

3

Systematic Review on Simulation Studies of Latent Interaction Models

The exploration of interaction effects in psychological research is increasingly popular. A few reviewing articles in family studies, business management, and school psychology, have been published particularly for instructing researchers to test and interpret interaction effects (Whisman & McClelland, 2005; Dawson, 2014; Fairchild & McQuillin, 2010). The interaction effect examines how a third variable, known as a moderator, alters the relationship between a predictor and an outcome variable, demonstrating the complexity of how individual traits, environmental factors, and situational contexts interact. Fairchild and McQuillin (2010) clarified that the terms “moderation” and “interaction” are often used interchangeably since both terms imply how a third variable can modify relations between two variables. One of the predominant approaches for conducting interaction analysis is through moderated multiple regression (MMR), where the regression model incorporates a product term composed of its explanatory variable and moderator, and is usually estimated using a least squares method (Aiken & West, 1991; Newsom et al., 2003; Yuan et al., 2014). Considering the most parsimonious MMR model with one predictor (X), one moderator (Z), and one criterion variable (Y):

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \cdot Z + \epsilon, \quad (1)$$

where β_0 is the intercept, β_1 is the regression coefficient for X , and β_2 is the regression coefficient for Z . $X \cdot Z$ is the product (or interaction) term created by multiplying the first-order variables X and Z and should imply information about the interaction between X and Z . β_3 is the regression coefficient for $X \cdot Z$. ϵ is a normally distributed random residual term and carries information of unaccounted variance not explained by X , Z and $X \cdot Z$. Assuming assumptions of typical multiple regression models are met (Osborne & Waters, 2019; Osborne et al., 2001), a significant interaction effect is determined by rejecting the null hypothesis (i.e., H_0) that $\beta_3 = 0$ through the t test.

Observed variables in classical regression models are assumed to be measured without

errors; however, this assumption usually does not hold true in real practice and empirical research. Measurement errors do play a crucial role in parameter estimation, and they can stem from various sources such as respondent misunderstanding, data entry mistakes, or instrument flaws, and can significantly distort the analysis if not properly addressed (Bollen, 1989). Low reliability (i.e., high amount of measurement error) in the predictor X and moderator Z may introduce bias when estimating the interaction effect β_3 , and subsequently inflated standard error, reduced statistical power, and attenuated effect size (Fisicaro & Lautenschlager, 1992; Carroll et al., 2006; Aguinis et al., 2005). Latent variable models, such as structural equation modeling (SEM), explicitly incorporate measurement errors into their framework, distinguishing between the true scores of the underlying construct and the observed scores that may be contaminated by errors (Jöreskog, 1970). Latent variables represent underlying factors or traits that are indicated by a set of observed variables but cannot be measured directly (Bollen, 1989). For instance, psychological concepts like intelligence or satisfaction are often quantified through various indirect measures, rather than being observed outright.

As for interaction analysis, the issue of measurement errors is more serious. Busemeyer and Jones (1983) mentioned that the reliability of the product term XZ is a function of the reliability of predicting variables (e.g., X and Z) and their correlation. Alternatively speaking, the estimation of interaction effect will be downward biased with a larger degree if predicting variables are highly correlated and contain more measurement errors, and such bias will not be remedied by increased sample size. Hence, an increasing number of latent variable models with interaction (i.e., latent interaction model) have been proposed to account for measurement error and produce more accurate parameter estimation.

In this section, I will provide a brief review of latent interaction models based on their assumptions, variable types, and estimation methods. Then, I will introduce the Study 1 of my dissertation proposal, in which I will conduct a systematic review of simulation studies

that have investigated each method of estimating latent interaction effects.

Development of Latent Interaction Models

Constrained Product Indicator Method (CPI): The Seminal Model by Kenny and Judd (1984)

The idea of modeling latent interaction was first formally proposed by Kenny and Judd (1984) that is commonly recognized as the seminal article on the development of latent interaction methods. The structural model looks similar to MMR, except that it is based on latent variables:

$$y = \alpha + \gamma_x \xi_x + \gamma_m \xi_m + \gamma_{xm} \xi_x \xi_m + \zeta, \quad (2)$$

where α is the constant intercept, ξ_x is the latent predictor, ξ_m is the latent moderator, and $\xi_x \xi_m$ is the latent interaction variable created by multiplying ξ_x and ξ_m . Note that ξ_x and ξ_m are named first-order latent variables as well. γ_x , γ_m , and γ_{xm} are the path coefficients of the first-order variables and the interaction term. ζ is the disturbance term. Each latent variable is assumed to be indicated by at least two observed items and hence follows a typical factor analysis model:

$$x_1 = \lambda_{x_1} \xi_x + \delta_{x_1}, \quad x_2 = \lambda_{x_2} \xi_x + \delta_{x_2}, \quad (3)$$

$$m_1 = \lambda_{m_1} \xi_m + \delta_{m_1}, \quad m_2 = \lambda_{m_2} \xi_m + \delta_{m_2}, \quad (4)$$

where x_1 , x_2 , m_1 , m_2 are observed indicators of ξ_x and ξ_m , two for each respectively. λ s are the factor loadings and δ s are the measurement errors. Kenny and Judd (1984) used the mean-deviation form for all observed indicators and therefore constant terms are omitted in the above equations. In this model, ξ_x and ξ_m are assumed to have multivariate normal distributions with allowed covariance. Based on this assumption, Jöreskog and Yang (1996) demonstrated that the mean of the interaction term $\xi_x \xi_m$ equals to the covariance between ξ_x and ξ_m . Besides, ξ_x , ξ_m , δ s, and ζ are multivariate normal with means of 0 and uncorrelated with each other. Alternatively demonstrating in mathematical symbols,

$$\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\kappa}, \boldsymbol{\Phi}), \quad (5)$$

where $\boldsymbol{\xi}$ is a vector of latent variables, and $\boldsymbol{\kappa}$ and $\boldsymbol{\Phi}$ are the mean vector and covariance matrix of latent variables (Jöreskog and Yang, 1996). Specifically,

$$\boldsymbol{\kappa} = \begin{pmatrix} 0 \\ 0 \\ \phi_{\xi_x \xi_m} \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} \phi_{\xi_x} & & \\ \phi_{\xi_m \xi_x} & \phi_{\xi_m} & \\ 0 & 0 & \phi_{\xi_x \xi_m} + \phi_{\xi_m \xi_x}^2 \end{pmatrix}. \quad (6)$$

To estimate the latent interaction term $\xi_x \xi_m$, Kenny and Judd (1984) suggested to use all possible cross product pairs of first-order latent variables' indicators, namely product indicator (PI), and they are $x_1 m_1$, $x_1 m_2$, $x_2 m_1$, $x_2 m_2$ in this case. Take $x_1 m_1$ as an example, the mathematical component of the PI will be:

$$x_1 m_1 = \lambda_{x_1} \lambda_{m_1} \xi_x \xi_m + \lambda_{x_1} \xi_x \epsilon_{m_1} + \lambda_{m_1} \xi_m \epsilon_{x_1} + \epsilon_{x_1} \epsilon_{m_1}, \quad (7)$$

in which the factor loading and the error term of $x_1 m_1$ can be obviously represented by two functions of parameters from the equations of original indicators. Specifically, the factor loading of $x_1 m_1$ is $\lambda_{x_1} \lambda_{m_1}$ while the error term of $x_1 m_1$ is $\lambda_{x_1} \xi_x \epsilon_{m_1} + \lambda_{m_1} \xi_m \epsilon_{x_1} + \epsilon_{x_1} \epsilon_{m_1}$. It implies that the components of factor loading and error term of the formed PI all come from existing parameters of first-order indicators and they serve as model constraints. As the number of formed PIs increases, the model constraints become extraordinarily complicated and excessive, which leads to a cumbersome model with potential convergence issue. Therefore Kenny and Judd's (1984) model is called constrained product indicator method (CPI).

As many past literature described while Kenny and Judd (1984) did not explicitly mention in their original work, the latent interaction term in nature is the product of two first-order latent predictors, and hence the error terms of PIs indicating the interaction term must be allowed to covary with those of first-order indicators (Cortina et al., 2021; Schoemann & Jorgensen, 2021). Alternatively speaking, the error terms of PIs that share the

98 first-order latent predictors may covary with the corresponding first-order indicators:

$$\begin{matrix} x_1m_1 \\ x_1m_2 \\ x_2m_1 \\ x_2m_2 \end{matrix} \begin{bmatrix} \theta_{x_1m_1} & a & b & \\ & \theta_{x_1m_2} & & b \\ & & \theta_{x_2m_1} & a \\ & & & \theta_{x_2m_2} \end{bmatrix}, \quad (8)$$

99 where the diagonal elements are error variance of each PI. The covaried error terms are
 100 labeled by a and b where the residual covariances are constrained with the same fixed values
 101 for PIs sharing the same first-order indicator. For example, x_1m_1 and x_1m_2 share the
 102 first-order indicator x_1 so that their error variance are constrained to equality.

103 The estimation of this model is based on maximum likelihood (ML) estimator that
 104 builds on normal distributions of observed indicators and latent constructs, and any
 105 non-normality in either observed indicators or latent variables may lead to biased estimates
 106 and underestimated standard errors due to the violation of normality assumption. Besides,
 107 Type I error rates may be inflated and lead to false positive detection of interaction effects
 108 (Jöreskog & Yang, 1996).

109 Although not many substantive research used the original method of Kenny and Judd
 110 (1984) due to complicated model constraints and difficulty in implementation, many later
 111 developed advanced latent interaction methods are mostly based on this seminal model.
 112 Kelava & Brandt (2023) in Chapter 23 of Handbook of Structural Equation Modeling (Hoyle,
 113 2022) categorized usually used latent interaction methods into four types: product indicator
 114 approaches, distribution analytic approaches, Bayesian parametric approaches, and method
 115 of moments approaches. The performance of each approach based on simulation studies and
 116 empirical research results has been summarized in Table 23.3 in the book. Below I will
 117 briefly summarize representative methods of each category of latent interaction models.

Product Indicator (PI) Method

There has been significant progress in PI methods since the foundational work of Kenny and Judd in 1984. These methodologies range from fully latent approaches, which involve multiple indicators for latent products and simultaneous estimation of measurement and structural components, to partially latent approaches that use single indicators (Cortina et al., 2021).

Building upon CPI, Wall and Amemiya (2001) presented that the latent covariance matrix Φ is based on the assumption of normally distributed latent variables. Specifically, it requires that ξ_x and ξ_m to have normal distribution so that the covariance between the first-order latent predictor and the latent interaction term can be 0 (i.e., $Cov[\xi_x, \xi_x \xi_m] = 0$). Besides, the variance of the interaction term can no longer be represented as $\phi_{\xi_x \xi_m} + \phi_{\xi_m \xi_x}^2$ with the violation of normal distribution. Hence, theoretically CPI should not produce unbiased parameter estimates with reliable standard errors with the constraints on latent covariance matrix based on the normality assumptions. With the theoretical reasoning and mathematical derivation, they proposed the Generalized Appended Product Indicator (GAPI) method in which the constraints on Φ are removed and each element in Φ is freely estimated, while the other model constraints remain the same. This innovative approach enhances the capacity to analyze interactions among non-normally distributed latent variables, improving the accuracy of parameter estimates in these contexts.

Marsh et al. (2004) advanced the GAPI approach by further reducing the model constraints and only kept the mean structure in equation (6), allowing for free estimation of parameters that were previously constrained. This modification, while simplifying the model by reducing its restrictions, also leads to a reduction in degrees of freedom due to the increase in freely estimated parameters. One example is that the factor loading and error term in equation (7) are freely estimated parameters instead of functions of existing parameters. Through their research, Marsh et al. (2004) stated that UPI theoretically should

have robustness to the violations of assumptions of multivariate normality since no model constraints based on these assumptions are specified in the UPI model. Their simulation studies effectively challenged the necessity of complicated model constraints in CPI and argued that UPI is simpler and more approachable. However, a few studies have shown that UPI may produce biased parameter standard error estimates with non-normally distributed latent variables and categorical observed indicators (Aytürk et al., 2020; Aytürk et al., 2021; Pieters et al., 2022; Cortina et al., 2021).

Another disadvantage of UPI is that the number of PIs may become unmanageable while the first-order indicators are large. Although Wen et al., (2013) discussed possible methods of reducing the number of PIs using matched-pair UPI and parceling UPI, there remains arbitrariness in selecting first-order PIs, which may lose information from unused PIs. Thus, single product indicator methods have been proposed to address this problem. In an effort to enhance the estimation accuracy of latent interaction effects while maintaining methodological simplicity, Lai and Hsiao (2018) introduced a reliability-adjustment product indicator (RAPI) method, wherein composite scores such as sum or mean scores are used as single indicators (SI). RAPI effectively captures all available information from first-order indicators and addresses the issue of unwieldy number of PIs. To account for measurement error, RAPI imposes error-variance constraints computed from observed first-order indicators on the factor loadings and error variances of composite scores as SIs. The simulation studies in Lai and Hsiao (2018) demonstrated that RAPI was able to produce unbiased estimates of latent interaction effects accompanied by acceptable standard errors under conditions of small sample sizes ($N = 100$) and low reliability ($\rho = .70$) for congeneric items.

The Two-Stage Path Analysis (2S-PA) method, introduced by Lai and Hsiao (2022), is an alternative SI method for estimating latent interaction effects. In contrast to conventional joint SEM practices, which typically require simultaneous estimation of measurement and structural models and demand substantial sample sizes for satisfactory

convergence rates, the 2S-PA method divides the process into two distinct stages. At the first stage, researchers use any appropriate psychometric methods to compute factor scores (e.g., expected-a-posterior scores, regression scores, Bartlett scores); these factor scores are then used as SIs to indicate latent variables at the second stage, and the structural model is estimated for relations between latent variables. This separation of estimation not only enhances the stability of model estimation and reliability of parameter estimation but also makes the method more accessible for studies with smaller sample sizes. A distinctive feature of the 2S-PA approach is its ability to assign specific estimated reliability to each observation, which broadens its usability to incorporate ordered categorical items and to accommodate non-normal distributions. This flexibility marks a significant advancement over traditional methods, addressing a wider range of data types and distributional challenges. Consequently, the 2S-PA method offers a practical and efficient alternative for researchers dealing with the complexities of measurement error and the limitations of traditional SEM in the context of latent variable modeling.

Distribution Analytic Approaches

As mentioned in Kenny and Judd's model (1984), the distribution of the latent interaction variable is not normally distributed even though the latent predictor and moderator have normal distributions. Therefore the assumption of multivariate distribution when using maximum likelihood (ML) estimation in product indicator methods are violated. To address this issue, latent moderated structural equations (LMS; Klein & Moosbrugger, 2000) and quasi-maximum-likelihood (QML; Klein & Muthén, 2007) approach have been proposed, which do not involve PIs. Instead, they directly effectively handle the non-normality of latent dependent variables that may arise from the inclusion of latent interaction terms.

A notable characteristic of the LMS method is that theoretically it does not violate the multivariate assumption and thus it is expected to handle the inherently non-normal

distribution of the interaction variable by generating accurate and precise parameter estimates. Specifically, LMS approximates the likelihood of the non-normal first-order indicator vector (e.g., $[x_1, x_2, m_1, m_2]$) using a finite mixture of conditional normal distributions. This approximation is grounded on the critical insight that the interaction effects stemming from the interaction terms (e.g., $\xi_x \xi_m$) can be examined by using the distribution of the indicator variable across different levels of the latent predictor and moderator that form the interaction term (e.g., ξ_x ; Klein & Moosbrugger, 2000; Marsh et al., 2004). In essence, LMS navigates the complexity of non-normal distributions by segmenting the distribution based on conditioning on a latent component, thus transforming a non-normal distribution into a series of normal distributions. For the estimation of the model's unknown parameters within this mixture distribution framework, LMS employs a modified version of the expectation maximization (EM) algorithm (Dempster et al., 1977), enabling the derivation of maximum likelihood (ML) estimates with enhanced accuracy and efficiency. LMS showed outstanding performance of estimating interaction effects when latent predicting variables are normally distributed, but showed biased estimates with standard errors and elevated Type I error when they are moderately or severely non-normal (Cham et al., 2012; Aytürk et al., 2020; Aytürk et al., 2021; Cortina et al., 2021).

Like LMS, quasi-maximum likelihood (QML) is another distribution analytic approach extends beyond the traditional ML estimation, especially when the strict assumptions of ML are not fully met. The QML method involves estimating the parameters of a model by maximizing a likelihood function that is “quasi” because it is based on an assumed (and possibly incorrect) distribution of first-order latent variables and their observed indicators. The development of the QML approach was motivated by dual objectives, accelerating algorithmic computations and strengthening methodological integrity. Grounded in the quasi-likelihood principle (Wedderburn, 1974), QML seeks to replace the traditional likelihood function with a quasi-likelihood that exhibits analogous characteristics, which facilitates the optimization of the quasi-(log-) likelihood function for the estimation of

parameters and enhances computational efficiency with enhanced robustness against the challenges posed by non-normality in latent variables. Comparative simulation studies reveal that QML and LMS share similar estimation properties, wherein QML demonstrates outstanding performance under normally distributed latent variables. Nonetheless, QML is susceptible to elevated Type I error rates with moderate to severe non-normality of latent predictor variables, as LMS showed (Kelava & Nagengast, 2012).

Bayesian Parametric Approaches

According to Kelava and Brandt's (2023) summary, Arminger and Muthén (1998) first introduced the application of nonlinear latent variable models by employing the Gibbs sampler (Geman & Geman, 1984) alongside the Metropolis–Hastings algorithm (Hastings, 1970). A key element of their methodology involved the use of prior distributions for parameter estimation, which are predefined hypotheses about the parameter distributions. Utilizing these priors allows researchers to integrate existing knowledge about the parameters into the analysis. The concept of priors ensures that the resulting posterior distribution of the parameters remains within the same distributional family (e.g., normal), facilitating a theoretically consistent estimation process. By framing the problem within a Bayesian context, the method allows for the incorporation of prior knowledge about the likely size and direction of interactions, enhancing the robustness and interpretiveness of the analyses.

Lee et al. (2007) proposed a generalized structural equation modeling (GSEM) approach that uses Bayesian modeling framework to estimate latent interaction effects. This Bayesian approach produces estimates that have similar optimal properties to ML estimates, but with the added flexibility of incorporating prior knowledge and handling small sample sizes more effectively. The authors validated the empirical performance of GSEM through a series of simulation studies, demonstrating its effectiveness across different sample sizes and prior information scenarios. These studies confirm GSEM's robustness and its ability to produce accurate parameter estimates and standard error estimates, even in complex models

with non-linear effects and covariates. Nevertheless, as general challenges in Bayesian methods, GSEM relies on strong distributional assumptions on latent predicting variables and requires informative prior distributions in general. The choices of prior distributions need to be based on past literature or reliable empirical research; otherwise, biased parameter estimates may be produced.

Method of Moment Approaches

Developed by Wall and Amemiya (2003), the two-stage method of moments estimator (2SMM) is a regression method that corrects for “errors-in-variables” by initially estimating factor scores through a confirmatory factor analysis for both exogenous and endogenous latent variables. In its subsequent phase, the regression analysis employs these factor scores, adjusting for the sums of squares and cross-products of the predictors, as well as the interaction between predictors and outcomes, effectively performing a regression similar to ordinary least squares (OLS) but with adjustments for variables’ moments (Kelava & Brandt, 2023). Variables’ moments are essential statistical measures that describe the characteristics and shape of a data distribution. Each moment relates to a specific aspect of the distribution, starting with the first moment, known as the mean. The mean reflects the central tendency of the dataset, offering an average value derived by summing all data points and dividing by their count. This foundational measure sets the stage for understanding more complex properties of the distribution, as each subsequent moment reveals deeper insights into its structure and behavior. Moments beyond the first expand our understanding from the central location to variability, skewness, and kurtosis, each layer adding to the comprehensive statistical portrait of the dataset (Casella & Berger, 2002). 2SMM achieves asymptotically unbiased estimates for regressions involving latent variables with non-normal distributions, under the assumption that the residuals from the measurement model follow a normal distribution. Empirical validations, such as those conducted by Brandt et al. (2014), have demonstrated that 2SMM provides unbiased and efficient estimates for both parameters and their standard errors with normally distributed predictors, which have similarly

outstanding performance to methods like LMS and QML. Notably, 2SMM distinguishes itself by delivering unbiased parameter estimates in scenarios where variables deviate from normality and the structural model is correctly specified, although it may slightly under-represent standard errors with non-normal predictors. This evidences 2SMM's effectiveness in accommodating non-normal latent predictor variables, underscoring the method's foundational robustness as highlighted by Wall & Amemiya (2003).

Mooijart and Bentler (2010) introduced a method-of-moments (MM) approach that extends traditional analysis by incorporating higher-order moments of variables. This approach allows for a more flexible handling of measurement error variables by not strictly requiring them to follow a normal distribution. By focusing on minimizing the difference between observed and the model's expected higher-order moments, this method aims to accurately fit nonlinear models. While it operates under the assumption that latent predictor variables are normally distributed to ensure the consistency of estimates, it acknowledges and adjusts for the non-normality introduced by interaction terms. However, Brandt et al. (2014) noted potential limitations of MM, such that it may yield biased results if the latent predictor variables themselves are not normally distributed. Additionally, the stability and reliability of higher-order moments could be compromised in analyses with small sample sizes, affecting the method's robustness (Kelava & Brandt, 2023).

Study 1: Systematic Review of Simulation Studies Evaluating Latent Interaction Methods

Simulation studies, essentially computer-based experiments drawing on pseudo-random sampling from known probability distributions, stand as a cornerstone in statistical research (Morris et al., 2017; Hallgren, 2013; Boomsma, 2013). They are used particularly in assessing new methods and comparing different statistical models, which helps researchers understand the behavior of these models by comparing the model-implied results with the known and preset "population" parameters (Sainani, 2015). Within these simulation

studies, researchers define the traits of the target population, including its distribution and the interrelations among variables, which provides much more possibilities of addressing methodological research questions that cannot be answered using real data. Furthermore, the resilience of various methods, how unaffected they are by deviations from their theoretical assumptions, can be evaluated across differing population and sample attributes (Hinds et al., 2018). Simulation studies have been popularly used in examining properties of estimators and procedures in SEM. Boomsma (2013) conducted a comprehensive review of published articles from 1994 to 2012 on the journal *Structural Equation Modeling*, and found that about 30% articles purely conducted Monte Carlo simulation studies to evaluate various properties of advanced statistical models in their methods. Typical evaluations include Type I error rate, statistical power, and bias of parameter estimation.

However, some researchers may face challenges in conducting simulation studies due to lack of understanding, whereas others may not devote sufficient attention to the study's design or the clarity of their reporting (Siepe et al., 2023). A solid design of simulation studies empowers researchers to not only undertake but also critically evaluate such studies, while encouraging use of simulation studies with increased diligence and transparency in their reporting. Acknowledging simulation studies as practical experiments underlines the necessity of employing experimental design and analytical principles. As Siepe et al. (2023) mentioned, a good quality of evidence from a simulation study in psychology depends on the study design, conduction of experiment, data analysis, and appropriate reporting. Frequent issues in the setup, analysis, and dissemination of these studies may lead to a shallow processing and misleading interpretation of the data (Burton et al., 2006). Hence, an in-depth grasp of the aims, structure, execution, analytical approaches, and reporting within simulation studies is essential. Simmons et al. (2011) pointed out that in the realm of research of pure simulation study, researchers have significantly more researcher degrees of freedom than empirical research, in terms of greater flexibility in data generation and less cost of adjusting study design. For example, researchers can freely choose data-generating

models or mechanisms, evaluation objects and criteria, and content of reporting, while there could be potential issues in such flexibility (Hoaglin & Andrews, 1975; Burton et al., 2006; Feinberg & Rubright, 2016). Besides, some simulation studies may offer recommendations of practice for using certain methodologies or models, which may have great impact on future studies (Siepe, 2023).

Guidelines and recommendations for conducting Monte Carlo simulation studies in general psychology and psychometric-specific research have been developed and published for years (Boomsma, 2013; Carsey & Harden, 2014; Feinberg & Rubright, 2016). In addition, Morris et al. (2019) proposed a systematic scheme called ADEMP (Aims, Data-generating mechanisms, Estimands and other targets, Methods, Performance measure) as a design and reporting structure for future simulation studies. In this study, I plan to conduct a systematic review of simulation studies about statistical methods/models for estimating latent interaction effects, and use ADEMP as a template structure to examine whether they follow best practice in study design, data generation, parameter evaluation, and report and interpretation. Given that currently there is not a systematic review that has been conducted on latent interaction studies, I plan to describe the characteristic and quality of existing simulation studies, and use ADEMP structure to propose a guideline for future simulation studies on latent interaction.

Why Is It Important to Review Simulation Studies of Latent Interaction Models

As review in the earlier sections, latent interaction models are methodologically intricate, often requiring sophisticated statistical techniques and assumptions about data distribution, measurement error, and model specification. Simulation studies are uniquely positioned to test these models under controlled conditions, varying assumptions and parameters to systematically investigate their robustness, and validity. Given the purpose of comparing different models, ideally researchers want to control as many conditions as possible to maximally fairly compare models. However, it is not possible to evaluate all

latent interaction models in one study with all possible conditions. Without a comprehensive synthesis of these studies, the research community risks a fragmented understanding of these models' behaviors and limitations. A systematic review can consolidate findings across simulation studies, highlighting consistent patterns, discrepancies, and gaps in how these models perform across different conditions and settings (Abell et al., 2023). This aggregation is crucial for refining model assumptions, improving estimation techniques, and enhancing the overall reliability and accuracy of conclusions drawn from real-world data.

Current simulation studies of latent interaction models have some nuanced discrepancies in study design across studies. In the popular paper by Marsh et al. (2004), their results indicated that the UPI model using all possible PIs (all-pair UPI) demonstrated decreasing bias with satisfying standard error, and the UPI method has been widely used in empirical research. However, the all-pair UPI has only been examined in Study 1 in which this model was tested on parallel items. Specifically, all the factor loadings of observed indicators were set to .7 and the error variances were set to the same values in their population-generating model. Besides, they used raw bias averaged through simulation cycles as the quality indicator of point estimate of the latent interaction effect, and the empirical standard deviation and standard error were separately reported. With the development of latent interaction research, the all-pair UPI needs to be examined in more sophisticated conditions that may be encountered in substantive research (e.g., congeneric observed items with various factor loadings and measurement errors). Besides, more evaluation criteria of estimators may need to be used for more comprehensive examination of quality. In the supplemental material of my qualifying example paper "Two-Stage Path Analysis with Interaction: A Good Alternative to Current Methods of Modeling Latent", it was found that the all-pair UPI did not perform as well as reported on congeneric items with standard bias as the evaluation criteria. This is particularly important because slight variations in model implementation can lead to significantly different outcomes, affecting the validity and generalizability of the research conclusions. Although the paper of Marsh et

al. (2004) was published 20 years ago, it has been regarded as a standard paper of product indicator methods and the all-pair UPI method has been widely used in both methodological and empirical research. Thus, it is motivated to synthesize the evaluation of current latent interaction models with all the details of study design and manipulated conditions across existing simulation studies.

Besides, given that the focus and aim of simulation studies can largely differ, a systematic review of these simulation studies can synthesize across diverse conditions to establish how latent interaction models can be best specified and interpreted. For instance, simulation studies by Muthén and Asparouhov explored different estimation techniques like the integration method and Monte Carlo integration, revealing their effects on statistical power and error rates (Muthén & Asparouhov, 2003). However, Lai & Hsiao (2018) largely focused on comparing two latent interaction models across various reliability measures and reliability levels. Given the complexity of real data and substantive research (e.g., data type, data quality, study design, etc), it is important to choose the most appropriate method by considering the conditions where the method shows promising performance. By synthesizing outcomes from a range of simulation studies, systematic reviews can help in crafting nuanced recommendations that address diverse research contexts. For example, they can clarify how latent variable interactions should be handled in longitudinal studies versus cross-sectional studies, or in studies with hierarchical data structures. The work by Preacher et al. (2010) on multilevel mediation models highlights how simulation studies can provide insights into the interplay of complex model specifications in multilevel settings. Thus, a systematic review in latent interaction models is expected to provide a guideline for empirical researchers under complex situations.

Methods

Data Source and Extraction

The systematic review will be conducted on published peer-reviewed literature on available electronic databases of psychology studies: PsycINFO, PubMed, EMBASE, and Scopus. Google Scholar will be used as a supplemental database to comprehensively search all relevant journal articles. The search strategy and key terms will be as follows: (“latent” OR “structural equation modeling” AND (“interaction” OR “moderat*” OR “nonlinear”) AND (“simulat*” OR “monte carlo”). The strings “moderat*” and “simulat*” should capture all conjugated words including them, such as “moderation”, “moderated”, “simulation”, and “simulating”. The above searching strings will be searched in the title, abstract, and keyword fields of journals on each database. In order to ensure that the finalist of searched journal articles is comprehensive, the reference lists of included articles will be searched as the last step. Besides, given that the history of latent interaction methods is not long, articles published in all years should be relevant and included for synthesis.

After the initial data collection, I will use the inclusion criteria to exclude articles that are not in the scope of current study. The included journal articles for the next reviewing stage must: (1) have conducted at least one computer-based simulation study to evaluate one latent interaction model; (2) have reported results of simulation study; (3) have written the article in English. Assume that irrelevant articles have been excluded, the rest of articles will be manually examined, evaluated, and coded using Excel. A data collection table will be generated to include all the journal articles as a supplemental material for the future paper, while a preferred reporting items for systematic reviews and meta-analysis (PRISMA; Page et al., 2021) flow chart that shows all the process of data collection will be included in the main paper. As the main author, I will conducted the literature search and finalize the included articles, and I plan to recruit a research assistant (RA) as a secondary coder and reviewer. We will independently review the articles, code the information into

data, and construct the coding sheet. Then, all the discrepancies will be labeled by comparing two coding sheets and the discrepancy rate will be computed and reported. All the articles where discrepancies of data coding appear will then be reviewed again and discussed until mutual agreement is reached (i.e., discrepancy rate is 0%).

Data Analysis

The collected journal articles will be coded in an Excel sheet according to ADEMP structure. Below I will provide potential variables for each component of ADEMP.

Aims. Aims may include study purpose and objectives of simulation studies. Possible objectives may include: (1) Evaluate a newly proposed latent interaction method/model, (2) compare and contrast multiple latent interaction methods/models, (3) Evaluate methods of parameter estimation (e.g., ML), (4) examine latent interaction method/model for difficult variable processing (e.g., missing data), (5) others.

Data-generating mechanisms. Data-generating mechanism may include various models for generating population data. Regarding studies about latent interaction, most structural models are represented as multiple regression models, such as equation (2). Measurement models will depend on variable types. Possible models for continuous observed first-order indicators will include continuous confirmatory factor analysis (CFA) model, classical test theory (CTT) model, and others. For categorical variables, item response theory (IRT) model, Rasch model, and graded response model (GRM) may be used.

Other factors of design may include: (1) Whether the study uses real data/empirical example to inform/guide the design of the simulation study, (2) whether specific equations are provided for the population data model, (3) justification of study design, (4) sample size, (5) reliability of first-order indicators, (6) factor loadings of first-order indicators, (7) number of category and threshold value for categorical first-order indicators, (8) correlation between latent predicting variables, (9) missing data conditions (e.g., missing not at a random, missing at random, missing completely at random), (10) normality conditions of latent

variables, (11) normality conditions of observed indicators.

Estimands and other targets. First, possible estimates for latent interaction models will include: (1) point estimates of path coefficients (first-order effects and interaction effects), (2) standard errors of path coefficients, (3) model fit indexes, (4) mean and standard deviation of model fit indexes, (5) others.

Second, various model parameters will be specified or constrained depending on specific methods included in the study. Possible assumptions may include: (1) Distributions of first-order observed indicators and PIs/SIs, (2) distributions for measurement error of first-order indicators, PIs/SIs, and indicators for latent endogenous variable being predicted, (3) constraints on factor loadings (freely estimated or constrained), (4) distribution of disturbance term in the structural model.

Methods. Possible methods for latent interaction may include but are not limited to what have been introduced in the previous sections, depending on specific methods.

As for the structure of simulation experiments, possible parameters may include whether the study: (1) specified and reported starting seed, (2) collected and reported information on non-convergence, (3) reported number of replications for each replication condition, (4) provided justification of number of replications, (5) reported softwares used to performance the simulation study.

Performance measure. All the performance measures that appear in the collected simulation studies will be recorded and coded for descriptive analysis. Possible measures will include: (1) Bias and standardized bias, (2) (relative) standard error bias, (2) coverage rate, (3) root mean squared error (RMSE), (4) Type I error rate, (5) measure of statistical power, (6) others.