

Accommodating and Extending Various Models for Special Effects Within the Generalized Partially Confirmatory Factor Analysis Framework

Applied Psychological Measurement
2024, Vol. 48(4-5) 208–229

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01466216241261704

journals.sagepub.com/home/apm



Yifan Zhang¹  and Jinsong Chen¹ 

Abstract

Special measurement effects including the method and testlet effects are common issues in educational and psychological measurement. They are typically covered by various bifactor models or models for the multiple traits multiple methods (MTMM) structure for continuous data and by various testlet effect models for categorical data. However, existing models have some limitations in accommodating different type of effects. With slight modification, the generalized partially confirmatory factor analysis (GPCFA) framework can flexibly accommodate special effects for continuous and categorical cases with added benefits. Various bifactor, MTMM and testlet effect models can be linked to different variants of the revised GPCFA model. Compared to existing approaches, GPCFA offers multidimensionality for both the general and effect factors (or traits) and can address local dependence, mixed-type formats, and missingness jointly. Moreover, the partially confirmatory approach allows for regularization of the loading patterns, resulting in a simpler structure in both the general and special parts. We also provide a subroutine to compute the equivalent effect size. Simulation studies and real-data examples are used to demonstrate the performance and usefulness of the proposed approach under different situations.

Keywords

special effect, generalized partially confirmatory factor analysis, bifactor, multiple traits multiple methods, testlet effect

¹The University of Hong Kong, Hong Kong

Corresponding Author:

Jinsong Chen, Faculty of Education, The University of Hong Kong, Room 420, 4/F, Meng Wah Complex, Pokfulam Road, Hong Kong.

Email: jinsong.chen@live.com

Introduction

Special measurement effects including testlet and method effects due to different testlet formats, wording, rater characteristic, and so on are common in educational and psychological measurement. Broadly speaking, these special effects can be divided into two groups, each with various psychometric models designed to handle them. Special effects for continuous data are usually analyzed from the perspective of factor analysis (FA, [Brown, 2015](#)) with various models for bifactor ([Reise et al., 2010](#)) and multi-trait multi-method (MTMM; [Campbell & Fiske, 1959](#)) structures available. In contrast, special effects for categorical data are typically addressed from the perspective of item response theory (IRT) with various testlet effect models ([Bradlow et al., 1999](#)) available. The two perspectives are usually discussed separately in the literature, although they are statistically communicable by considering uni- or multidimensional IRT as categorical FA ([Takane & De Leeuw, 1987](#)).

Recently, a great deal of modeling developments can be found from both perspectives. For special effects with continuous data, various new models for bifactor and MTMM structures have been proposed (e.g., [Esra & Atar, 2021](#); [Geiser & Simmons, 2021](#); [Wang et al., 2018](#)). Note that both bifactor and MTMM-type models are usually adopted under the context of method effects. Standard-type bifactor models are confirmatory in nature, with one general factor, multiple orthogonal special factors, and a fully specified loading pattern (e.g., [Brown, 2015](#); [Chen et al., 2006](#); [Holzinger & Swineford, 1937](#)). When the loading pattern is partially specified, it becomes an exploratory bifactor model which has enjoyed a rediscovery more recently (e.g., [Giordano & Waller, 2020](#); [Jennrich & Bentler, 2011](#); [Reise, 2012](#); [Reise et al., 2010](#)). Although several methods are available, the preferred approach to exploratory bifactor analysis is the classic or derived Schmid-Leiman (SL; [Schmid & Leiman, 1957](#)) methods, which are subject to the proportionality constraints and hierarchical bifactor structure ([Giordano & Waller, 2020](#)). Taking into account that the general factor is fully specified (e.g., usually measured by all items), these models are partially rather than fully exploratory. Moreover, there is only one general factor and local independence is assumed in all existing bifactor models, which could further limit the applications of this type of models. Compared to bifactor models, models for MTMM are confirmatory only, but have more complex structures, which can be used to model the relationships among multiple traits (e.g., general factors) and methods (e.g., special factors). In its most general form, MTMM structure allows for multiple general factors, correlated special factors, and correlated uniqueness or measurement errors (i.e., local dependence). However, MTMM-type models are only identified under certain conditions. [Eid et al. \(2006\)](#) have detailed the structure and restriction of several identified MTMM-type models, including the correlated trait correlated uniqueness (CTCU) model, correlated trait uncorrelated method (CTUM) model, and correlated trait correlated method (CTCM) model. These three models are most widely used in the MTMM context (e.g., [Brown, 2015](#); [Kyriazos, 2018](#); [Marsh & Byrne, 1993](#)). Specifically, CTCU allows for local dependence. CTUM can be considered as an extension of the standard bifactor model with multiple general factors. CTCM can be problematic, especially when the number of special factors is small, and the correlated trait correlated method model with one method less (CTC(M-1)) is usually adopted instead. However, all these models are confirmatory by nature, making them inappropriate for any exploratory settings.

For special effects with categorical data, testlet effect models under the context of IRT are often employed to model the effect of a common stimulus within a bundle of items. The Bayesian random effect model on dichotomous responses proposed by [Bradlow et al. \(1999\)](#) is an early prototype of a testlet effect model. [Wang et al. \(2002\)](#) further extended the random effects model to polytomous responses and released the restriction of constant testlet effects. After that, [Wang and Wilson \(2005\)](#) proposed the Rasch testlet effect model as a special case of the multidimensional

Rasch model with the ability to deal with both dichotomous and polytomous responses. Based on Rasch testlet effect model, some multidimensional testlet effect models have been proposed recently (Zhan et al., 2014, 2015). In general, testlet effect models are confirmatory by nature, with only one general factor (i.e., trait) measured by all items and the assumption of local independence.

For the bifactor and MTMM-type models, the maximum likelihood estimation (MLE) implemented through the expectation-maximization algorithm is widely used. In comparison, the Bayesian approach is often employed in testlet effect models, which has several benefits (Fukuhara & Kamata, 2011; Wainer et al., 2007) including its flexibility and scalability to complex scenarios. Recently, a generalized partially confirmatory factor analysis (GPCFA) framework with mixed Bayesian Lasso methods was introduced (Chen, 2021). This Bayesian-based approach provides a regularized and flexible framework for factor analysis, with clear advantages such as making the exploratory and confirmatory approaches two ends of a continuum, handling both continuous and categorical data, and regularizing the complex loading structure and local dependence simultaneously.

In this research, we will modify the GPCFA framework to accommodate special effects for continuous and categorical data with an additional method to measure the effect size. As a result, the revised GPCFA can cover and extend many traditional models for special effects with added benefits. Specifically, bifactor models can be extended with multiple general factors and local dependence, and MTMM-type models can be extended to address partially exploratory settings with partially specified loading patterns, while extended testlet effect models can enjoy all above benefits. It will also be easier to explore and compare different models for special effects within a unified framework and effect size measure. Moreover, all models will automatically inherit GPCFA's other benefits such as the partially confirmatory design and addressing mixed-type variables with missingness and local dependence.

While building on the GPCFA framework, this research introduces several new contributions. First, we extend the GPCFA framework by dividing the factor structure into the general and special factors to accommodate various special effects. Second, we clarify how the revised GPCFA can incorporate or be connected with different bifactor, MTMM, or testlet structures, providing not only a unified framework to understand different models for special effects but also a chance to specify novel models for potential applications. For example, one can specify new bifactor, MTMM or testlet effect models that can handle data mixed with continuous, dichotomous, and polytomous responses and missingness, with or without regularization or local dependence. Third, we present a method to homogeneously measure and compare effect sizes across various settings. Fourth, we demonstrate the effectiveness of the revised framework and effect size measure through two simulation studies for continuous and categorical data, with illustrations of two real-life examples for applied researchers.

Theoretical Framework

GPCFA for special effects

The GPCFA is a factor analytic model with regularization of the loading pattern and local dependence, which makes it in between the exploratory and confirmatory ends (Chen, 2021). For complete description of this partially confirmatory methodology and related algorithms, readers can refer to the GPCFA literature (e.g., Chen, 2020, 2021; Chen et al., 2021).

Assume there are N respondents and J observed variables (i.e., items) with K latent factors. The observed response matrix $\mathbf{X} = (x_{ij})_{N \times J}$ contains both continuous and categorical variables. To

provide a unified framework for analysis, latent response matrix $\mathbf{Y} = (y_{ij})_{N \times J}$ is introduced using different link functions:

$$x_{ij} = \begin{cases} y_{ij} & \text{for continuous items} \\ \sum_{m=1}^{M_j} m \cdot I(\alpha_{j,m} \leq y_{ij} \leq \alpha_{j,m+1}) & \text{for categorical items} \end{cases} \quad (1)$$

where categorical item x_{ij} has M_j categories defined by the threshold vector $\alpha_j = (-\infty = \alpha_{j,1} < \alpha_{j,2} < \dots < \alpha_{j,M_j+1} = +\infty)^T$ and $I(A)$ is an indicator function that takes 1 if A is true and 0 otherwise. In this way, data mixed with both continuous and categorical items can be addressed homogeneously. The latent responses \mathbf{Y} can be expressed as:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{E} \quad (2)$$

where the vector $\boldsymbol{\mu}$ represents the $J \times 1$ intercept vector, matrix $\boldsymbol{\Lambda}$ represents the $J \times K$ loading matrix, \mathbf{F} represents K factors with the $K \times K$ factorial covariance matrix $\boldsymbol{\Phi}$, and \mathbf{E} represents the $J \times 1$ residuals with the $J \times J$ residual covariance matrix $\boldsymbol{\Psi}$.

Loading estimation structure is conducted through different settings of the design matrix $\mathbf{Q} = (q_{jk})_{J \times K}$, where $q_{jk} = -1, 0$ and 1 standing for the unspecified (Lasso parameters), zero-fixed and specified loadings, respectively. Specified loadings are free to be estimated by regular estimation (i.e., free parameters), and regularization for unspecified loadings and local dependence can be addressed simultaneously by Bayesian adaptive Lasso and covariance Lasso, as shown in [Appendix A](#).

To accommodate the GPCFA framework to deal with the method and testlet effects, several adjustments of the structure are needed. The factors are separated into two parts: the construct part (general factors) and the special effects part (special factors). We will use subscript “g” and “s” to denote the general and special factors and related loadings, respectively. Here, we use the following notations: total $K = K_g + K_s$ factors with K_g and K_s for the general and special factors, respectively. The latent responses \mathbf{Y} become:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}_g \mathbf{F}_g + \boldsymbol{\Lambda}_s \mathbf{F}_s + \mathbf{E} \quad (3)$$

where matrix $\boldsymbol{\Lambda}_g$ represents $J \times K_g$ general loading matrix, \mathbf{F}_g represents K_g general factors; matrix $\boldsymbol{\Lambda}_s$ represents $J \times K_s$ special loading matrix, \mathbf{F}_s represents K_s special factors; and \mathbf{E} represents $J \times 1$ measurement errors or residuals. Both loading matrices can be partially unspecified depending on the context. When all loadings are either specified or zero-fixed, the model is fully confirmatory. When more and more loadings are unspecified, the model becomes increasingly exploratory-inclined. The factorial covariance matrix $\boldsymbol{\Phi}$ can be separated as:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_g & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_s \end{bmatrix} \quad (4)$$

where $\boldsymbol{\Phi}_g$ is the $K_g \times K_g$ general factorial covariance, $\boldsymbol{\Phi}_s$ is a $K_s \times K_s$ diagonal matrix represents the special factor covariance matrix, as different special effects are usually independent of each other, and $\mathbf{0}$ is a zero matrix.

In practice, the general factor usually refers to latent construct or trait and the special factor usually derived from the study design, test format, or measurement method. [Figure 1](#) illustrates an example of the revised GPCFA model with 9 items and two general and two special factors. In this example, one loading per item is specified in both general and special factors, while the other loadings in the general factors are unspecified, and those in the special factors are set to zero.

Additionally, the general factors are correlated while the special factors are orthogonal. Meanwhile, the items can be continuous or categorical and local independence is assumed. In general, one can specify the Q-matrix, factor structure (correlated or orthogonal), local dependence, and type of each item to accommodate various bifactor, MTMM, or testlet structures.

The matrix equations and corresponding Q matrix is represented as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \\ \mu_7 \\ \mu_8 \\ \mu_9 \end{bmatrix} + \begin{bmatrix} \lambda_{g1,1} & \lambda_{s3,1} & 0 \\ \lambda_{g1,2} & \lambda_{s3,2} & 0 \\ \lambda_{g1,3} & \lambda_{s3,3} & 0 \\ \lambda_{g1,4} & 0 & \lambda_{s4,4} \\ \lambda_{g1,5} & 0 & \lambda_{s4,5} \\ & \lambda_{g2,6} & \lambda_{s3,6} & 0 \\ & \lambda_{g2,7} & \lambda_{s3,7} & 0 \\ & \lambda_{g2,8} & 0 & \lambda_{s4,8} \\ & \lambda_{g2,9} & 0 & \lambda_{s4,9} \end{bmatrix} \cdot \begin{bmatrix} F_{g1} \\ F_{g2} \\ F_{s1} \\ F_{s2} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \\ -1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{bmatrix} \quad (5)$$

For model identification, two constraints are necessary: setting all factorial variances to one, which constrains Φ as a correlation matrix; and specifying a few loadings for each factor when local independence is assumed (Chen, 2021). To address local dependence, at least one loading per item must be specified. Additionally, all latent response vectors for both continuous and categorical variables will be standardized to maintain consistency. All parameters will be iteratively implemented by MCMC. It is worth noting that identification is less stringent under the Lasso regularization: GPCFA models with or without local independence can be identified with a few loadings specified per factor or one loading specified per item, respectively (e.g., Chen, 2020, 2021; Chen et al., 2021).

Accommodating Bifactor Models

The revised GPCFA framework can accommodate various bifactor models. Specifically, a standard bifactor model with one general factor can be obtained by setting $K_g = 1$ and having each item load on only one of several method dimensions. The first column of the loading matrix in equation (5) is then fully specified as a general factor, and the factorial correlation matrix becomes a diagonal matrix. It is a fully confirmatory model and the design matrix \mathbf{Q} consists of 0 and 1 only (i.e., all loadings are specified). Figure 2 provides an example of the standard bifactor model.

The model becomes partially confirmatory when the loading matrix for the special factors are partially unspecified, which is similar to an exploratory bifactor model allowing cross-loadings on special factors. It is worth noting that the SL-type exploratory bifactor models are subject to the proportionality constraint due to the SL-transformation procedure (Yung et al., 1999). This constraint requires a linear combination between general factor loadings and special factor loadings. Considering that the general factor is fully specified (i.e., no exploration needed), these models are partially rather than fully exploratory.

In comparison, the revised GPCFA is subject to the identification constraint of a few specified loadings per factor (under the assumption of local independence), rather than the SL constraint. But one can evaluate the constraint post hoc. Under the GPCFA framework, the standard bifactor

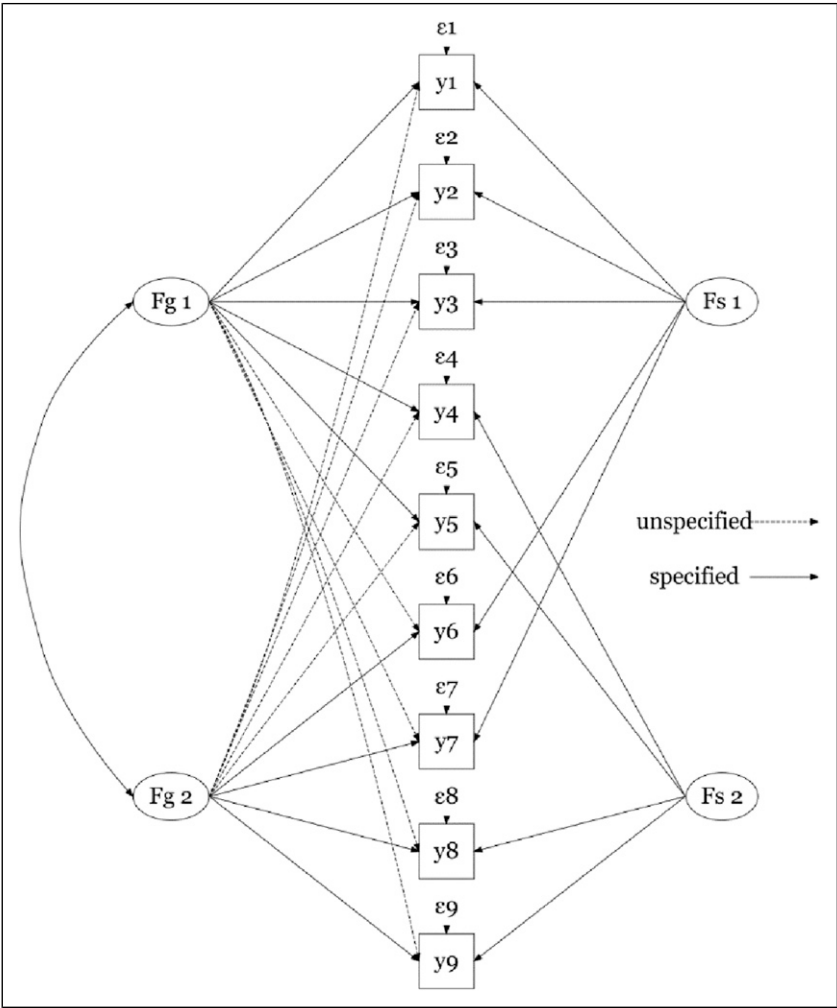


Figure 1. An example of revised GPCFA.

model can be extended with multiple general factors which can be correlated. In addition, both the loading matrices for the general and special factors can be partially unspecified, making the extended bifactor model exploratory-inclined. Compared to exploratory bifactor models, the revised GPCFA offers more flexibility and scalability by covering multiple general factors, mixed-type data, missingness, and local dependence.

Accommodating MTMM structures

The separation of general and special factors in the revised GPCFA framework allows us to accommodate several MTMM structures with four examples shown in [Figure 3](#).

The correlated trait correlated uniqueness (CTCU) model can be directly estimated by the original GPCFA model by considering different traits as different factors with local dependence. The observed variable \mathbf{Y} can be decomposed into common trait variable (general factor) \mathbf{F}_g and residual \mathbf{E} with intercept $\boldsymbol{\mu} = 0$:

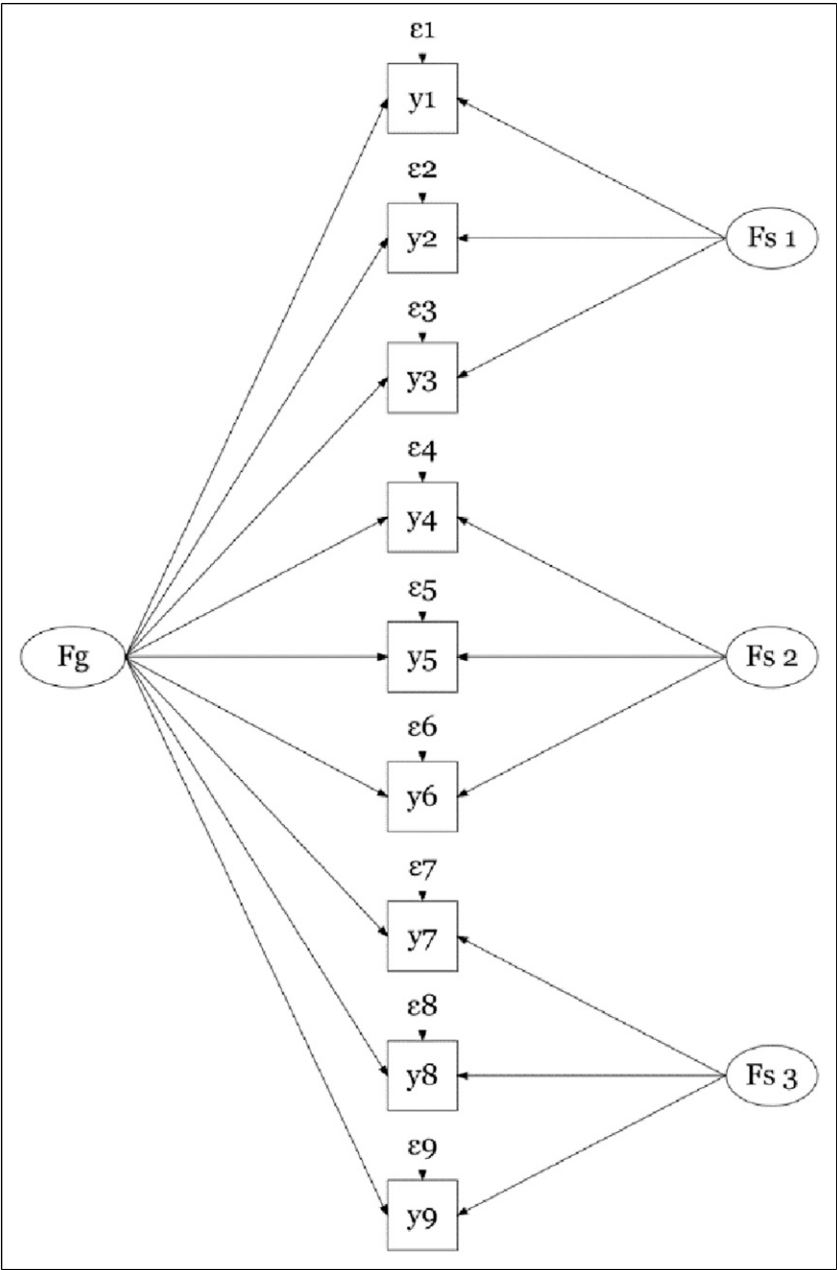


Figure 2. Standard bifactor model.

$$Y = \Lambda_g F_g + E \tag{6}$$

The residual in CTCU can be correlated between items within the same method effects. Namely, the correlated residual is equivalent to the special factor substantively, both of which refer to the method effect. However, the effect size is incalculable unless one can summarize multiple pairs of correlated residuals within the method reasonably.

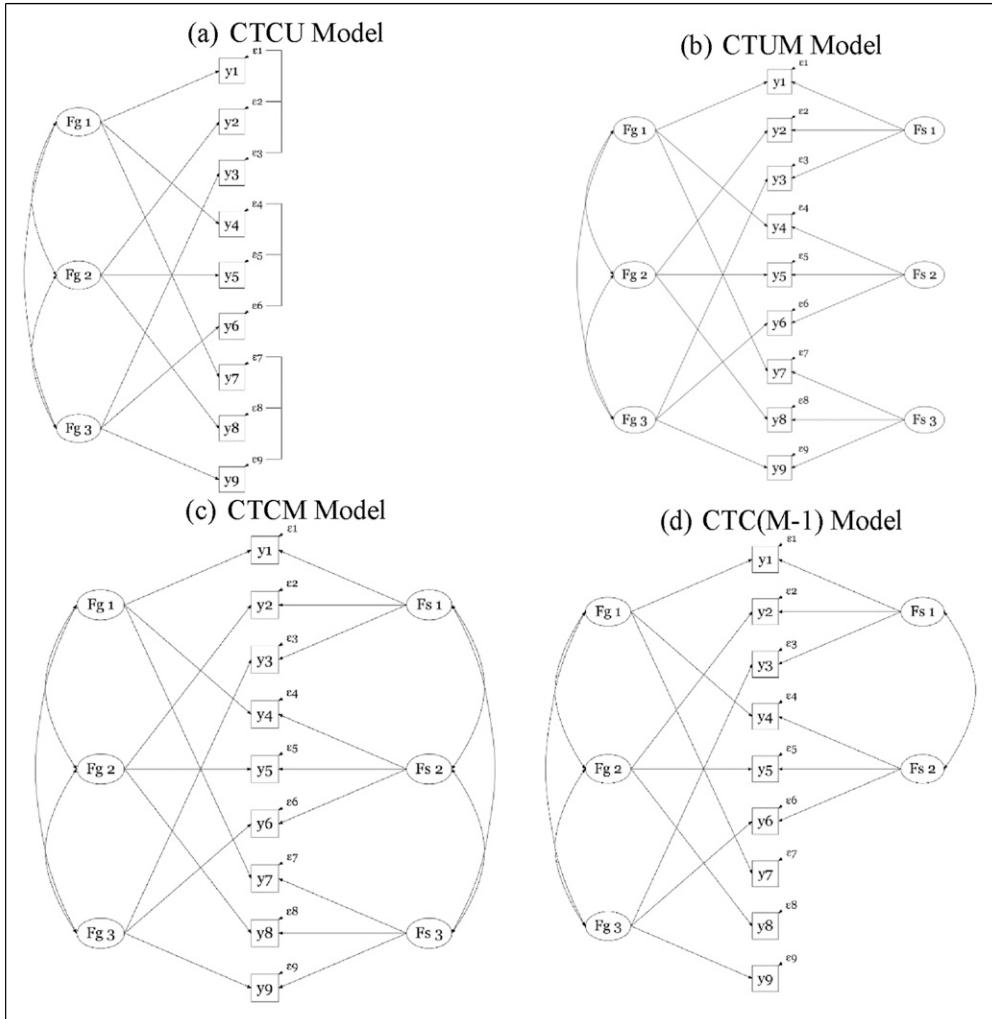


Figure 3. Structures of MTMM Related Models: (a) Correlated Trait Correlated Uniqueness model; (b) Correlated Trait Uncorrelated Method model; (c) Correlated Trait Correlated Method model; (d) Correlated Trait Correlated Method model with One method factor less than methods considered.

The correlated trait uncorrelated method (CTUM) is equivalent to the extended bifactor model, with multiple general factors that are correlated and multiple special factors that are uncorrelated:

$$\mathbf{Y} = \mathbf{\Lambda}_g \mathbf{F}_g + \mathbf{\Lambda}_s \mathbf{F}_s + \mathbf{E} \quad (7)$$

where the general factors and special factors denote the trait and method components, respectively. The general and special loadings $\mathbf{\Lambda}_g$, $\mathbf{\Lambda}_s$ are equivalent to trait and method loadings.

Correlated trait correlated method (CTCM) and CTC(M-1) are more generalized than CTUM by allowing all special factors to be correlated. In CTCM, the covariance matrix of special factors, $\mathbf{\Phi}_s$ can be non-diagonal. However, it is often subjected to the issue of identification. CTC(M-1) is a special variant of the CTCM which contains one method factor less than that in CTCM to identify the model. Even though CTC(M-1) solves the identification problem of CTCM, it is

difficult to distinguish between the sizes of different method effects due to the correlated methods. The transformation of GPCFA to different MTMM models will not change the characteristic of these models and scholars can still refer to the guidelines of MTMM to choose the appropriate model (Eid et al., 2006). On the other hand, one can make the models partially confirmatory by allowing unspecified loadings in the general or special loading matrices.

Accommodating Testlet Effect Models

This section introduces the transformation of GPCFA to accommodate several testlet effect models under the IRT context which are mainly for dichotomous responses. With the local independence assumption in IRT, the residual covariance matrix Ψ of GPCFA restricted to diagonal matrix. The response matrix of GPCFA in equation (3) is conditionally distributed as:

$$\mathbf{Y}|\mathbf{F}, \boldsymbol{\omega} \sim N(\Lambda_g \mathbf{F}_g + \Lambda_s \mathbf{F}_s + \boldsymbol{\mu}, \Psi) \quad (8)$$

It leads to the item response function with the normal ogive model, namely, using the cumulative function of the standard normal distribution Φ^* as the link function:

$$P(x_{ij} = 1|F_{gi}, F_{si}, \boldsymbol{\omega}) = P(y_{ij} > 0|F_{gi}, F_{si}, \boldsymbol{\omega}) = \Phi^* [\lambda_{gj}F_{gi} + \lambda_{sj}F_{si} + \mu_j] / \psi_{jj}^{1/2} \quad (9)$$

where $\boldsymbol{\omega}$ contains all unknown model parameters in $\Lambda_g, \Lambda_s, \boldsymbol{\mu}, \Phi$ and Ψ . The normal ogive function and logistic function are almost indistinguishable with linear transformation (Dinero & Haertel, 1977) and they can be connected by a constant 1.702 (Camilli, 1994). In the above model, we set the scale of the latent response y as one for model identification, which is equivalent to the categorical confirmatory factor analysis (CCFA) parameterization with the same normal ogive link function.

Alternatively, one can make model identified by setting the $\psi_{jj} = 1$, resulting in the multi-dimensional item response theory (MIRT) parameterization:

$$P(x_{ij} = 1|\theta_{gi}, \theta_{si}) = \Phi^* [a_{gj}\theta_{gi} + a_{sj}\theta_{si} - d_j] \quad (10)$$

The a_{gj} and a_{sj} indicate the discrimination parameters for the general factor and special (testlet) factor, while d_j indicates the location of the item under MIRT, which is similar to the intercept μ_j in GPCFA. θ_{gi} is the general latent trait in IRT, which follows multivariate normal distributions. θ_{si} (i.e., F_{si} in revised GPCFA) indicates the random effect from response i due to testlet s . It is evident that the above solutions are transformable with revised GPCFA:

$$\begin{aligned} a_{gj} &= \lambda_{gj} / \psi_{jj}^{1/2} \\ a_{sj} &= \lambda_{sj} / \psi_{jj}^{1/2} \\ d_j &= -\mu_j / \psi_{jj}^{1/2} \end{aligned} \quad (11)$$

Different testlet effect models can be obtained as follows: The general testlet effect model (Li et al., 2006) can be transformed from GPCFA by restricting the number of general factors to one and keeping the residual covariance matrix Ψ as a diagonal matrix, which is essentially the same as the bifactor factor analysis model for dichotomous responses (Gibbons & Hedeker, 1992). The formula is similar to equation (10) with $\theta_{si} \sim N(0, 1)$.

The two-parameter normal ogive (2PNO) testlet effect model proposed by Bradlow et al. (1999) can be written for each latent response y_{ij} as:

$$y_{ij} = a_j(\theta_{gi} + \theta_{si} - b_j)$$

where $\theta_{si} \sim N(0, \sigma_{\theta_{si}}^2)$. Compared to equation (10), it adds a proportional constraint to testlet θ_{si} (special factor F_{si} in GPCFA) and replaces the location parameter d_j with difficulty parameter b_j , while keeping the testlet effect constant among all testlets. The 2PNO testlet effect model can be re-expressed as the extended 2PNO testlet effect model (Chen et al., 2006; Rijmen, 2010):

$$P(x_{ij} = 1 | \theta_{gi}, \theta_{si}) = \Phi^* (a_{gj}\theta_{gi} + a_{gj}C_s\theta_{si} - a_{gj}b_j) \quad (13)$$

The transformation between GPCFA and the extended 2PNO testlet effect model is:

$$\begin{aligned} a_{gj} &= \lambda_{gj} / \psi_{jj}^{1/2} \\ a_{gj}\sigma_{\theta_{si}} &= \lambda_{sj} / \psi_{jj}^{1/2} \\ a_{gj}b_j &= -\mu_j / \psi_{jj}^{1/2} \end{aligned} \quad (14)$$

The constant C_s is for testlet s and equals to the standard deviation of θ_{si} . Alternatively, one can use the testlet discrimination parameter a_{sj} in equation (13), which equals to the product of discrimination on the general trait and the standard deviation of testlet effect: $a_{sj} = a_{gj}\sigma_{\theta_{si}}$. Since the variance for all testlets is the same in this model ($\sigma_{\theta_{si}}^2 = \sigma_{\theta_s}^2$), the corresponding variances for all special factors will be equal in GPCFA.

In the extended 2PNO testlet effect model, if we restrict the discrimination parameters a_{gj} to one and release the constraint of testlet variance, it can be converted to the one-parameter Rasch testlet effect model (Wang & Wilson, 2005):

$$P(x_{ij} = 1 | \theta_{gi}, \theta_{si}) = \Phi^* (\theta_{gi} + \theta_{si} - b_j) \quad (15)$$

Here, the general trait loadings are equal to one for all items. The loading represents the discriminating power, and the variance of true-score equals to the square of the loading (McDonald, 2013). The variance of testlet ($\sigma_{\theta_{si}}^2$) represents the testlet effect, which is equivalent to the average of square loadings of the special factor in GPCFA.

In addition, GPCFA can also estimate the extended Rasch testlet effect model, like within-item multidimensional testlet effect model (Zhan et al., 2014, 2015) by releasing the constraint of special factor allocation (i.e., adding the number of special factors). Moreover, polytomous responses or mixed-type formats, with missingness and local dependence, can be readily addressed within the GPCFA framework. However, the GPCFA cannot be transformed into the three-parameter testlet effect model with guessing parameter yet.

Equivalent Effect Size

Typically, the effect size refers to the amount of variance due to the random effect when the general factor or trait is standardized (Wang & Wilson, 2005). Since all factors are standardized under GPCFA, we define the equivalent effect size measure D , the average of the square loadings on the random effect (i.e., special factor). Moreover, due to the partially confirmatory setting, we separate the random effect into two parts. For specified loadings in the loading vector of the special factor, all loading estimates will be included. For unspecified loadings, we only consider those loading estimates with absolute values greater than the cutoff of 0.1, which is typically used under the Lasso scenarios. We denote the $J \times 1$ loading estimation vector for special factor k as $\Lambda_k = (\Lambda'_k, \Lambda''_k)^T$ where Λ_k is the k th column of Λ . The $J' \times 1$ vector Λ'_k and $J'' \times 1$ vector Λ''_k are the

specified and unspecified parts within the cutoff value, respectively. The effect size is calculated as:

$$D_k = \frac{\sum_{j=1}^{J'} (\lambda'_{jk})^2 + \sum_{j=1}^{J''} (\lambda''_{jk})^2}{J' + J''}, \text{ where } |\lambda''_{jk}| \geq 0.1 \quad (16)$$

Simulation Studies

Two simulation studies were conducted to evaluate the performance of the proposed model on continuous and categorical data. Based on previous research (Chen, 2021) and real-life examples, a sample size of $N = 1000$ was used, and two effect sizes ($D = 0.1$ and 0.2) were simulated. Different conditions of local dependence and other settings were manipulated as shown in the following studies. The true diagram for two simulation studies is given in Appendix C. For each condition cell, 200 replications were simulated and analyzed. The performance assessment for each parameter includes the bias (BIAS), the mean of the standard error (SE), the root mean square error (RMSE) between the estimates and the true values, and the percentage of estimates that differed significantly from zero ($\alpha = .05$) based on the highest posterior density (HPD) interval (SIG%). The SIG% indicates the percentage of Type I error for zero Lasso loadings.

To stabilize most Markov chains (i.e., the estimated potential scale reduction (EPSR) < 1.1 (Gelman et al., 2014)), 20,000 iterations of burn-in draws were performed, followed by additional 20,000 iterations. All studies used the LAWBL package (Chen, 2022) in the R (R Development Core Team, 2021) computing environment. The *sim_lvm* and *pcfa* functions in the package were used to generate and analyze the data, respectively. A sample implementation code is provided in Appendix D.

Study 1: Model Performance and Comparisons of Special Effects for Continuous Data under Local Dependence

In this study, we evaluated the performance of the proposed models for special effects for continuous data under local dependence. We set the number of general factors and special factors as $K_g = K_s = 3$, with six items per special factor (i.e., $J = 18$). The true loading matrix was:

$$\Lambda^T = \begin{bmatrix} \lambda_{g1} & \lambda_{g2} & \lambda_{g3} & \lambda_{g4} & \lambda_{g5} & \lambda_{g6} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} \\ \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g1} & \lambda_{g2} & \lambda_{g3} & \lambda_{g4} & \lambda_{g5} & \lambda_{g6} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} \\ \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g0} & \lambda_{g1} & \lambda_{g2} & \lambda_{g3} & \lambda_{g4} & \lambda_{g5} & \lambda_{g6} \\ \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} \\ \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} \\ \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} \end{bmatrix}$$

where major loadings for each general factor were set from 0.5 to 0.75 with a 0.05 interval for λ_{g1} to λ_{g6} . Each special factor had 6 identical nonzero loadings determined by effect size (e.g., if the effect size $D = 0.1$, all nonzero loading would be set around $\lambda_{s1} = 0.3162$ in special factors). Other loadings were set as $\lambda_{g/s0} = 0$. All nonzero loadings in the true matrix were estimated as specified loadings, others in general factors were set as unspecified and in special factors were set as zero. The design matrix Q was:

$$\mathbf{Q}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Two factorial correlations between general factors were investigated: $\phi_{kk'} = 0.3$ and 0.6 , where k and $k' = 1$ to 3 and $k \neq k'$. For local dependence, nonzero off-diagonal elements of Ψ were $\psi_{13} = \psi_{31} = \psi_{23} = \psi_{32} = \psi_{46} = \psi_{64} = \psi_{56} = \psi_{65} = 0.2$. Other off-diagonal terms were $\psi_0 = 0$.

Table 1 presents the simulation results for continuous variables. Due to the symmetry of the design, the loading estimates are averaged across the general or special factors, and details of loading recovery for general factors are shown in Appendix B. The overall estimation for small effect size was better than the large one. Specifically, the general factor loading recovery was satisfactory for small effect size and small factorial correlation. While in condition of large factorial correlation, the model could overestimate SE (i.e., ~ 1) and have relatively low SIG% for large effect size. The recovery of special factor loadings and local dependence were similar, which was acceptable in terms of BIAS, RMSE and SE, while the SIG% was poor. The power of local dependence decreased when the related true loadings increased, which informed that it was increasingly difficult to detect local dependence with large loadings. All three special factors' effect size (i.e., D) were estimated perfectly. For factorial correlations' estimation, the recovery was satisfactory except for the SIG% and small factorial correlation was slightly better than the large one.

Study 2: Model Performance and Comparisons of Special Effects for Categorical Data

In this study, we evaluated the performance of the revised GPCFA for special effects with categorical data under the assumption of local independence, which are common in testlet effect models. We investigated two levels of the number of categories, $M = 2$ and 4 , for all items. The number of general factors and special factors were set as $K_g = 1$ and $K_s = 3$, respectively, with six items per special factor, namely, $J = 18$. The true loading matrix was set as:

$$\mathbf{\Lambda}^T = \begin{bmatrix} \lambda_{g1} & \lambda_{g2} & \lambda_{g3} & \lambda_{g4} & \lambda_{g5} & \lambda_{g6} & \lambda_{s1} & \lambda_{s2} & \lambda_{s3} & \lambda_{s4} & \lambda_{s5} & \lambda_{s6} & \lambda_{s1} & \lambda_{s2} & \lambda_{s3} & \lambda_{s4} & \lambda_{s5} & \lambda_{s6} \\ \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} \\ \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} \\ \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s0} & \lambda_{s1} & \lambda_{s1} \end{bmatrix}$$

General factor loadings were set from 0.5 to 0.75 with an interval of 0.05 , repeated three times. All 6 nonzero loadings for each special factor were determined by the effect size, as in study 1. All these loadings with nonzero true values were set as specified loadings which were freely estimated. Other loadings were set as $\lambda_{g/s0} = 0$ and were estimated by regularization.

Table 2 summarizes the simulation results for categorical variables. The general factor loading estimates were similar across three parts (λ_{g1} to λ_{g6}) due to the repeated design, and the average of three parts was displayed to save space. The estimation recovery for loadings on general factors was satisfactory for all four conditions. More specifically, estimation on small effect size was

Table 1. Simulation Results for Study 1.

Par	True	$\phi = 0.3$				$\phi = 0.6$			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
$D = 0.1$									
λ_{g1}	0.5	0.002	0.036	0.086	0.998	-0.013	0.044	0.139	1.000
λ_{g2}	0.55	0.001	0.032	0.087	1.000	-0.019	0.044	0.140	1.000
λ_{g3}	0.6	-0.002	0.042	0.086	1.000	-0.019	0.055	0.135	1.000
λ_{g4}	0.65	0.005	0.049	0.089	1.000	-0.021	0.060	0.136	1.000
λ_{g5}	0.7	-0.018	0.038	0.082	1.000	-0.054	0.064	0.127	1.000
λ_{g6}	0.75	-0.003	0.025	0.085	1.000	-0.042	0.052	0.131	1.000
λ_{s1}	0.316	-0.022	0.101	0.122	0.650	-0.025	0.102	0.122	0.632
D_1	0.1	0.003	0.012	0.025	1.000	0.003	0.012	0.025	1.000
D_2	0.1	0.013	0.018	0.024	1.000	0.013	0.018	0.024	1.000
D_3	0.1	0.030	0.033	0.027	1.000	0.026	0.030	0.027	1.000
$\lambda_{g/s0}$	0	0.010	0.031	0.097	0.000	0.025	0.042	0.131	0.000
ψ_{13}	0.2	-0.036	0.047	0.078	0.620	-0.037	0.047	0.076	0.625
ψ_{23}	0.2	-0.038	0.048	0.082	0.450	-0.040	0.049	0.081	0.425
ψ_{46}	0.2	-0.052	0.058	0.101	0.025	-0.035	0.041	0.096	0.190
ψ_{56}	0.2	-0.139	0.139	0.068	0.000	-0.132	0.133	0.072	0.000
$\phi_{kk'}$	0.3/0.6	-0.040	0.054	0.168	0.082	-0.088	0.094	0.196	0.968
$D = 0.2$									
λ_{g1}	0.5	-0.009	0.061	0.127	0.938	-0.059	0.102	0.198	0.707
λ_{g2}	0.55	-0.008	0.060	0.135	0.942	-0.063	0.106	0.211	0.740
λ_{g3}	0.6	-0.037	0.085	0.119	0.938	-0.078	0.133	0.192	0.805
λ_{g4}	0.65	0.015	0.145	0.118	0.947	-0.044	0.191	0.181	0.812
λ_{g5}	0.7	-0.034	0.079	0.119	0.958	-0.109	0.140	0.191	0.825
λ_{g6}	0.75	0.009	0.068	0.121	0.958	-0.068	0.126	0.191	0.838
λ_{s1}	0.447	-0.064	0.186	0.118	0.704	-0.081	0.194	0.126	0.596
D_1	0.2	-0.042	0.053	0.042	1.000	-0.047	0.057	0.043	1.000
D_2	0.2	0.061	0.064	0.033	1.000	0.068	0.071	0.034	1.000
D_3	0.2	-0.011	0.027	0.032	1.000	-0.039	0.046	0.035	1.000
$\lambda_{g/s0}$	0	0.007	0.060	0.131	0.000	0.042	0.089	0.194	0.000
ψ_{13}	0.2	-0.029	0.046	0.065	0.830	-0.018	0.042	0.071	0.830
ψ_{23}	0.2	-0.030	0.045	0.069	0.760	-0.018	0.042	0.075	0.735
ψ_{46}	0.2	-0.152	0.153	0.050	0.000	-0.133	0.135	0.065	0.000
ψ_{56}	0.2	-0.195	0.195	0.014	0.000	-0.191	0.191	0.022	0.000
$\phi_{kk'}$	0.3/0.6	-0.025	0.079	0.210	0.158	-0.166	0.203	0.298	0.348

Note. λ_{g1} – λ_{g6} averaged across all general factors; λ_{s1} averaged across all special factors; $\lambda_{g/s0}$ averaged across all zero loading estimates; D : effect size; For $\phi_{kk'}$, k and $k' = 1$ to 3 and $k \neq k'$; RMSE: root mean square error; SE: standard error; SIG%: percent of estimates differed from zero significantly ($\alpha = .05$).

slightly better than large one. The recovery of special factor loadings was poor with large SE and lower power (i.e., SIG%). However, the effect size estimates were significant and satisfactory for all conditions, similar to study 1. The results for $M = 2$ and 4 were similar. The desirable results for zero loadings were found with nearly zero Type I error rates. The parameter set in this study can be seen as an extension of the classic Rasch testlet effect model. The results suggest that GPCFA has satisfactory performance in complex testlet cases.

Table 2. Simulation Results for Study 2.

Par	True	M = 2				M = 4			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
D = 0.1									
λ_{g1}	0.5	0.004	0.037	0.064	1.000	0.001	0.030	0.060	1.000
λ_{g2}	0.55	−0.002	0.035	0.063	1.000	−0.005	0.029	0.059	1.000
λ_{g3}	0.6	0.000	0.032	0.064	1.000	0.002	0.026	0.061	1.000
λ_{g4}	0.65	−0.005	0.030	0.064	1.000	−0.004	0.024	0.061	1.000
λ_{g5}	0.7	0.004	0.029	0.064	1.000	0.002	0.023	0.063	1.000
λ_{g6}	0.75	0.001	0.027	0.064	1.000	−0.001	0.021	0.062	1.000
λ_{s1}	0.316	−0.053	0.075	0.167	0.101	−0.038	0.058	0.148	0.243
D_1	0.1	−0.019	0.023	0.030	1.000	−0.015	0.019	0.030	1.000
D_2	0.1	−0.016	0.020	0.034	1.000	−0.015	0.019	0.037	1.000
D_3	0.1	−0.017	0.022	0.042	1.000	−0.015	0.019	0.044	1.000
$\lambda_{g/s0}$	0	0.019	0.037	0.113	0.000	0.018	0.032	0.099	0.000
D = 0.2									
λ_{g1}	0.5	0.003	0.040	0.082	1.000	−0.003	0.034	0.078	1.000
λ_{g2}	0.55	−0.005	0.039	0.080	1.000	−0.012	0.035	0.077	1.000
λ_{g3}	0.6	−0.007	0.037	0.083	1.000	−0.006	0.031	0.080	1.000
λ_{g4}	0.65	−0.014	0.038	0.083	1.000	−0.013	0.031	0.081	1.000
λ_{g5}	0.7	0.021	0.041	0.088	1.000	0.012	0.032	0.086	1.000
λ_{g6}	0.75	0.014	0.037	0.088	1.000	0.004	0.027	0.086	1.000
λ_{s1}	0.447	−0.048	0.079	0.165	0.567	−0.028	0.058	0.145	0.775
D_1	0.2	−0.039	0.045	0.054	1.000	−0.029	0.035	0.054	1.000
D_2	0.2	−0.036	0.043	0.061	1.000	−0.030	0.036	0.065	1.000
D_3	0.2	−0.059	0.066	0.078	1.000	−0.049	0.055	0.079	1.000
$\lambda_{g/s0}$	0	0.010	0.034	0.106	0.001	0.013	0.029	0.094	0.000

Note. λ_{gi} = average of three parts of general factor loadings; λ_{s1} averaged across all special factors; D: effect size; $\lambda_{g/s0}$ averaged across all zero loading estimates; M: number of categories; RMSE: root mean square error; SE: standard error; SIG %: percent of estimates differed from zero significantly ($\alpha = .05$).

Empirical Examples

Study 1: Humor Styles Questionnaire: Special Effect for Continuous Data

It is common to encounter special effects such as method effects due to wording, item formats, or reverse items in multidimensional psychological scales. In this study, the Humor Styles Questionnaire (HSQ) (Martin et al., 2003) was used to test if there’s a method effect for the reverse items using the GPCFA model. The HSQ consists of 32 items with four general factors, including 11 reverse items (Appendix E). The public dataset from 1070 respondents can be found at https://openpsychometrics.org/_rawdata/, which include 130 missing values.

Taking into account the method effect, we can specify each item to load exclusively on one general factor, with the 11 reverse items for a special factor for reverse wording. This design represents a special case of the MTMM structure with multiple traits, local independence and one method, which will be referred to as the baseline model. With all other loadings set as unspecified and estimated with regularization, we can evaluate two GPCFA models with and without local dependence between items. Estimates of local dependence and loading can be found in Table 3. Considering local dependence, the number of cross-loadings for general factors reduced from

18 to 8, and 12 significant correlated residual terms were found. The effect sizes for both cases were significant and were all around 0.05, which is small. But the identified loading patterns for the special factor were rather different. Two standard CFA models were fitted with cross-loadings and residual structure identified in GPCFA and were compared with the baseline model using Mplus (Muthén & Muthén, 2017). Table 4 shows that both GPCFA-suggested models fitted the data better than the baseline model, and the GPCFA with local dependence considered was the best and sole acceptable model. Table 5 shows that the factorial correlations are similar between the two models. The results suggest that both cross-loadings and residual covariance contribute to model fitness when we consider the special effect for the reverse items in HSQ.

Study 2: PISA Reading Assessment: Special Effect for Categorical Data

Educational assessments with testlet effects are common in psychometrics. In this study, we used the PISA reading assessment for UK in 2000 to explore the testlet effects by GPCFA (Chen & de la Torre, 2014). The dataset contains 1039 responses with 26 released items from six independent articles from booklet 8 and 9. For the baseline model, there was one general factor as reading literacy and 5 special factors as 5 independent articles (the last article was excluded due to too few items). The correspondent articles with specified items can be found in Appendix F (there is no cross-loading).

For the GPCFA, we can specify the loadings following the baseline model but leave all other loadings unspecified (rather than being fixed as zero). As shown in Table 6, the first article had the biggest effect size of around 0.1, and the fifth article had a moderate special effect (0.076). The effect sizes of the other three articles were around 0.05 which can be considered as trivial. Five significant cross-loadings were estimated among special factors without being pre-specified and were all around .1. Statistically speaking, these items share common stimulation even if they are not in the same article. The explanations can be sought in many ways. For example, the significant items estimated in the first special factor might all belong to number sense (Chen & de la Torre, 2014), and these findings might provide a reference for future research. We adopt categorical CFA to compare the baseline model with GPCFA-suggested model in Mplus. Fit evaluation in Table 7 suggest both models were acceptable with a small difference, but the GPCFA-suggested model still fitted the data better.

Discussion

Special effects including the method and testlet effects are common issues in psychological and educational measurement. This research extends the GPCFA framework to accommodate special effects for continuous and categorical data by modifying the factor structure. The revised GPCFA can be related to different bifactor, MTMM-type, and testlet effect models by setting different constraints. A useful indicator D was produced to measure and compare the special effect sizes. Models for special effects under the revised GPCFA framework offer multidimensionality for both the general and special factors (or traits) and automatically inherit GPCFA's benefits to accommodate the partially confirmatory design with regularizations, local dependence, mixed-type formats, and missingness jointly. As a result, it provides a chance to easily specify novel models for potential applications within one framework.

Compared with traditional bifactor, MTMM and testlet effect models, the revised GPCFA framework is more flexible in three ways. First, the proposed model allows for multiple general factors and special factors with different constraints on factorial correlation and local dependence flexibly. Second, different from traditional rotation and MLE, the Bayesian Lasso method in GPCFA can estimate loading matrix and local dependence at the same time while dealing with

Table 3. Significant Loadings and Residual Estimates for the Humor Styles Questionnaire.

Item	Local independent					Local dependent					LD	Effect
	F_{gl}	F_{g^2l}	F_{g^3}	F_{g^4}	F_{sl}	F_{gl}	F_{g^2}	F_{g^3}	F_{g^4}	F_{sl}		
1	0.668					0.657				0.146 ^a	$\Psi'_{24,4}$	0.114
2		0.636					0.618				$\Psi'_{27,4}$	-0.066
3			0.536	0.094				0.550			$\Psi'_{6,5}$	0.071
4				0.626					0.618		$\Psi'_{29,5}$	0.149
5	0.653				-0.136 ^a	0.612					$\Psi'_{30,6}$	0.247
6	0.215	0.379				0.187	0.382				$\Psi'_{20,8}$	0.141
7	-0.169		0.601			-0.139		0.550		0.177 ^a	$\Psi'_{18,10}$	0.131
8				0.784					0.735		$\Psi'_{19,11}$	0.098
9	0.476			0.092		0.447	0.735			0.192	$\Psi'_{21,11}$	-0.076
10	-0.118 ^a	0.816						0.542			$\Psi'_{25,13}$	0.214
11	-0.122		0.550						0.638		$\Psi'_{29,23}$	-0.073
12				0.640							$\Psi'_{28,27}$	0.078
13	0.620				0.269 ^a	0.613						
14		0.657					0.677					
15			0.680					0.596		0.309		
16			0.104	0.556					0.596	0.285		
17	0.773					0.760				0.166		
18	-0.167	0.827					0.732					
19	0.133	0.125	0.444			0.195		0.464				
20				0.782					0.719			
21	0.696		-0.135			0.702						
22		0.375	0.081				0.419			0.233		
23			0.497		0.149 ^a			0.463		0.252 ^a		
24	-0.168			0.478		-0.111*			0.461	-0.106 ^a		
25	0.703				0.366 ^a	0.646				0.174		
26		0.686					0.692					
27			0.548					0.552				

(continued)

Table 3. (continued)

Item	Local independent				Local dependent						Effect
	F_{g1}	F_{g2}	F_{g3}	F_{g4}	F_{s1}	F_{g1}	F_{g2}	F_{g3}	F_{g4}	F_{s1}	LD
28		0.178	0.154	0.240		0.108 ^a	0.136	0.163	0.241		
29	0.595		0.149	-0.192		0.532			-0.160	0.232	
30		0.454		-0.123			0.447				
31			0.692		0.110 ^a			0.597		0.342	
32				0.660					0.655		
D					0.039					0.060	

Note. F_{g1} = affiliative humor; F_{g2} = self-enhancing humor; F_{g3} = aggressive humor; F_{g4} = self-defeating humor; F_{s1} = special effect for reverse items; LD = local dependence (only significant terms were presented); D: effect size; only significant or above .1 loadings are presented; underscored are unspecifed loadings.

^aIndicates non-significant loadings.

Table 4. Different CFA Models' Fitness for the Humor Styles Questionnaire.

Model	RMSEA	CFI	TLI	SRMR	AIC	BIC	CHI ²	DF
M0	0.057	0.858	0.842	0.063	96029.639	96591.861	2017.008	447
M1	0.047	0.906	0.893	0.039	95509.458	96131.385	1472.827	435
M2	0.037	0.944	0.935	0.039	95096.709	95763.415	1042.078	426

Note. RMSEA = root mean square error of approximation; CFA = confirmatory factor analysis; M0 = baseline (no cross-loading or residual covariance); M1 & M2 = all significant loadings identified in the GPCFA; M1 = GPCFA with local independent; M2 = GPCFA with all significant residual covariance.

Table 5. Factorial Correlation for the Humor Styles Questionnaire.

	Local independent			Local dependent		
	F _{g1}	F _{g2}	F _{g3}	F _{g1}	F _{g2}	F _{g3}
F _{g2}	0.484			0.504		
F _{g3}	0.246	0.165		0.206	0.177	
F _{g4}	0.234	0.251	0.227	0.251	0.261	0.256

Note. All correlation estimates are significant.

mixed types of variables and missingness in a unified framework. Third, the regularization of loading structure covers a wide range of the substantive continuum. This partially confirmatory approach allows for regularization of the loading patterns, resulting in a simpler structure in both the general and effect parts. Unspecified loadings for both the general and special factors also offer us more flexibility to incorporate uncertainty (e.g., addressing cross-loadings) during the modeling process. Moreover, one can analyze both the testlet effect in IRT and method effect in factor analysis with the equivalent effect size in a unified way.

Two simulation studies and corresponding real-life studies were adopted to evaluate and illustrate how the revised GPCFA framework can address special effects for both continuous and categorical cases under different conditions. Specifically, the small effect size ($\sim .1$) achieved better model estimation in simulation studies and large effect size ($\sim .2$) might lead to over-estimating for some parameters. The real-life examples based on the Humor Styles Questionnaire and PISA reading assessment illustrate how GPCFA can be used to test different special effects and calculate the effect size. In practice, the special effects are usually around 0.1, and we can consider random effects with an effect size less than that value as negligible. Finally, the R package LAWBL (Chen, 2022) is free and powerful in implementing GPCFA with different types of constraints.

There are a few limitations in this research. From the estimation perspective, Bayesian Lasso is time-consuming and requires raw data to estimate the procedure. In contrast, the frequentist approach based on MLE is faster, and only needs summary statistics for typical models. Future research can explore alternative algorithms that can combine the flexibility of the MCMC and the efficiency of the MLE. The variational inference (e.g., Anderson & Peterson, 1987; Hinton & Van Camp, 1993) based on the Bayesian approach is promising, which can inherit many of MCMC's benefits and balance computational efficiency and accuracy at the same time. Recently, this algorithm had been introduced under the FA and structural equation modeling context (Dang & Maestrini, 2022; Khan et al., 2010), which can provide a basis for its implementation under the revised GPCFA framework. From the structure perspective, extensions of the GPCFA to

Table 6. Parameter Estimates of the PISA Reading Assessment.

Item	F _{gl}	F _{s1}	F _{s2}	F _{s3}	F _{s4}	F _{s5}
1	0.543	0.339				
2	0.625	0.406		<u>0.118^a</u>		
3	0.752	0.448				
4	0.608	0.324				
5	0.555	0.287				
6	0.656		0.141 ^a			
7	0.710		0.219			
8	0.560					
9	0.540		0.393			
10	0.515		0.265			
11	0.684			0.128 ^a		
12	0.681	<u>0.112^a</u>		0.131 ^a		
13	0.647	<u>0.107^a</u>		0.221 ^a		
14	0.639			0.288 ^a		
15	0.648			0.195 ^a		
16	0.645				0.187 ^a	
17	0.737				0.276	
18	0.765				0.263 ^a	
19	0.568				0.296	
20	0.647					0.282
21	0.803					
22	0.829					0.106 ^a
23	0.750					0.319
24	0.600					0.486
25	0.691				<u>0.139^a</u>	
26	0.639					<u>0.100^a</u>
D		0.102	0.058	0.050	0.062	0.076

Note. F_{gl}: Reading literacy; F_{s1} - F_{s5}: 5 different articles; D: effect size; only significant and above .1 loadings are presented; underscored are unspecified loadings.

^aIndicates non-significant loadings.

Table 7. Different CCFA Models' Fitness for the PISA.

Model	RMSEA	CFI	TLI	SRMR	CHI ²	DF
M0	0.023	0.990	0.988	0.039	431.199	275
M1	0.017	0.995	0.994	0.035	352.562	272

Note. RMSEA = root mean square error of approximation; CCFA = categorical confirmatory factor analysis; M0 = baseline; M1 = GPCFA suggested.

accommodate more variants such as the three-parameter testlet effect model are worth exploring. Further research can also empower the revised GPCFA framework for research design in complex settings by incorporating both structural and measurement components. By regularizing different structural and measurement parametric matrices flexibly, one is allowed to create many more partially confirmatory designs that can be used for different purposes. Finally, more empirical

evidence across different real-life scenarios is still desirable to demonstrate the capacity of GPCFA to accommodate various special effects in practice.

Acknowledgments

We wish to thank the editor and reviewers for their helpful comments on the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a General Research Fund Grant (17603022) from the Hong Kong Research Grants Council.

ORCID iDs

Yifan Zhang  <https://orcid.org/0009-0002-7345-2561>

Jinsong Chen  <https://orcid.org/0000-0002-0157-5469>

Supplemental Material

Supplemental material for this article is available online.

References

- Anderson, J. R., & Peterson, C. (1987). A mean field theory learning algorithm for neural networks. *Complex Systems*, 1(5), 995–1019. https://doi.org/10.1007/978-94-011-5014-9_20
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64(2), 153–168. <https://doi.org/10.1007/bf02294533>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Camilli, G. (1994). Teacher's corner: Origin of the scaling constant $d = 1.7$ in item response theory. *Journal of Educational Statistics*, 19(3), 293–295. <https://doi.org/10.3102/10769986019003293>
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Chen, J. (2020). A partially confirmatory approach to the multidimensional item response theory with the Bayesian Lasso. *Psychometrika*, 85(3), 738–774. <https://doi.org/10.1007/s11336-020-09724-3>
- Chen, J. (2021). A generalized partially confirmatory factor analysis framework with mixed Bayesian Lasso methods. *Multivariate Behavioral Research*, 57(6), 879–894. <https://doi.org/10.1080/00273171.2021.1925520>
- Chen, J. (2022). Partially confirmatory approach to factor analysis with Bayesian learning: A LAWBL Tutorial. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(5), 800–816. <https://doi.org/10.1080/10705511.2022.2039660>
- Chen, J., & de la Torre, J. (2014). A procedure for diagnostically modeling extant large-scale assessment data: The case of the Programme for International Student Assessment in reading. *Psychology*, 5(18), 1967–1978. <https://doi.org/10.4236/psych.2014.518200>

- Chen, J., Guo, Z., Zhang, L., & Pan, J. (2021). A partially confirmatory approach to scale development with the Bayesian Lasso. *Psychological Methods*, 26(2), 210. <https://doi.org/10.1037/met0000293>
- Dang, K.-D., & Maestrini, L. (2022). Fitting structural equation models via variational approximations. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(6), 839–853. <https://doi.org/10.1080/10705511.2022.2053857>
- Dinero, T. E., & Haertel, E. (1977). Applicability of the Rasch model with varying item discriminations. *Applied Psychological Measurement*, 1(4), 581–592. <https://doi.org/10.1177/014662167700100413>
- Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data.
- Esra, O., & Atar, H. Y. (2021). Examination of wording effect of the TIMSS 2015 mathematical self-esteem scale through the bifactor models. *International Journal of Assessment Tools in Education*, 8(2), 326–341. <https://doi.org/10.21449/ijate.718670>
- Fukuhara, H., & Kamata, A. (2011). A bifactor multidimensional item response theory model for differential item functioning analysis on testlet-based items. *Applied Psychological Measurement*, 35(8), 604–622. <https://doi.org/10.1177/0146621611428447>
- Geiser, C., & Simmons, T. G. (2021). Do method effects generalize across traits (and what if they don't)? *Journal of Personality*, 89(3), 382–401. <https://doi.org/10.1111/jopy.12625>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. CRC Press.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57(3), 423–436. <https://doi.org/10.1007/bf02295430>
- Giordano, C., & Waller, N. G. (2020). Recovering bifactor models: A comparison of seven methods. *Psychological Methods*, 25(2), 143–156. <https://doi.org/10.1037/met0000227>
- Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In Proceedings of the sixth annual conference on Computational learning theory. ACM.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1), 41–54. <https://doi.org/10.1007/bf02287965>
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, 76(4), 537–549. <https://doi.org/10.1007/s11336-011-9218-4>
- Khan, M. E. E., Bouchard, G., Murphy, K. P., & Marlin, B. M. (2010). Variational bounds for mixed-data factor analysis. In *Advances in Neural Information Processing Systems*, 23. MIT Press.
- Kyriazos, T. A. (2018). Applied psychometrics: The application of CFA to multitrait-multimethod matrices (CFA-MTMM). *Psychology*, 9(12), 2625–2648. <https://doi.org/10.4236/psych.2018.912150>
- Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30(1), 3–21. <https://doi.org/10.1177/0146621605275414>
- Marsh, H. W., & Byrne, B. M. (1993). Confirmatory factor analysis of multitrait-multimethod self-concept data: Between-group and within-group invariance constraints. *Multivariate Behavioral Research*, 28(3), 313–449. https://doi.org/10.1207/s15327906mbr2803_2
- Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37(1), 48–75. [https://doi.org/10.1016/s0092-6566\(02\)00534-2](https://doi.org/10.1016/s0092-6566(02)00534-2)
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Muthén, B., & Muthén, L. (2017). Mplus. In *Handbook of item response theory* (pp. 507–518). Chapman and Hall/CRC.
- Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>

- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559. <https://doi.org/10.1080/00223891.2010.496477>
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47(3), 361–372. <https://doi.org/10.1111/j.1745-3984.2010.00118.x>
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53–61. <https://doi.org/10.1007/bf02289209>
- Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393–408. <https://doi.org/10.1007/bf02294363>
- R Development Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.
- Wang, W.-C., & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, 29(2), 126–149. <https://doi.org/10.1177/0146621604271053>
- Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *ETS Research Report Series*, 26(1), 109–128. <https://doi.org/10.1002/j.2333-8504.2002.tb01869.x>
- Wang, Y., Kim, E. S., Dedrick, R. F., Ferron, J. M., & Tan, T. (2018). A multilevel bifactor approach to construct validation of mixed-format scales. *Educational and Psychological Measurement*, 78(2), 253–271. <https://doi.org/10.1177/0013164417690858>
- Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64(2), 113–128. <https://doi.org/10.1007/bf02294531>
- Zhan, P., Li, X., Wang, W.-C., Bian, Y., & Wang, L. (2015). The multidimensional testlet-effect cognitive diagnostic models. *Acta Psychologica Sinica*, 47(5), 689–701. <https://doi.org/10.3724/SP.J.1041.2015.00689>
- Zhan, P., Wen-Chung, W., Wang, L., & Li, X. (2014). The multidimensional testlet-effect Rasch model. *Acta Psychologica Sinica*, 46(8), 1208. <https://doi.org/10.3724/sp.j.1041.2014.01208>