are more familiar with an ordinary, exploratory use of mixture IRT modeling may feel that such a confirmatory approach seems somewhat unusual. However, as discussed earlier in this paper, a confirmatory approach could reasonably be adopted in mixture IRT modeling, as it can be used to verify a researcher's hypothesis on the character of postulated latent classes and latent class differentiation. This purpose is unique and differentiated from the purpose of an exploratory approach.

Various extensions of the Saltus model were discussed with discrimination parameters, person predictors, and ordinal response data, and the estimation of an extended Saltus model with Mplus was illustrated with an empirical dataset. Those extensions are not exhaustive and other types of extensions are certainly possible. For instance, the model could be expanded to handle multidimensional test items and multilevel/longitudinal data structures. Future studies will investigate the applications and estimation of these extensions.

Authors' note

I am grateful to Dr. Perman Gochyyev for his very helpful comments on an earlier version of the paper.

References

Ackerman, T. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, *29*, 674–691. doi:10.1111/j.1745-3984.1992.tb00368.x

Alexeev, N., Templin, J., & Cohen, A. (2011). Spurious latent classes in the mixture rasch model. *Journal of Educational Measurement*, *3*, 313–332. doi:10.1111/j.1745-3984.2011.00146.x

Archer, J. (2004). Sex differences in aggression in real-world setting: A meta-analytic review. *Review of General Psychology*, *8*, 291–222. doi:10.1037/1089-2680.8.4.291

Boeckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665–678. doi:10.1037/a0028111

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A mixture item response model for multiple-choice data. *Journal of Educational and Behavioral Statistics*, *26*, 381–409. doi:10.3102/10769986026004381

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture rasch model with ordinal constraints. *Journal of Educational Measurement*, *39*, 331–348. doi:10.1111/j.1745-3984.2002.tb01146.x

Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, *72*, 393–411. doi:10.1007/s11336-007-9005-4

Cho, S.-j. & Cohen, A. S. (2010). A multilevel mixture IRT model with an application to dif. *Journal of Educational and Behavioral Statistics*, *35*, 336–370. doi:10.3102/1076998609353111

Cho, S.-j., Suh, Y., & Lee, W.-y. (2016). An ncme instructional module on latent dif analysis using mixture item response models. *Educational Measurement: Issues and Practice*, *35*, 48–61. doi:10.1111/emip.12093

Choi, I.-h. & Wilson, M. (2015). Multidimensional classification of examinees using the mixture random weights linear logistic test model. *Educational and Psychological Measurement*, *75*, 78–101. doi:10.1177/0013164414522124

Cohen, A. S. & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, *4*, 133–148. doi:10.1111/j.1745-3984.2005.00007

Daniel, R. C. & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem-solving items. *Applied Psychological Measurement*, *35*, 348–364. doi:10.1177/0146621609349801

De Boeck, P., Cho, S.-j., & Wilson, M. (2011). Explanatory secondary dimension modeling of latent differential item functioning. *Applied Psychological Measurement*, *35*, 583–603. doi:10.1177/0146621611428446

De Boeck, P. & Partchev, I. (2012). IRTrees: tree-based item response models of the GLMM family. *Journal of Statistical Software*, *48*, 1–28. doi:10.18637/jss.v048.c01

De Boeck, P. & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer-Verlag.

Draney, K. (2007). The Saltus model applied to proportional reasoning data. *Journal of Applied Measurement, 8*, 438–455.

Draney, K. & Jeon, M. (2011). Investigating the Saltus model as a tool for setting standards. *Psychological Testing and Assessment Modeling*, *53*, 486–498.

Draney, K. & Wilson, M. (2008). A lltm approach to the examination of teachers' ratings of classroom assessment tasks. *Psychology Science*, *50*, 417–432.

Finch, W. H. & Pierson, E. E. (2011). A mixture IRT analysis of risky youth behavior. *Frontiers in Psychology*, *2*(98), 1–15. doi:10.3389/fpsyg.2011.00098

Fischer, G. H. (1973). Linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*, 359–374. doi:10.1016/0001-6918(73)90003-6

Holland, P. W. & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Huang, G.-H. & Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, *69*, 5–32. doi:10.1007/BF02295837

Huang, H.-Y. (2016). Mixture IRT model with a higher-order structure for latent traits. *Educational and Psychological Measurement*, *77*, 275–304. doi:10.1177/0013164416640327

Jeon, M. & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods*, *48*, 1070–1085. doi:10.3758/s13428-015-0631-y

Jeon, M. & Rijmen, F. (2016). A modular approach for item response theory modeling with the R package FLIRT. *Behavior Research Methods*, *48*, 742–755. doi:10.3758/s13428-015-0606-z

Jin, K.-y., Chen, H.-f., & Wang, W.-c. (2018). Mixture item response models for inattentive responding behavior. *Organizational Research Methods*, *2*, 197–225. doi:10.1177/1094428117725792

Kubinger, K. D. (2008). On the revival of the rasch model-based lltm: from constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly*, *20*, 311–327.

Mislevy, R. J. & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215. doi:10.1007/BF02295283

Mislevy, R. J. & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, *61*, 41–71. doi:10.1007/BF02296958

Molenaar, D., Oberski, D., Vermunt, J., & De Boeck, P. (2016). Hidden Markov IRT models for responses and response times. *Multivariate Behavioral Research*, *51*, 606–626. doi:10.1080/00273171.2016.1192983

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176. doi:10.1177/014662169201600206

Muthén, B. & Lehman, J. (1985). Multiple group IRT modeling: Applications to item bias analysis. *Journal of Educational Statistics*, *10*, 133–142. doi:10.3102/10769986010002133

Muthén, L. & Muthén, B. (2017). *Mplus user's guide*. Angeles, CA: Muthen & Muthen.

Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14,* 535–569. doi:10.1080/10705510701575396

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271–282. doi:10.1177/014662169001400305

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, *34*, 100–114. doi:10.1002/j.2333-8504.1968.tb00153.x

Schnipke, D. L. & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213–232. doi:10.1111/j.1745-3984.1997.tb00516.x

Smit, A., Kelderman, H., & van der Flier, H. (2000). The mixed birnbaum model: estimation using collateral information. *Methods of Psychological Research Online*, *5*, 31–43.

Tijmstra, J., Bolsinova, M. A., & Jeon, M. (2018). Generalized mixture IRT models with different item-response structures: A case study using likert-scale data. *Behavior Research Methods*, *55*, 1–20. doi:10.3758/s13428-017-0997-0

Vansteelandt, K. (2000). *Formal models for contextualized personality psychology*. K.U. Leuven, Belgium: Unpublished doctoral dissertation.

Vansteelandt, K. & Van Mechelen, I. (2004). The personality triad in balance: multidimensional individual differences in situation-behavior profiles. *Journal of Research in Personality*, *38*, 367–393. doi:10.1016/j.jrp.2003.08.001

Vermunt, J. K. (1999). A general class of nonparametric models for ordinal categorical data. *Sociological*
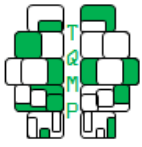
Methodology, *29*, 187–223. doi:10.1111/0081-1750. 00064

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*, 33–51. doi:10.1177/0962280207081238

Vermunt, J. K. & Magidson, J. (2005). *Latent gold 4.0 [computer program]*. Belmont, MA: Statistical Innovations Inc.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307. doi:10.1348/000711007X193957

von Davier, M. & Yamamoto, K. (2007). Mixture distribution rasch models and hybrid rasch models. In M. von Davier & C. Carstensen (Eds.), *Multivariate and mixture distribution rasch models* (pp. 99–115). New York: Springer.

Wanders, R., van Loo, H., Vermunt, J. K., Meijer, R., Hartman, C., Schoevers, R., & de Jonge, P. (2016). Casting wider nets for anxiety and depression: disability-driven cross-diagnostic subtypes in a large cohort. *Psychological Medicine*, *46*, 3371–3382. doi:10.1017/S0033291716002221

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, *105*, 276–289. doi:10.1037/0033-2909.105.2.276

Appendix

Here I provide the full Mplus syntax for estimating the extended, 2PL ordinal Saltus model that was illustrated in this paper. As discussed, the Saltus model and its extensions can be estimated as constrained mixture IRT models by imposing linear constraints on item parameters. I will show how this operation can be done in Mplus with this example.

```
<2PL ordinal Saltus model>
!! Header of input file
TITLE: 2PL ordinal Saltus model for verbal aggression data

!! Data file specification
DATA: FILE = verbal_pol.dat;

!! Define variables and specify number of latent classes
VARIABLE:
NAMES = male u1-u24; ! male is dummy variable for male subjects
CATEGORICAL = u1-u24; ! ordinal item responses
MISSING = ALL(99); ! missing data are coded as 99
CLASSES = c (2); ! number of latent classes

!! Estimation settings
ANALYSIS: TYPE = MIXTURE; ! estimate finite mixture model
ALGORITHM = INTEGRATION; ! use integration method (with 15 default quadrature
    points)
STARTS = 50 10 ; ! use multiple random starts (can be increased depending on
    studies)

!! Model specification
MODEL:
! Overall model
%OVERALL%
f1 BY u1-u24*; ! free loading parameters
f1@1; ! factor variance is fixed at 1
! Multinomial regression of class membership with male variable
c ON male; ! male variable is used to predict subjects' class membership

! Model for Class 1
%c#1%
f1 BY u1-u24* (x1-x24); ! loading parameters
```
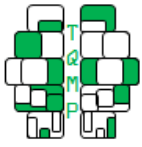
```
[u1$1-u24$1](a1-a24); ! threshold parameters for response category 1
[u1$2-u24$2](ax1-ax24); ! threshold parameters for response category 2
f1@1; ! factor variance is fixed at 1
[f1@0]; ! factor mean fixed at 0 (reference latent class)

! Model for Class 2
%c#2%
! Use different loading parameter labels for Do items (u13-u24)
f1 BY u1-u24* (x1-x12 y13-y24);
! Use different threshold parameter labels for 'Do' items (u13-u24)
[u1$1-u24$1](a1-a12 b13-b24); ! for response category 1
[u1$2-u24$2](ax1-ax12 bx13-bx24); ! for response category 2
f1@1; ! factor variance is fixed at 1
[f1]; ! factor mean freely estimated

! Set model constraints
MODEL CONSTRAINT:
NEW(tau_d tau_a); ! define Saltus parameters
!! Define Saltus parameter for threshold parameters as difference
!! in threshold parameters for 'Do' items between Class 1 and Class 2
! Category 1
tau_d = a13-b13;
tau_d = a14-b14;
tau_d = a15-b15;
tau_d = a16-b16;
tau_d = a17-b17;
tau_d = a18-b18;
tau_d = a19-b19;
tau_d = a20-b20;
tau_d = a21-b21;
tau_d = a22-b22;
tau_d = a23-b23;
tau_d = a24-b24;
! For Category 2
tau_d = ax13-bx13;
tau_d = ax14-bx14;
tau_d = ax15-bx15;
tau_d = ax16-bx16;
tau_d = ax17-bx17;
tau_d = ax18-bx18;
tau_d = ax19-bx19;
tau_d = ax20-bx20;
tau_d = ax21-bx21;
tau_d = ax22-bx22;
tau_d = ax23-bx23;
tau_d = ax24-bx24;

!! Define Saltus parameter for discrimination parameters as difference
!!in discrimination parameters for Do items between Class 2 and Class1
tau_a = y13-x13;
tau_a = y14-x14;
tau_a = y15-x15;
```

```
tau_a = y16-x16;
tau_a = y17-x17;
tau_a = y18-x18;
tau_a = y19-x19;
tau_a = y20-x20;
tau_a = y21-x21;
tau_a = y22-x22;
tau_a = y23-x23;
tau_a = y24-x24;


!! Save posterior probability in each class
SAVEDATA:
SAVE is cprob;
FILE is prob_pol.txt ;
```

Citation

Jeon, M. (2018). A constrained confirmatory mixture IRT model: Extensions and estimation of the Saltus model using Mplus. *The Quantitative Methods for Psychology*, *14*(2), 120–136. doi:10.20982/tqmp.14.2.p120