Properties of Second-Order Exponential Models as Multidimensional Response Models

Carolyn J. Anderson and Hsiu-Ting Yu

Abstract Second-order exponential (SOE) models have been proposed as item response models (e.g., Anderson et al., J. Educ. Behav. Stat. 35:422–452, 2010; Anderson, J. Classif. 30:276–303, 2013. doi: 10.1007/s00357-00357-013-9131-x; Hessen, Psychometrika 77:693–709, 2012. doi:10.1007/s11336-012-9277-1 Holland, Psychometrika 55:5–18, 1990); however, the philosophical and theoretical underpinnings of the SOE models differ from those of standard item response theory models. Although presented as reexpressions of item response theory models (Holland, Psychometrika 55:5–18, 1990), which are reflective models, the SOE models are formative measurement models. We extend Anderson and Yu (Psychometrika 72:5–23, 2007) who studied unidimensional models for dichotomous items to multidimensional models for dichotomous and polytomous items. The properties of the models for multiple latent variables are studied theoretically and empirically. Even though there are mathematical differences between the second-order exponential models and multidimensional item response theory (MIRT) models, the SOE models behave very much like standard MIRT models and in some cases better than MIRT models.

Keywords Dutch Identity • Log-multiplicative association models • Formative models • Reflective models • Composite indicators • Skew normal • Bi-variate exponential

1 Introduction

Philosophical, theoretical, and empirical differences between second-order exponential (SOE) models and multidimensional item response theory (MIRT) models exist; however, these differences that have not been fully discussed nor

---

C.J. Anderson (✉)
University of Illinois, Urbana-Champaign, Champaign, IL, USA
e-mail: cja@illinois.edu; http://faculty.education.illinois.edu/cja/homepage

H.-T. Yu
National Chengchi University, Taipei City, Taiwan
e-mail: hsiutingyu@gmail.com

L.A. van der Ark et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 196, DOI 10.1007/978-3-319-56294-0_2

widely recognized in the literature on SOE models are derived based on the Dutch Identity (Holland 1990; Hessen 2012). Equivalent to SOE models, log-multiplicative association (LMA) models were derived as latent variable models from statistical graphical models (Anderson and Vermunt 2000), as well as from item response models using rest scores in lieu of the latent variables (Anderson and Yu 2007; Anderson et al. 2010). Anderson and Yu (2007) studied unidimensional LMA models for dichotomous data. The LMA models are formative measurement models, and they are item response models in their own right. A better understanding of the properties of LMA models as item response models leads to implications regarding the use and performance of LMA models for analyzing response data. The LMA models have a number of advantages, including maximum likelihood estimation does not require an assumption for the marginal distribution of the latent variables and the models can be fit directly to response patterns using Newton-Raphson. The goal of this paper is to extend Anderson and Yu (2007) to study the properties of multidimensional LMA (or equivalently SOE) models for dichotomous and polytomous items.

Holland (1990) proposed and used the Dutch Identity to derive SOE models for data based on underlying uni- and multidimensional IRT models for dichotomous items. The SOE models are equivalent to LMA models, which are special cases of a log-linear model with two-way interactions. Hessen (2012) extended the Dutch Identity to polytomous items and derived an LMA model; however, he focused on models analogous to the partial credit model (i.e., models in the Rasch family), even though his extension of the Dutch Identity is more general. For models in the Rasch family, category scores are set to fixed values (e.g., consecutive integers). Hessen (2012) mentioned that the category scores could be treated as parameters and estimated. We treat category scores as parameters that are estimated. We extend and generalize the results in Anderson and Yu (2007) and Hessen (2012) to the case of multidimensional models for dichotomous and polytomous items. We highlight the philosophical, theoretical, and empirical differences between LMA and MIRT models.

In the first section of this paper, we discuss the philosophical and theoretical differences between standard MIRT and LMA models. In the following two sections, two properties of LMA are theoretically and empirically studied: the downward collapsibility of LMA models and the effect of different marginal distributions of the latent variables on the models' performance. We conclude with a discussion the potential uses of LMA models in measurement contexts.

2 Reflective and Formative Models

The differences between reflective and formative latent variable models have been discussed by Markus and Borsboom (2013), Bollen and Bauldry (2011), and others. Our intent here is to show the philosophical differences between LMA and MIRT models and how they lead to different mathematical models.

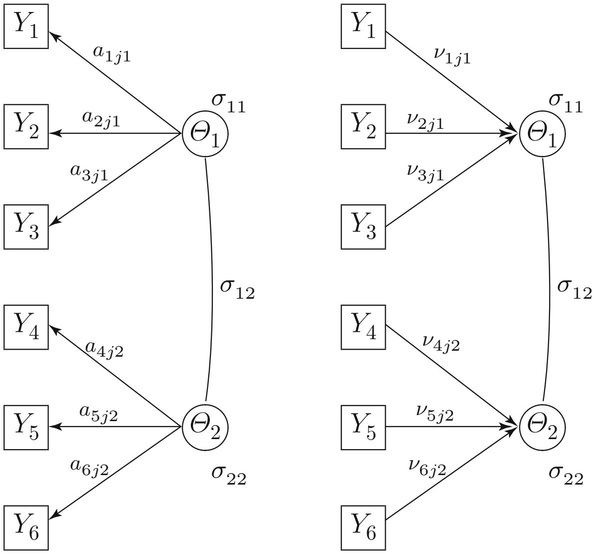

Fig. 1 Graphs corresponding to reflective (*left*) and formative (*right*) models for six items and two latent continuous variables

A reflective model posits that latent variables are prior to behavior, and the latent variables are conceived of as existing whether they are measured or not. A reflective model is illustrated by the graph on the left in Fig. 1. The values on the latent variables lead to observed responses; therefore, behavior indicates or reflects a person's value on the unobserved quantity. A change in the value of a latent variable causes a change in the response behavior. The items are *effect indicators* (Bollen and Bauldry 2011).

To algebraically take into account the directional nature of the relationship between $\boldsymbol{\theta}$ and $\boldsymbol{y}$, models are developed by writing the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{y}$ as $f(\boldsymbol{y}, \boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})$. For a MIRT model, the marginal distribution of the latent variables $f(\boldsymbol{\theta})$ is typically assumed to be multivariate normal, and the distribution for the responses conditional on the latent variables $f(\boldsymbol{y}|\boldsymbol{\theta})$ is a product of multinomial logistic regression models. The model for responses to items is found by numerically integrating over the latent variables; that is, the probability of response pattern $\boldsymbol{y}$ is

$$P(\boldsymbol{y}) = \int_{\theta_1} \cdots \int_{\theta_M} \prod_{i=1}^{I} \frac{\exp[\beta_{ij} + \sum_m \alpha_{ijm}\theta_m]}{\sum_h \exp[\beta_{ih} + \sum_m \alpha_{ihm}\theta_m]} f(\boldsymbol{\theta})d(\boldsymbol{\theta}), \tag{1}$$

where β_{ij} is a location parameter for response option j of item i, and α_{ijm} is the slope parameter for response option j of item i on latent variable θ_m.

In a formative model, the direction of the relationship between $\boldsymbol{\theta}$ and $\boldsymbol{y}$ is reversed relative to the reflective model. A graph representing a formative model is illustrated on the right in Fig. 1. Items define and give meaning to latent variables. The items are *composite indicators* because $\boldsymbol{\theta}$ are composites of the values of the items (Bollen and Bauldry 2011). The joint distribution of $\boldsymbol{y}$ and $\boldsymbol{\theta}$ is found by first specifying the distribution for $f(\boldsymbol{y})$ and then the distribution for $f(\boldsymbol{\theta}|\boldsymbol{y})$; that is, $f(\boldsymbol{y}, \boldsymbol{\theta}) = f(\boldsymbol{\theta}|\boldsymbol{y})f(\boldsymbol{y})$. Assuming that $f(\boldsymbol{y})$ is multinomial and $f(\boldsymbol{\theta}|\boldsymbol{y})$ is a homogeneous conditional, Gaussian distribution leads to an LMA model for the probabilities of observed response patterns $\boldsymbol{y}$ (Anderson and Vermunt 2000; Anderson et al. 2010). The model for data is

$$P(\boldsymbol{y}) = \exp\left[\lambda + \sum_{i=1}^{I} \lambda_{ij} + \sum_{i}\sum_{k \geq i}\sum_{m}\sum_{m' \neq m} \sigma_{mm'} \nu_{ijm} \nu_{kjm'}\right], \qquad (2)$$

where λ ensures probabilities sum to 1 over response patterns, λ_{ij} is the marginal effect term for response option j to item i, ν_{ijm} is the category scale value for response j to item i on latent variable m, and $\sigma_{mm'}$ is a within response pattern variance or covariance of the latent variable(s). The λ_{ij}s and ν_{ijm}s in (2) are analogous to the β_{ij}s and α_{ijm}s, respectively, in (1). Based on the LMA model, the conditional means of the latent variables given $\boldsymbol{y}$ equal:

$$E(\theta_m|\boldsymbol{y}) = \sum_{m'=1}^{M} \sigma_{mm'} \left(\sum_{i=1}^{I} \nu_{ijm'} \right). \qquad (3)$$

Models (1) and (2) are very general models. In this paper, we study the case where each item is directly related to one and only one latent variable, that is, $\nu_{ijm} \neq 0$ and $\alpha_{ijm} \neq 0$ for one and only one m. We expect that the results we find will be the same for more complex models, but we leave this for future study.

The MIRT model given in (1) is not only philosophically different but mathematically different from the LMA model given in (2).

3 Downward Collapsibility of LMA Models

If an item is dropped from data generated from a MIRT model, the data excluding the item still follow a MIRT model and theoretically yield the same estimates of item parameters for the remaining items. If an item is dropped from (or added to) an LMA model, the resulting model is a different model with different parameter estimates. We theoretically and empirically study the effect on LMA model parameter estimates when dropping an item from data (i.e., collapse data over an item). In the first section, we consider the case when data are generated from an LMA model (not collapsible), and in the second section, we consider the case when data are generated from a MIRT model (downward collapsible).

3.1 LMA-Generated Data

Suppose that item 1 is directly related to θ_1 and it is dropped from the data. Let $\boldsymbol{y}_{-1}$ indicate the data excluding item 1. Rather than (3), the conditional means for θ_1 and θ_m are

$$E(\theta_1|\boldsymbol{y}_{-1}) = \sigma_{11}\sum_{i\neq 1}\nu_{ij1} + \sum_{m>1}\left(\sigma_{1m}\sum_{k}\nu_{kjm}\right) + \sigma_{11}\sum_{j}\nu_{1j1}P(Y_1 = j|\boldsymbol{y}_{-1})$$

and

$$E(\theta_m|\boldsymbol{y}_{-1}) = \sigma_{1m}\sum_{i\neq 1}\nu_{ij1} + \sum_{m'>1}\left(\sigma_{mm'}\sum_{k}\nu_{kjm'}\right) + \sigma_{1m}\sum_{j}\nu_{1j1}P(Y_1 = j|\boldsymbol{y}_{-1}),$$

respectively. The last term in each of these equations for the conditional means is unobserved and equals the expected biases of the means due to dropping item 1.

Dropping an item that is directly related to θ_1 changes the conditional variances of θ_1 and any θ_m directly related to θ_1 (i.e., $\sigma_{1m} \neq 0$). In particular, the conditional variances after collapsing over item 1 are

$$\mathrm{var}(\theta_1|\boldsymbol{y}_{-1}) = \sigma_{11} + \sigma_{11}^2\left(\sum_{j}\nu_{ij1}^2 P(Y_1 = j|\boldsymbol{y}_{-1}) - \left(\sum_{j}\nu_{ij1}P(Y_1 = j|\boldsymbol{y}_{-1})\right)^2\right),$$

and

$$\mathrm{var}(\theta_m|\boldsymbol{y}_{-1}) = \sigma_{mm} + \sigma_{1m}^2\left(\sum_{j}\nu_{ij1}^2 P(Y_1 = j|\boldsymbol{y}_{-1}) - \left(\sum_{j}\nu_{ij1}P(Y_1 = j|\boldsymbol{y}_{-1})\right)^2\right).$$

The conditional variances will increase for larger values of σ_{11} and σ_{1m}. The change of $\mathrm{var}(\theta_m|\boldsymbol{y}_{-1})$ is smaller than that for $\mathrm{var}(\theta_1|\boldsymbol{y}_{-1})$ because $\sigma_{1m}^2 \leq \sigma_{11}^2$. Regardless of the value of σ_{11} and σ_{1m}, the conditional means and variances are affected the most when an item with the largest values of ν_{ij1} is dropped, and they are least affected when the item with the smallest values of ν_{ij1} is dropped.

Our interest is in the theoretical behavior of the LMA models; therefore, $P(\boldsymbol{y})$s were computed from an LMA (six items, three response options per item), so the LMA model fits the data perfectly. The size of the scale values for an item was measured by $\sum_j \nu_{ijm}^2$. Two additional data sets were created by collapsing over the item with the smallest value and the largest value of $\sum_j \nu_{ijm}^2$. The item with the weakest relationship to a θ_m should have the smallest effect on the results, and collapsing over the item with the strongest relationship to a θ_m should have the largest effect.

Throughout this paper, maximum likelihood estimation was used to estimate parameters of LMA and MIRT models. The LMA models were fit to data using SAS$^{\text{ő}}$ PROC NLP (version 9.4, SAS Institute Inc. 2015). The MIRT models were fit to data using *flexMIRT* (Houts and Cai 2013) assuming bivariate (multivariate) normality.[1]

In terms of goodness of fit, the likelihood ratio goodness-of-fit statistic (G^2) is used as an index but is not compared to a χ^2 distribution because there is no sampling variability. As a second index, we used the dissimilarity index:

$$D = \sum_{\boldsymbol{y}} \frac{|P(\boldsymbol{y}) - \hat{P}(\boldsymbol{y})|}{2},$$

where the sum is over all response patterns, $P(\boldsymbol{y})$ is the probability of response pattern $\boldsymbol{y}$, and $\hat{P}(\boldsymbol{y})$ is the fitted value of the probability of response pattern $\boldsymbol{y}$ from a model. The index D is interpretable as the proportion of data that would have to be moved from one response pattern to another for the model to fit perfectly (Agresti 2013).

Any misfit of the LMA model fit to the six items is due to numerical inaccuracy in the data generation and/or model estimation. The LMA model fits the probabilities of response patterns for the six items nearly perfectly. When collapsing over the weak item, the parameter estimates and goodness-of-fit statistics of the LMA model were nearly identical to those when the model was fit to all six items. Specifically, collapsing over the weak item had a smaller impact on the goodness of fit than collapsing over the strong item (i.e., $G^2 = 0.0000$ versus $G^2 = 0.0002$ and $D = 0.0002$ versus $D = 0.0049$). All of the LMA models fit the probabilities better than all of the MIRT models.

When collapsing over the strong item, there were noticeable differences between the estimated parameters from the LMA model fit to those used to generate the data. The variance of θ_m increased the most when the item dropped is the strong item. Specifically, when the strong item is dropped, the variance of the latent variable to which it is connected goes from 0.87 to 1.66, but when the weak item is dropped, the variance of the latent variable to which it is connected goes from 0.77 to 0.88. As predicted, both $\hat{\sigma}_{11}$ and $\hat{\sigma}_{22}$ increased when collapsing over either the weak or strong item. The change in both variances occurs because when we collapse over an item related to, say θ_1, leads to less information to estimate the latent variable θ_2, which increases uncertainty (i.e., larger σ_{22}).

[1]Files containing code and data that reproduce all analyses can be downloaded from http://faculty.education.illinois.edu/cja/homepage.

3.2 MIRT-Generated Data

If θ_1 and θ_2 were discrete, then we could collapse over, for example, item 1 and expect $\hat{\nu}_{ijm}$ for $i \neq 1$ to remain the same. Since for the LMA models $\hat{\boldsymbol{\theta}}$ equals a weighted sum of category scores, $\hat{\boldsymbol{\theta}}$ is empirically discrete and LMA models might be collapsible. When data are generated from a model that implies collapsibility, whether LMA scale values are affected by dropping items is an open question. Since MIRT models imply collapsibility, probabilities were generated from a two-dimensional MIRT model with $\boldsymbol{\theta} \sim MVN(\boldsymbol{\mu} = (0, 0), \rho = 0.5)$ for eight items where items 1–4 were related to θ_1, and items 5–8 were related to θ_2. The generated probabilities were collapsed over one item at a time until there were only four items remaining. We alternated collapsing over an item related to θ_1 and one related to θ_2.

Since LMA models are formative measurement models, we are primarily interested in the $\hat{\nu}_{ijm}$s, which are used to compute estimates of the conditional means of the latent variables (i.e., $\hat{\boldsymbol{E}}(\theta_m|y)$). The scale values $\hat{\nu}_{ijm}$ were essentially unaffected by collapsing the data. When data were collapsed over an item, both of the $\hat{\sigma}_{mm}$s increased. When the first item was dropped, which was related to θ_2, the increase of $\hat{\sigma}_{22}$ was greater than that for $\hat{\sigma}_{11}$. When the second item was dropped, which was related to θ_1, the increase in $\hat{\sigma}_{11}$ was greater than that for $\hat{\sigma}_{22}$. This pattern continued until there are only four items remaining.

In sum, if data are generated from a MIRT model, which collapsibility, then the LMA model yields nearly the same $\hat{\nu}_{ijm}$s when items are dropped. Conversely, we can consider adding items. If the data come from a model that implies collapsibility and then when adding items (assuming that the added items are related to the underlying latent variable(s)), the $\hat{\nu}_{ijm}$s are not expected to change, and $\hat{\sigma}_{mm}$s are expected to be smaller.

4 Different Marginal Distributions

A property often given as an advantage of LMA models is that a marginal distribution of the latent variables is a mixture of normals, which can take on many different shapes. The goal of this section is to determine whether and when an LMA model may perform well in terms of goodness of fit and parameter recovery and compare LMA model performance with a corresponding MIRT model.

In this study, we generated probabilities for response patterns by numerically integrating out the latent variables from a MIRT model assuming one of four different underlying distributions. The multivariate normal (MVN) was chosen because this is the typical assumption made when fitting a MIRT model. The multivariate skew normal was chosen because the MVN is a special case of the skew normal, and there has been some interest in using the skew normal as an alternative to the normal distribution (Azevedo et al. 2011; Casabianca and Junker 2016; Lee 2002). Marshall-Olkin bivariate exponential distribution (Mardia et al. 1979) was chosen because some variables in the data that we often analyze are very skewed. Lastly, a mixture of two multivariate normal distributions was chosen to mimic a situation where individuals have opposite attitudes or views. This also reflects a situation where there is an important group variable that has not been included in the model, and there is differential item functioning.

As the number of items increases to ∞, $\sigma_{mm} \longrightarrow 0$, an LMA model will yield the actual marginal distribution of the latent variables. The behavior of LMA models for short tests or subscales is less certain; therefore, we empirically examine the behavior of the models when fit to generated data for short tests. Probabilities of response patterns were generated using the MIRT model in (1) for $M = 2$ latent variables and $I = 4$ or 6 items with $J = 2$, 3, or 4 response options. For the multivariate normal distribution, we also fit models to data with $M = 3$ and $I = 6$ items.

Both the MIRT and the LMA models were fit to all of the data sets. Albeit naive, when the distribution generating the data is not normal, the MIRT models were fit to data assuming multivariate normality. Although not reported here, two additional models were fit as baseline models: the log-linear model of independence and the homogeneous (all two-way interaction) log-linear model. The probabilities of response patterns were multiplied by 1,000,000 to retain more decimal places and accuracy. Besides the dissimilarity index D, a second measure of goodness of fit is reported for the models: the percent of association accounted for by a model,

$$\text{Percent association} = \frac{G^2_{independence} - G^2_{model}}{G^2_{independence}} \times 100\%,$$

where the likelihood ratio statistic G^2 from the independence model is a measure of the amount of association in the data.

To examine parameter recovery, we used the correlation between the parameters used to generate the data and the estimated parameters from the LMA and MIRT models. Given our focus on LMA models, we are primarily interested in the estimation of the ν_{ijm} parameters. The marginal effect terms λ_{ij} generally are viewed as nuisance parameters from an LMA model framework, but the correlations for marginal terms are reported for the sake of completeness.

The results for different numbers of items and response options are all very similar; therefore, we only report the result for one case (i.e., six items, three response options, and two latent variables). Goodness-of-fit statistics are reported in Table 1, and correlations between estimated parameters and those used to generate the data are reported in Table 2.

When data were generated using the bivariate normal distribution ($\boldsymbol{\mu} = \mathbf{0}$, $\rho = 0.5$), the MIRT model should fit perfectly. Any misfit is due to numerical inaccuracy in generating the probabilities and/or estimating the model. The MIRT models essentially fit perfectly; however, the goodness-of-fit indices for the LMA models are just shy of perfect. When data were generated using a skew normal (i.e., $\boldsymbol{\mu} = \mathbf{0}$, $\rho = 0.75$, and shape parameters 2 and 3) or a bivariate exponential distribution (i.e.,

Table 1 Goodness-of-fit statistics for LMA and MIRT models fit to date generated from a MIRT model with different underlying distributions for $f(\boldsymbol{\theta})$

	Dissimilarity		Percent association	
Underlying distribution	LMA	MIRT	LMA	MIRT
Bivariate normal	0.0016	0.0002	99.99	100.00
Skew normal	0.0268	0.0268	97.14	97.16
Bivariate exponential	0.0127	0.0129	98.44	98.39
Mixture of normals	0.0346	0.0708	99.43	96.05

Table 2 Correlations between LMA and MIRT model parameter estimates and parameters used generated MIRT model probabilities for different $f(\boldsymbol{\theta})$s

	LMA	MIRT	LMA	MIRT
Underlying distribution	$r(\alpha_{ijm}, \hat{\nu}_{ijm})$	$r(\alpha_{ijm}, \hat{a}_{ijm})$	$r(\beta_{ij}, \hat{\lambda}_{ij})$	$r(\beta_{ijm}, \hat{b}_{ij})$
Bivariate normal	0.9980	1.0000	0.9839	1.0000
Skew normal	0.9950	0.9962	0.8361	0.7506
Bivariate exponential	0.9326	0.9077	0.8257	0.7894
Mixture of normals	0.9971	0.9430	0.9665	0.9428

$f(\boldsymbol{\theta}) = \exp(-1.0\theta_1 - 0.5\theta_2 - 0.2\max(\theta_1, \theta_2))\kappa$ where κ normalized the function), the LMA and MIRT models both provide good representations of the data, and there are no systematic differences in terms of which model fits the data better. When data were generated from the mixture of two normals (i.e., $\boldsymbol{\mu}_1 = (-2, -2)'$, $\boldsymbol{\mu}_2 = (2, 2)'$, $\rho = 0.5$, and mixing weight of 0.5), the LMA models clearly fit the data better than the MIRT models.

More differences between the models' performance were found in terms of parameter recovery. When data were from the bivariate normal, MIRT parameters are perfectly correlated with those used to generate the data; however, the LMA parameters were just short of perfect. For the skew normal, the correlations between the α_{ijm}s used to generate the data and the estimated ν_{ijm}s parameters from the LMA models were about the same as the corresponding correlations of parameters from the MIRT models; however, the correlations for the β_{ijm}s were much larger for the LMA model than the MIRT model. For the exponential and mixture of normal distributions, the correlations for the estimated ν_{ijm}s and λ_{ij}s from the LMA models were considerably larger than those for parameters from the MIRT models.

5 Discussion

The LMA models and standard MIRT models were shown to be philosophically and mathematically different models; however, they share some important properties. For short tests, the LMA models performed nearly as well as standard MIRT models when the underlying distribution of the latent variables is multivariate normal, and the LMA and MIRT models empirically perform equally well when the underlying distribution is skew normal. With the skew normal, the goodness of fit is about the same for both the LMA and MIRT models; however, the estimation of the β_{ij}s parameters had lower correlations with the parameters used to generate the data than the LMA model parameters. The LMA models perform better than MIRT models in terms of goodness of model fit to data and parameter recovery when data arise from an LMA model and when $f(\boldsymbol{\theta})$ follows either a bivariate exponential distribution or a mixture of two normal distributions.

The LMA models are more flexible than discussed in this paper. The LMA models can include covariates for the latent variables, the marginal effect terms (i.e., the λ_{ij}), and the conditional variances and covariances of the latent variables (Anderson 2013). The models also permit various restrictions on parameters, including equality, ordinal, partially ordinal, linear transformations, and/or any desired transformation (Anderson 2013). The LMA models can also represent more complex latent variable structures than those studied in this paper, such as those where items "load" on multiple correlated or uncorrelated latent variables (e.g., bi-factor models). Since the assumptions and theory are the same, we expect the same results for more complex models such as those that we found for the simpler models reported in this paper.

Our focus was on short tests because these are cases where LMA and MIRT models may differ. Although we used common commercial software (SAS) to fit the LMA models to data, one bottleneck to more widespread applications of LMA models is a limitation to the size of the problem that can be handled. The size of the cross-classification of items (i.e., number of response patterns) increases exponentially when adding items and/or categories per item. When scores are input, the pseudo-likelihood method given in Anderson et al. (2007) works well and can be implemented in any program that fits conditional multinomial logistic models. Recently Paek (2016), Paek and Anderson (2017) proposed a solution to the more general problem where scores are estimated. In simulations, Paek (2016) showed that the algorithm yields nearly identical parameter estimates as MLE of LMA models for short tests and that the algorithm recovers parameters used to simulate the data in longer tests (i.e., 20 and 50 items). The more general algorithm also can be implemented in any software program that fits conditional multinomial logistic regression models.

We do not advocate that LMA models replace MIRT models because they are philosophically and theoretically different measurement models. The LMA models actually may be complimentary to applications of MIRT models. Suppose a researcher desires a reflective model but does not know what marginal distribution of the latent variable(s) should be used when fitting a MIRT model to data. The LMA models can be used to estimate the marginal distribution of the latent variables, which could confirm or suggest a distribution to be used when fitting the MIRT model to data.

The empirical studies in this paper imply that one cannot conclusively determine whether the model should be formative or reflective. Whether one performs better than the other is an empirical question. The choice between using an LMA model or a MIRT model for a particular case depends on a researcher's conceptualization of the latent variable.

References

A. Agresti, *Categorical Data Analysis,* 3rd edn. (Wiley, New York, 2013)

C.J. Anderson, Multidimensional item response theory models with collateral information as Poisson regression models. J. Classif. **30**, 276–303 (2013). doi: 10.1007/s00357-00357-013-9131-x

C.J. Anderson, J.K. Vermunt, Log-multiplicative association models as latent variable models for nominal and/or ordinal data. Sociol. Methodol. **30**, 81–121 (2000)

C.J. Anderson, H.-T. Yu, Log-multiplicative association models as item response models. Psychometrika **72**, 5–23 (2007)

C.J. Anderson, Z. Li, J.K. Vermunt, Estimation of models in a Rasch family for polytomous items and multiple latent variables. J. Stat. Softw. **20** (2007). http://www.jstatsoft.org/v20/i06/v20i06.pdf

C.J. Anderson, J.V. Verkuilen, B.L. Peyton, Modeling polytomous item responses using simultaneously estimated multinomial logistic regression models. J. Educ. Behav. Stat. **35**, 422–452 (2010)

C.L.N. Azevedo, H. Bolfarine, D.F. Andrade, Bayesian inference for a skew-normal IRT model under the centred parameterization. Comput. Stat. Data Anal. **55**, 353–365 (2011)

K.A. Bollen, S. Bauldry, Three C's in measurement models: causal indicators, composite indicators, and covariates. Psychol. Methods **16**, 265–284 (2011)

J.M. Casabianca, B.W. Junker, Multivariate normal distribution, in *Handbook of Item Response Theory, Volume Two: Statistical Tools*, ed. by W.J. van der Linden (Talyor & Fransics/CRC Press, Boca Raton, 2016), pp. 35–46

D.J. Hessen, Fitting and testing conditional multinomial partial credit models. Psychometrika **77**, 693–709 (2012). doi:10.1007/s11336-012-9277-1

P.H. Holland, The Dutch identity: a new tool for the study of item response models. Psychometrika **55**, 5–18 (1990)

C.R. Houts, L. Cai, *flexMIRT: Flexible Multilevel Multidimensional Item Analysis and Test Scoring* (Vector Psychometric Group, LLC, Chapel Hill, 2013)

J. Lee, Multidimensional item response theory: an investigation of interaction effects between factors on item parameter recovery using Markov Chain Monte Carlo. Unpublished Doctoral Dissertation, Michigan State University (2002)

K.V. Mardia, J.M. Kent, J.M. Bibby, *Multivariate Analysis* (Academic, Orlando, 1979)

K.A. Markus, D. Borsboom, *Frontiers of Test Validity Theory: Measurement, Causation and Meaning* (Routlege, New York, 2013)

Y. Paek, Pseudo-likelihood estimation of multidimensional polytomous item response theory models. Unpublished Doctoral Dissertation, University of Illinois at Urbana-Champaign (2016)

Y. Paek, C.J. Anderson, Pseudo-likelihood estimation of multidimensional response models: polytomous and dichotomous items, in *The 81st Annual Meeting of the Psychometric Society, Asheville, NC*, ed. by L.A. van der Ark, S.A. Culpepper, J.A. Douglas, W.C. Wang, M. Wiberg (Springer, New York, 2017)

SAS Institute Inc., *Statistical Analysis System*, version 9.4 (SAS Institute, Cary, 2015)

http://www.springer.com/978-3-319-56293-3

Quantitative Psychology
The 81st Annual Meeting of the Psychometric Society, Asheville, North Carolina, 2016
van der Ark, L.A.; Wiberg, M.; Culpepper, S.A.; Douglas, J.A.; Wang, W.C. (Eds.)
2017, IX, 430 p. 49 illus., 24 illus. in color., Hardcover
ISBN: 978-3-319-56293-3