Dissertation Proposal

Gengrui (Jimmy) Zhang[1]

[1] University of Southhern California

Author Note

The authors made the following contributions. Gengrui (Jimmy) Zhang: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Gengrui (Jimmy) Zhang. E-mail: gengruiz@email.com

Abstract

My research intends to address current research gap in latent interaction studies, explore extensions of the Two-Stage Path Analysis (2S-PA) method in estimating latent interaction effects with categorical observed indicators, and investigate the two-stage corrected standard error of 2S-PA. In my previous work and the original paper of 2S-PA (Lai & Hsiao, 2022), 2S-PA showed promising properties of estimating path models and incorporate latnet interaction effects with continuous observed indicators. However, categorical items may bring obvious difficulty in terms of parameter estimation and model convergence. Besides, current 2S-PA model uses measurement parameters at stage 1 as known parameters, while these parameters may have uncertainty. Ignoring these uncertainty may result in biased estimation of structural parameters with underestimated standard errors. In this research proposal, I will systemtically review all existing methodological paper in which at least one simulation study has been conducted to evaluate latent interaction models (Study 1), use Monte Carlor simulation experiments to evaluate the performance of 2S-PA with interaction on categorical observed indicators (Study 2), and conduct another Monte Carlor simulation experiment to examine the performance of 2S-PA with corrected standard errors (Study 3).

Dissertation Proposal

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 (Dawson, 2014; Fairchild & McQuillin, 2010; Whisman & McClelland, 2005). 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). Considering the most parsimonious MMR model with one predictor (X), one moderator (Z), and one criterion variable (Y):

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \cdot Z + \epsilon, \tag{1}$$

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, 2002), 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 (Aguinis et al., 2005; Carroll et al., 2006; Fisicaro & Lautenschlager, 1992). 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, 1971). 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 predictor variables (e.g., X and Z) and their correlation. Alternatively speaking, the estimation of interaction effect will be downwardly biased with a larger degree if predictor 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 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.

Literature Review 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:

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

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:

$$x_1 = \lambda_{x_1} \xi_x + \delta_{x_1}, \qquad x_2 = \lambda_{x_2} \xi_x + \delta_{x_2}, \tag{3}$$

$$m_1 = \lambda_{m_1} \xi_m + \delta_{m_1}, \qquad m_2 = \lambda_{m_2} \xi_m + \delta_{m_2}, \tag{4}$$

where x_1, x_2, m_1, m_2 are observed indicators of ξ_x and ξ_m, two for each respectively. λs are the factor loadings and δs 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, δs, and ζ are multivariate normal with means of 0 and uncorrelated with each other. Alternatively demonstrating in mathematical symbols,

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

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 & Yang, 1996). Specifically,

$$\boldsymbol{\kappa} = \begin{pmatrix} 0 \\ 0 \\ \phi_{\xi_x \xi_m} \end{pmatrix} \qquad \boldsymbol{\Phi} = \begin{pmatrix} \phi_{\xi_x} & & \\ \phi_{\xi_m \xi_x} & \phi_{\xi_m} & \\ 0 & 0 & \phi_{\xi_x \xi_m} + \phi^2_{\xi_m \xi_x} \end{pmatrix}. \tag{6}$$

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:

$$x_1 m_1 = \lambda_{x_1} \lambda_{m_1} \xi_x \xi_m + \lambda_{x_1} \xi_x \epsilon_{m_1} + \lambda_{m_1} \xi_m \epsilon_{x_1} + \epsilon_{x_1} \epsilon_{m_1}, \tag{7}$$

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

$$\begin{array}{c} x_1m_1 \\ x_1m_2 \\ x_2m_1 \\ x_2m_2 \end{array} \begin{bmatrix} \theta_{x_1m_1} & a & b & \\ & \theta_{x_1m_2} & & b \\ & & \theta_{x_2m_1} & a \\ & & & \theta_{x_2m_2} \end{bmatrix}, \tag{8}$$

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

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

Although not many substantive research used the original method of Kenny and Judd (1984) due to complicated model constraints and difficulty in implementation, many later developed advanced latent interaction methods are mostly based on this seminal model. Kelava and Brandt (2019) in Chapter 23 of Handbook of Structural Equation Modeling (Hoyle, 2022) categorized commonly used latent interaction methods into four types: product indicator approaches, distribution analytic approaches, Bayesian parametric approaches, and method of moments approaches. The performance of each approach based on simulation studies and empirical research results has been summarized in Table 23.3 in the book. Below I will 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 $\mathbf{\Phi}$ 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 $\phi_{\xi_x\xi_m} + \phi^2_{\xi_m\xi_x}$ with the violation of normal distribution. Hence, theoretically CPI will 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, 2021; Cortina et al., 2021; Pieters et al., 2022).

Another disadvantage of UPI is that the number of PIs may become unmanageable while the first-order indicators are large. Although Wu 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, Hsiao et al. (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 Hsiao et al. (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 of using maximum likelihood (ML) estimation in PI methods is 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 directly involve PIs. Instead, they handle non-normality of latent dependent variables that may arise from the inclusion of latent interaction terms.

Traditional SEM approaches often assume that variables follow a normal distribution. However, A. Klein and Moosbrugger (2000) highlighted that latent interaction models typically induce non-normal distributions in their indicators due to the inclusion of product terms of latent variables. The LMS method specifically accounts for these non-normal distributions by modeling the joint indicator vector as a finite mixture of normal distributions, thus enhancing the accuracy and robustness of 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 observed indicators 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 (Aytürk et al., 2020, 2021; Cham et al., 2012b; Cortina et al., 2021).

Like LMS, quasi-maximum likelihood (QML) is another distribution analytic approach that extends beyond the traditional ML estimation, especially when the strict assumptions of ML are not fully met. It is particularly notable for its robustness against the non-normal distributions that typically arise from the product terms of latent variables (A. G. Klein & Muthén, 2007). 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. By approximating the density function of indicators using a variance function model approach, QML is able to produce accurate estimation of model parameters even when the underlying distribution deviates from normality. Unlike LMS which results in an exponential increase in computational complexity with the addition of interaction terms, QML manages such complexity without significant increases in computational demand. It shows enhanced efficiency in parameter estimation and maintains robustness in hypothesis testing across various scenarios where traditional normality assumptions do not hold.

Bayesian Parametric Approaches. According to Kelava and Brandt (2019) summary, Arminger and Muthén (1998) first introduced the application of nonlinear latent variable models that employs Bayesian methods, notably the Gibbs sampler and the Metropolis-Hastings algorithm, to derive the posterior distributions of parameters and latent variables, offering a robust alternative to traditional ML estimation methods that often fail to appropriately account for non-normality in such models. This framework allows for the use of conjugate priors informing predefined hypotheses about the parameter distributions and the estimation of posterior distributions through Markov chain Monte Carlo (MCMC) methods. One of the main advantages of this Bayesian method is its robustness in situations where the data do not follow a normal distribution. Besides, the Gibbs sampler and the Metropolis-Hastings algorithm are efficient in sampling from the posterior distributions of the model parameters for real-world datasets, which may be large or highly complex.

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. Compared to Arminger and Muthén's method (1998), which also uses a Bayesian framework but primarily focuses on simpler interaction models, GSEM extends usability to more intricate models, such as the inclusion of covariates might influence both the independent and dependent variables in a study, complex nonlinearities, and higher-order effects for curvilinear relationships among variables. 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.

Asparouhov and Muthén (2021) newly introduced another Bayesian approach that can be applied for single and multilevel models with latent interaction effects. Compared to traditional ML methods that often require high computational demands and limitations when numerical integration is required in higher dimensions, this method allows for more efficient handling of the high-dimensional integration typically needed for models with multiple interactions and hierarchical data layers. Moreover, the method extends the utility of Bayesian estimation to models with categorical dependent variables and incomplete data sets. The ability to incorporate categorical outcomes and manage missing data without compromising the robustness of the estimation process is particularly beneficial in applied social sciences, where such data issues are prevalent.

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, 2019). 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 mehtods 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 and Amemiya (2003).

Mooijaart 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, 2019).

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 (Boomsma, 2013; Hallgren, 2013; Morris et al., 2019). 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 (Burton et al., 2006; Feinberg & Rubright, 2016; Hoaglin & Andrews, 1975). Besides, some simulation studies may offer recommendations of practice for using certain methodologies or models, which may have great impact on future studies (Siepe et al., 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; 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. 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 et al., 2003). However, Hsiao et al. (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 tile, 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.

Study 2: Two-Stage Path Analysis with Interaction for Categorical Observed Indicators

In educational and social psychology research, the utilization of ordered categorical data is prevalent due to its efficacy in capturing the complex and subjective nature of human experiences (Azen & Walker, 2021). Among various tools employed to collect such data, questionnaires and surveys usually measured in Likert scales are frequently used due to their capacity to transform subjective opinions into quantifiable data and articulate responses in a sequential manner (Croasmun & Ostrom, 2011). This methodological approach allows for a rigorous analysis of human attitudes, bridging the gap between qualitative experiences and quantitative assessments. For instance, participants might respond by indicating their level of agreement with statements like "I feel confident in my abilities" using categories ranging from "Strongly Disagree" to "Strongly Agree" (Norman, 2010). Similarly, clinical assessments often rely on ordered categories to gauge the severity of symptoms, behaviors, or conditions. Diagnostic frameworks categorize mental health disorders with specifiers such as "mild," "moderate," or "severe" (American Psychiatric Association, 2013), while other assessments might rate levels of distress or impairment across various life domains. Sometimes although continuous data is available for researchers, rough categorizations of continuous variables should be used to indicate several levels, such as ranges of income and levels of grades, for discovering patterns. Ettner (1996) divided participants' income into five quantiles ranging from lowest to largest and revealed a gradient relationship where health status tends to improve with each rising income quantile. Park et al. (2002) similarly divided ages into groups such as young adults (20-39), middle-aged adults (40-59), and older adults (60+), and investigated changes in memory function across different age groups. The examples demonstrate that ordered categories offer the flexibility to impose a degree of structure and order on the data that continuous variables cannot achieve.

Interaction effects, wherein the effect of predictors depends on a third variable, are suggested to model within the latent variable framework due to its ability to account for measurement error. Most of latent interaction methods have been examined with non-normally distributed first-order indicators with continuous data. Cham et al. (2012a) compared the performance of estimating the interaction effect by constrained product indicator (CPI; Jöreskog & Yang, 1996), matched-pair UPI (Marsh et al., 2004), GAPI (Generalized Appended Product Indicator; Wall & Amemiya, 2001), and LMS (Latent Moderated Structural Equation; Klein & moosbrugger, 2000), with various degrees of non-normality of first-order variables' observed indicators. Their results showed that GAPI and UPI generally were able to produce unbiased estimates of the latent interaction effect with acceptable Type I error rates when sample size is larger than 500, while CPI and LMS tended to yield biased estimates. The promising performance of matched-pair UPI echoed the stemming article by Marsh et al. (2004) that matched-pair UPI had significant potential of generating unbiased estimates and demonstrating superior statistical properties in estimating latent interaction effects. For ordered categorical first-order indicators, Aytürk et al. (2020) first studied UPI with four matching strategies and LMS on ordered categorical indicators across sample sizes, indicators' distributions, and category conditions. They found that the UPI method with parceling strategy and LMS were able to produce reasonably unbiased estimates when the first-order indicators were ordered categorical variables with symmetric distributions. However, under non-normally distributed categorical indicators, both methods tended to generate largely overestimated interaction effects with underestimated standard errors, elevated Type I error rates, and unacceptable coverage rates. Aytürk et al. (2020) pointed out one explanation of inadequate performance was that both UPI and LMS assume that the first-order indicators of latent predictors have continuous and multivariate normal distributions. It implied that the two popular methods among current latent interaction models, to our knowledge, cannot show satisfactory performance for ordered categorical items with non-normal distribution.

Unfortunately, researchers often encounter non-normal ordered categorical data in psychology research, which poses unique challenges for latent interaction analysis. Ordered categorical data often arise from survey responses, behavioral classifications, or diagnostic categories, where the data can take on a limited number of categories but do not follow a normal distribution (Agresti & Kateri, 2011). This deviation from normality is especially common in assessments and questionnaires, where Likert scales and similar ordinal measures are used to capture attitudes, symptoms, or behaviors (Liddell & Kruschke, 2018). For example, an impactful study examining responses of the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) within a Japanese national survey highlighted a common pattern among depressive symptom items, such that a distinctive intersection point was found between "rarely" and "some of the time" categories (Tomitaka et al., 2018). It demonstrated that the item responses of the CES-D scale were mostly heavily right-skewed with most responses at the left extreme category. In the field of psychology research, numerous scales with ordered categorical items, including but not limited to the CES-D, are popularly utilized. It is therefore important to explore reliable methods for modeling latent interaction effects within highly non-normal categorical data for observed first-order indicators.

In the current study, I plan to replicate the results in Aytürk et al. (2020), and compare the performance of two-stage path analysis with interaction (2S-PA-Int) with that of UPI (both matched-pair and parceling strategies; Marsh et al., 2004) and LMS-cat (LMS with categorical variable; Muthén & Muthén, 2017) on estimating interaction effects with order categorical first-order indicators. The performance of 2S-PA-Int on continuous variables has been evaluated and it shows that 2S-PA-Int is able to generate unbiased estimates of the interaction effect with minimally biased standard error, low Type I error rate, acceptable coverage rate, and low RMSE values. Given its promising performance, I am motivated to continue to evaluate 2S-PA-Int with ordered-categorical variables in this study.

Below, I will introduce the three methods with technical details, and briefly discuss tentative procedures of conducting a simulation study.

A General Model of Latent Interaction

Kenny and Judd (1984) laid the groundwork for a structural model designed to estimate latent interaction effects, including two first-order latent predictors and their interaction term:

$$y = \alpha + \gamma_x \xi_x + \gamma_m \xi_m + \gamma_{xm} \xi_x \xi_m + \zeta. \tag{9}$$

In the model, α is the intercept, ξ_x and ξ_m are the first-order latent predictors with their product, $\xi_x \xi_m$, defining the interaction effect. Notably, ξ_x and ξ_m can correlate. The disturbance term, ζ, follows a normal distribution $\zeta \sim N(0, \psi)$, where ψ is its variance reflecting unexplained variance in the dependent variable. The coefficients γ_x, γ_m, and γ_{xm} quantify the effects of the predictors and their interaction on the dependent variable, which could be either observed or latent.

The measurement model for the first-order latent predictors, such as ξ_x, is framed within the confirmatory factor analysis (CFA) approach:

$$\mathbf{x} = \boldsymbol{\tau_x} + \boldsymbol{\lambda_x} \xi_x + \boldsymbol{\delta_x}. \tag{10}$$

For each indicator i ranging from 1 to p associated with ξ_x, $\mathbf{x}$ represents a vector of observed indicators of dimensions $p \times 1$. The vector $\boldsymbol{\tau_x}$ consists of constant intercepts, $\boldsymbol{\lambda_x}$ encompasses factor loadings, and $\boldsymbol{\delta_x}$ includes measurement errors, all of dimension $p \times 1$. Each error δ_{x_i} follows a normal distribution with a mean of zero and a variance θ_{x_i}. Under the assumption of local independence, the error variance-covariance matrix, $\boldsymbol{\Theta}_{\delta_\mathbf{x}}$, is diagonal, containing variances $\theta_{x_1}, \theta_{x_2}, ..., \theta_{x_p}$. The same measurement framework is applicable to ξ_m.

Unconstrained Product Indicator

The unconstrained product indicator (UPI) method operates by constructing product indicators from observed first-order indicators that relate to their corresponding latent constructs. Unlike the constrained approaches that impose nonlinear constraints on parameters related to the interaction term (Jöreskog & Yang, 1996), the UPI method does not require those constraints and thus simplify the model specification. UPI facilitates the estimation of interaction effects between latent variables directly within any commercial SEM software (e.g., the `R` package `lavaan`), enhancing accessibility for applied researchers.

The structural model of UPI is the same as equation (1). Suppose ξ_x is indicated by three items (i.e., $x_1 \sim x_3$) and ξ_m is indicated by six items (i.e., $m_1 \sim m_6$), the measurement models are presented as:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1} \\ \tau_{x_2} \\ \tau_{x_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1} \\ \lambda_{x_2} \\ \lambda_{x_3} \end{bmatrix} \begin{bmatrix} \xi_x \end{bmatrix} + \begin{bmatrix} \delta_{x_1} \\ \delta_{x_2} \\ \delta_{x_3} \end{bmatrix}, \tag{11}$$

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix} = \begin{bmatrix} \tau_{m_1} \\ \tau_{m_2} \\ \tau_{m_3} \\ \tau_{m_4} \\ \tau_{m_5} \\ \tau_{m_6} \end{bmatrix} + \begin{bmatrix} \lambda_{m_1} \\ \lambda_{m_2} \\ \lambda_{m_3} \\ \lambda_{m_4} \\ \lambda_{m_5} \\ \lambda_{m_6} \end{bmatrix} \begin{bmatrix} \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{m_1} \\ \delta_{m_2} \\ \delta_{m_3} \\ \delta_{m_4} \\ \delta_{m_5} \\ \delta_{m_6} \end{bmatrix}, \tag{12}$$

where τs are the intercepts, λs are the factor loadings, and δs are the error terms. In this study, we will explore the performance of two UPI models with the matching (matched-pair UPI) and parceling (parceling UPI) strategies.

Matched-pair UPI. The matching strategy involves selecting indicators from the first-order indicators of ξ_x and ξ_m, and pairing them up to form PIs. If the numbers of first-order indicators of ξ_x and ξ_m are equal, all the indicators will be used; Otherwise, Wu et al. (2013) recommended to select first-order indicators in the order of reliability from the latent variable with more indicators and discard the rest of indicators. For example, suppose that the sequences of $x_1 \sim x_3$ and $m_1 \sim m_6$ are sorted decreasingly by their reliability (i.e., x_1 has the highest reliability and x_3 has the lowest for ξ_x; m_1 has the highest reliability and m_6 has the lowest for ξ_m), the formed PIs for the interaction term will be x_1m_1, x_2m_2, x_3m_3. The items $m_4 \sim m_6$ should be discarded due to low reliability. The reliability of any single indicator (e.g., m_1) is:

$$\frac{\lambda_{m_1}^2 \sigma_{\xi_m}^2}{\lambda_{m_1}^2 \sigma_{\xi_m}^2 + \theta_{m_1}}, \tag{13}$$

where $\sigma_{\xi_m}^2$ is the variance of ξ_m and θ_{m_1} is the error variance of m_1 (Aytürk et al., 2020). Then, the model for matched-pair UPI will be represented as:

$$\begin{bmatrix} x_1m_1 \\ x_2m_2 \\ x_3m_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1m_1} \\ \tau_{x_2m_2} \\ \tau_{x_3m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1m_1} \\ \lambda_{x_2m_2} \\ \lambda_{x_3m_3} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1m_1} \\ \delta_{x_2m_2} \\ \delta_{x_3m_3} \end{bmatrix}. \tag{14}$$

Due to the nature of discrete data for first-order indicators, the indicators do not follow multivariate normal assumptions and their error variances are not homogeneous. However, UPI is robust to violations of multivariate normal distributions of the first-order latent variables, the disturbance term, and the measurement errors of first-order indicators, since these parameters are freely estimated in UPI. The mean of the latent interaction term is equal to the covariance between the first-order latent variables (i.e., $E[\xi_x\xi_m] = Corr[\xi_x, \xi_m]$ where $Corr[\cdot]$ is the correlation term; Jöreskog & Yang, 1996). Although UPI is theoretically less statistically powerful than CPI in the condition of non-normal first-order indicators, Marsh et al. (2004) showed that the performance of UPI on simulated non-normal data was acceptable with unbiased estimates of interaction effects and reasonable standard errors.

Parceling UPI. According to the original idea of Wu et al. (2013), another strategy of dealing with unequal numbers of first-order indicators is to take average of two or more indicators from the same source latent construct and form parcels. The number of parcels will be equal to the number of indicators from the latent variable with fewer indicators. Parcels can be formed based on considerations such as thematic consistency among indicators, item-to-total correlations, or factor loadings. In this study, we follow the factorial algorithm (Rogers & Schmitt, 2004) in the study design of Aytürk et al. (2020) by taking the average of the indicators with the highest and lowest reliability until all indicators have been used for parceling. In the example of $m_1 \sim m_6$, three parcels will be created to match three items of ξ_x and the measurement model will be represented as:

$$\begin{bmatrix} p_{m_1m_6} \\ p_{m_2m_5} \\ p_{m_3m_4} \end{bmatrix} = \begin{bmatrix} \tau_{p_{m_1m_6}} \\ \tau_{p_{m_2m_5}} \\ \tau_{p_{m_3m_4}} \end{bmatrix} + \begin{bmatrix} \lambda_{p_{m_1m_6}} \\ \lambda_{p_{m_2m_5}} \\ \lambda_{p_{m_3m_4}} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{p_{m_1m_6}} \\ \delta_{p_{m_2m_5}} \\ \delta_{p_{m_3m_4}} \end{bmatrix}, \tag{15}$$

where $p_{m_1m_6}$, $p_{m_2m_5}$, $p_{m_3m_4}$ are three formed parcels according to the factorial algorithm, in which m_1 has the highest reliability and m_6 has the lowest reliability, m_2 has the second highest reliability and m_5 has the second lowest reliability, and m_3 has the third highest reliability and m_4 has the third lowest reliability. Then, the parcels are matched to the indicators of another first-order latent variable according to the order of reliability. For instance, $p_{m_1m_6}$ will be matched to x_1, and so on. The model is represented as:

$$\begin{bmatrix} x_1p_{m_1m_6} \\ x_2p_{m_2m_5} \\ x_3p_{m_3m_4} \end{bmatrix} = \begin{bmatrix} \tau_{x_1p_{m_1m_6}} \\ \tau_{x_2p_{m_2m_5}} \\ \tau_{x_3p_{m_3m_4}} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1p_{m_1m_6}} \\ \lambda_{x_2p_{m_2m_5}} \\ \lambda_{x_3p_{m_3m_4}} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1p_{m_1m_6}} \\ \delta_{x_2p_{m_2m_5}} \\ \delta_{x_3p_{m_3m_4}} \end{bmatrix}, \tag{16}$$

For UPI with matching and parceling strategies, the first-order indicators will be simulated as continuous variables and standardized. Additionally, they will be mean-centered before forming PIs to enhance model convergence rates and reduce model complexity (Marsh et al., 2004). Alternatively speaking, the formed PIs are treated and estimated as continuous variables, which are liberal to multivariate normal distributions (Cham et al., 2012)

Latent Moderated Structural Equation

The LMS method, developed by Klein and Moosbrugger (2000), introduces a generalized interaction model that directly models the interaction effect using maximum likelihood estimation (ML) with the Expectation-Maximization (EM) algorithm. LMS does not form PIs for the interaction term and hence largely reduces model complexity for specification. The measurement models for the first-order latent variables are the same as those of UPI as demonstrated in equations (2) ~ (3).

LMS estimates the interaction effect by analyzing distributions of first-order indicators. LMS uses the Cholesky decomposition method to decompose the $p \times 1$ vector of latent variables, ξ, into a $p \times p$ lower triangular matrix $\mathbf{A}$ and a $p \times 1$ vector $\mathbf{z}$ of independently distributed variables with standard normal distributions, as p is the number of first-order latent predictors. Using the example of ξ_x and ξ_m

$$\xi = \mathbf{A}\mathbf{z}, \tag{17}$$

where

$$\xi = \begin{pmatrix} \xi_x \\ \xi_m \end{pmatrix}, \tag{18}$$

$$\mathbf{A} = \begin{pmatrix} a_{xx} & 0 \\ a_{xm} & a_{mm} \end{pmatrix}, \tag{19}$$

and

$$\mathbf{z} = \begin{pmatrix} z_x \\ z_m \end{pmatrix}. \tag{20}$$

The decomposition of ξ variables into independent z variables is integral to LMS as it addresses the non-normal distribution of the interaction effect. The structural equation of LMS, adapted on equation (1), is:

$$y = \alpha + \mathbf{\Gamma}\xi + \xi^T \mathbf{\Omega}\xi + \zeta, \tag{21}$$

where $\mathbf{\Gamma}$ is a $(1 \times p)$ vector of first-order path coefficients of ξ, and $\mathbf{\Omega}$ is a $(p \times p)$ upper-diagnol matrix of the interaction effects' coefficients. In the example of $\xi = (\xi_x,\ \xi_m)^T$ (T represents transpose), the expanded structural equation can be shown as:

$$y = \alpha + (\gamma_x\ \gamma_m)\begin{pmatrix}\xi_x \\ \xi_m\end{pmatrix} + (\xi_x\ \xi_m)\begin{pmatrix}0 & \gamma_{xm} \\ 0 & 0\end{pmatrix}\begin{pmatrix}\xi_x \\ \xi_m\end{pmatrix} + \zeta. \tag{22}$$

By substituting the ξ vector into the structural model of LMS, the structrual model can be further expanded to:

$$\begin{aligned}
\eta &= \gamma_0 + \mathbf{\Gamma}\mathbf{A}\mathbf{z} + \mathbf{z}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{\Omega}\mathbf{A}\mathbf{z} + \zeta \\
&= \gamma_0 + \mathbf{\Gamma}\mathbf{A}\begin{pmatrix}z_1 \\ 0\end{pmatrix} + \mathbf{\Gamma}\mathbf{A}\begin{pmatrix}0 \\ z_2\end{pmatrix} + \begin{pmatrix}z_1 \\ 0\end{pmatrix}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{\Omega}\mathbf{A}\begin{pmatrix}z_1 \\ 0\end{pmatrix} + \begin{pmatrix}z_1 \\ 0\end{pmatrix}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{\Omega}\mathbf{A}\begin{pmatrix}0 \\ z_2\end{pmatrix} + \zeta \\
&= \left(\gamma_0 + \mathbf{\Gamma}\mathbf{A}\begin{pmatrix}z_1 \\ 0\end{pmatrix} + \begin{pmatrix}z_1 \\ 0\end{pmatrix}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{\Omega}\mathbf{A}\begin{pmatrix}z_1 \\ 0\end{pmatrix}\right) + \left(\mathbf{\Gamma}\mathbf{A} + \begin{pmatrix}z_1 \\ 0\end{pmatrix}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{\Omega}\mathbf{A}\right)\begin{pmatrix}0 \\ z_2\end{pmatrix} + \zeta,
\end{aligned} \tag{23}$$

wherein the interaction effect demonstrates that the dependent measure η is linearly related to z_2 but non-linearly related to z_1, and hence η is not normally distributed. LMS then estimates the model by finding the maximum likelihood solution of the mixture distribution of the first-order indicators and η using the EM algorithm with numerical integration (Klein & Moosbrugger, 2000).

The assumptions of LMS in Klein and Moosbrugger's work include: (1) the first-order indicators are multivariate normal; (2) The errors of first-order indicators are normally and independently distributed with means of 0 (which is violated for order categorical items); (3) The disturbance term (i.e., ζ) in the structural equation is normally distributed with a mean of 0, and independent of latent variables and The errors of first-order indicators. Since LMS is estimated using maximum likelihood method that is based on normal distribution, it has been reported that LMS could generate biased estimates of the interaction effect when the first-order items did not follow normal distributions, for either continuous or ordered categorical indicators (Aytürk et al., 2020; Cham et al., 2012a; Cheung et al., 2021; Cheung & Lau, 2017; Marsh et al., 2004; Maslowsky et al., 2015). Aytürk et al. (2021) have shown that LMS with categorical variables in `Mplus` is able to generate unbiased estimates for both first-order and interaction effects across sample size, interaction effect size, missing data conditions, and numbers of item response category, but could produce highly biased measurement parameters such as factor loadings and item category thresholds.

Two-Stage Path Analysis with Interaction

The two-stage path analysis with interaction (2S-PA-Int) model is extended based on the 2S-PA model proposed by Lai and Hsiao (2022). By separating the estimation of measurement and structural models, 2S-PA simplifies model specification with reduced convergence issue while accounting for measurement errors in observed first-order indicators. At the first stage, factor scores are estimated from a measurement model within the confirmatory factor analysis framework, and the corresponding measurement errors are obtained as the error constraints. Then the factor scores serve as single indicators (SIs) to indicate their respective latent constructs, and the impact of latent exogenous variables on endogenous variables can be estimated as path coefficients. The standard errors of measurement for each factor scores should be used as error variance constraints in the model specification. Similar to UPI, 2S-PA is also a product indicator method in which the product of two latent variables' factor scores is the single indicator (SI) of the latent interaction effect.

Using the example of ξ_x and ξ_m , the factor scores will be obtained from equations (2) and (3). Then the factor scores are SIs of their corresponding latent variables:

$$\begin{bmatrix} \tilde{x}_j \\ \tilde{m}_j \\ \widetilde{xm}_j \end{bmatrix} = \begin{bmatrix} \tau_{\tilde{x}_j} \\ \tau_{\tilde{m}_j} \\ \tau_{\widetilde{xm}_j} \end{bmatrix} + \begin{bmatrix} \lambda_{\tilde{x}_j} & 0 & 0 \\ 0 & \lambda_{\tilde{m}_j} & 0 \\ 0 & 0 & \lambda_{\widetilde{xm}_j} \end{bmatrix} \begin{bmatrix} \xi_{x_j} \\ \xi_{m_j} \\ \xi_{x_j}\xi_{m_j} \end{bmatrix} + \begin{bmatrix} \delta_{\tilde{x}_j} \\ \delta_{\tilde{m}_j} \\ \delta_{\widetilde{xm}_j} \end{bmatrix}, \tag{24}$$

where for each observation j from $j = 1, 2, ..., n$, $\tilde{x}_j$, $\tilde{m}_j$ are factor scores that are SIs of ξ_x and ξ_m. The PI $\widetilde{xm}_j$ is the SI of the latent interaction term created by multiplying the SIs of ξ_x and ξ_m. The structural model of 2S-PA-Int is the same as equation (1).

Researchers can use multiple ways to calculate factor scores. According to the study 1 in Lai and Hsiao (2022), a one-factor confirmatory factor (CFA) model can be fitted to ordered categorical items using maximum likelihood estimation to obtain factor scores. Alternatively, researchers can use a unidimensional item response model with the expected-a-posterior (EAP) method for binary or multiple categorical items instead.

The 2S-PA-Int method has the potential to account for non-normal categorical variables due to its ability to model observation-specific standard errors of factor scores. Take the factor score $\tilde{x}_j$ as an example, it is assumed that its error term follows a normal distribution $\delta_{\tilde{x}_j} \sim N(0, \theta_{\tilde{x}_j})$ where $\theta_{\tilde{x}_j}$ is the estimated error variance. For continuous variables, 2S-PA is similar to the conventional structural equation modeling (SEM) approach where model parameters (like factor loadings, path coefficients, and measurement error variances) are assumed to be constant across the sample. The assumption of parameter constancy leads to a single likelihood function that is used to estimate the model parameters for all observations and simplifies the model estimation process. However, it may not always be appropriate especially in complex datasets where heterogeneity among individuals is expected. For ordered categorical data, the measurement error cannot be assumed to be constant across observations because the relationship between the observed categories and the underlying latent variables they are intended to measure is not linear, and the precision of measurement can vary across the spectrum of the latent trait. For example, in a 5-point Likert scale measuring agreement, the difference between "strongly agree" and "agree" might not represent the same magnitude of change in the underlying attitude as the difference between "neutral" and "agree." In the case of ordered categorical variables, 2S-PA uses definition variables to fixed specific values that may vary across individuals or groups since the variance of the measurement error can vary across observations' levels of underlying latent construct, and the likelihood function depends on the observation-specific standard error of measurement.

Since latent variables do not have meaningful units in nature, it is suggested to scale the variance of latent variables to unit when fitting the measurement model (i.e., $\sigma^2_{\xi_x} = 1$. In the example that factor scores using the EAP method from item response models, let $\hat{\sigma}_{\tilde{x}_j}$ be the estimates standard error of $\tilde{x}_j$ for the individual j. The estimated variance of the true score, also the estimated observation-specific reliability, can be computed as $\hat{\rho}_{\tilde{x}_j} = \sigma^2_{\xi_x} - \hat{\sigma}^2_{\tilde{x}_j} = 1 - \hat{\sigma}^2_{\tilde{x}_j}$, where $\hat{\rho}_{\tilde{x}_j}$ represents the reliability. This reliability is used as a constraint on the factor loading of $\tilde{x}_j$ (i.e., $\lambda_{\tilde{x}_j} = \hat{\rho}_{\tilde{x}_j} = 1 - \hat{\sigma}^2_{\tilde{x}_j}$), and the error variance constraint is set as $\theta_{\tilde{x}_j} = \hat{\sigma}^2_{\tilde{x}_j}\hat{\rho}_{\tilde{x}_j} = \hat{\sigma}^2_{\tilde{x}_j}(1 - \hat{\sigma}^2_{\tilde{x}_j})$ accordingly. The parameter constraints for the factor score of ξ_m can be obtained in the similar way. Assuming that the error variance of ξ_m has been calculated as $\theta_{\tilde{m}_j}$, the error variance constraint for the PI of the interaction term can be calculated as the equation (6) ~ (8) in Hsiao et al. (2018):

$$\theta_{\widetilde{xm}_j} = \hat{\rho}^2_{\tilde{x}_j}\theta_{\tilde{m}_j} + \hat{\rho}^2_{\tilde{m}_j}\theta_{\tilde{x}_j} + \theta_{\tilde{m}_j}\theta_{\tilde{x}_j}. \tag{25}$$

Method and Simulation Design

Based on the study design of Aytürk et al. (2020) and Hsiao et al. (2021), I intend to compare the performance of UPI with two strategies of forming PIs, LMS for categorical items, and 2S-PA-Int, on estimating latent interaction effects. The observed first-order indicators (e.g., x_{ij}) will be generated from a graded response model (Samejima, 1969) with differential factor loadings:

$$x^*_{ij} = \lambda_{x_i}\xi_{x_j} + \delta_{x_{ij}}, \tag{26}$$

where x^*_{ij} is the score of underlying latent continuous variable for each observed categorical item i. $\delta_{x_{ij}}$ is the individual-specific error term for each observed indicator i and follows standard logistic distribution, assuming the item factor model is estimated with a cumulative logit link (Wirth & Edwards, 2007). Given x^*_{ij}, the observed categorical item x_{ij} can be created with multiple categories with ordinal categories:

$$x_{ij} = \begin{cases} 0 & \text{if } x^*_{ij} < \beta_{x_{i1}} \\ k & \text{if } \beta_{x_{ik}} \leq x^*_{ij} < \beta_{x_{i(k+1)}} \\ K-1 & \text{if } \beta_{x_{i(K-1)}} \leq x^*_{ij} \end{cases}, \tag{27}$$

where β_{ik} is the threshold parameter between the kth and $(k+1)$th category for $k = 1, 2, ..., K$. The framework for observed categorical items of ξ_{m_j} is similar to ξ_{x_j}.

I will use `R 4.3.2` (R Core Team, 2022) to conduct the simulation studies in this project. To simulate the population dataset, the first-order latent variables ξ_{x_j} and ξ_{m_j} will be first simulated with standard normal distributions (i.e., $N[0,\ 1]$). The population structural model is the same as the equation (1), where the path coefficients are set to fixed values (i.e., $\gamma_x = 0.3$, $\gamma_m = 0.3$ and $\gamma_{xm} = 0.3$). The disturbance term ζ will be simulated with normal distribution and a mean of 0. When the correlation between ξ_{x_j} and ξ_{m_j} is 0, the variance of disturbance (i.e., ψ) will be 0.73. The interaction term will be computed as the product of ξ_{x_j} and ξ_{m_j}, and the exogenous variable y will be simulated with a distribution of $N(0,\ \psi)$.

Next, the observed ordered categorical indicators will be generated according to the equations (18) ~ (19). Three items will be simulated for ξ_{x_j} (i.e., $x_{1j} \sim x_{3j}$) and 12 items for ξ_{m_j} (i.e., $m_{1j} \sim m_{12j}$). Regarding factor loadings, I plan to adapt the design in Aytürk et al. (2020) in which the unstandardized factor loadings are decreasing in magnitude with equally-spaced intervals across indicators. Specifically, the factor loadings for $x_{1j} \sim x_{3j}$ are set to (.60, .70, .80), and those for $m_{1j} \sim m_{12j}$ are set to (.30, .35, .40, .45, .50, .55, .60, .65, .70, .75, .80, .85). The reliability measure for categorical indicators can be computed using

Green and Yang (2009) alternative reliability estimate, ω_{cat}, for unidimensional categorical items. Flora (2020) presented the way of computing ω_{cat} using the function `reliability` from the R package `semTools` (Jorgensen et al., 2022), and the output of estimate `omega` and `omega2` is based on Green and Yang (2009) method. Hence, the scale internal consistency can be obtained using `reliability` function.

To better approximate the real substantive study, $x_{1j} \sim x_{3j}$ will be simulated with 2 categories (i.e., $K = 2$), which means that they are binary indicators, and $m_{1j} \sim m_{12j}$ will be simulated with 5 categories (i.e., $K = 4$), which means that they are indicators with multiple categories. In the case of symmetric distribution, the thresholds for $x_{1j} \sim x_{3j}$ will be set to 0 so that the success probability of indicators is .5, calculated by the R function `plogis`. For $m_{1j} \sim m_{12j}$, ks are simulated for symmetric distribution and skewed distribution. the threshold values will be centered around 0. Specifically, $\beta_{\mathbf{m_1}}$ is a threshold vector of (-1.5, -.5, .5, 1.5) across indicators. For the skewed condition, the thresholds for $x_{1j} \sim x_{3j}$ will be set to -2.2 and the success probability of indicators is .9; the threshold vector for $m_{1j} \sim m_{12j}$ is set to (.05, .75, 1.55, 2.55) across indicators and the distributions of simulated observed indicators are apparently right-skewed.

The final dataset should be composed of 15 observed indicators of ξ_{x_j} and ξ_{m_j}, and one exogenous observed variable y_j predicted by ξ_x, ξ_m, and the interaction term $\xi_{x_j}\xi_{m_j}$. Note that for UPI and LMS, the subscript j can be removed because they assume the same likelihood function for each observation.

The 2S-PA-Int method will be implemented by user-defined function based on the R package `lavaan` (Rosseel, 2012) and `OpenMx` (Boker et al., 2023), and estimated using maximum likelihood (`ML`). The factor scores using the EAP method can be calculated using the function `fscores` with the argument `full.scores.SE = TRUE` from the `mirt` package (Chalmers, 2012). The UPI method will be implemented using the function `IndProd` in the R package `semTools`, and the `sem` function in `lavaan`. The UPI models will be estimated using maximum likelihood with robust standard error (`MLR`). Since there is not a reliable R package for LMS, the LMS model will be estimated in `Mplus` (Muthén & Muthén, 1998-2024) using the Gauss-Hermite quadrature integration algorithm with 16 integration points. The robust standard errors will be used as well.

To investigate Type I error rate of latent interaction effect, a condition of $\gamma_{xm} = 0$ will be added in the study design to compare with $\gamma_{xm} = 0.3$. The latent variables ξ_x and ξ_m are assumed to follow standard normal distributions (i.e., $N \sim [0, 1]$) and they are allowed to correlate with three levels of correlation (i.e., $Corr[\xi_x, \xi_m] = 0,\ 0.3,\ 0.6$). Three sample size conditions (i.e., 100, 250, and 500) will be included in the study design to examine the impact of sample size on the estimation of latent interaction effect across methods, according to the study design of Hsiao et al. (2021). Besides, past studies have examined the impact of reliability of observed indicators, I plan to manipulate the reliability of underlying continuous indicators (i.e., x^*_{ij}) with three levels: 0.7, 0.8, and 0.9. According to the equation (5) of composite reliability, and given that the factor loadings are fixed values, I can adjust the error variances of each observed continuous indicator before categorizing them. Suppose the items $x_{1j} \sim x_{3j}$ are manipulated to have a reliability of 0.7, the sum of error variances will be 1.89. To achieve congeneric items, the distribution of error variance proportions will be 44%, 33%, and 23%, and the specific error variances of $x_{1j} \sim x_{3j}$ will be 0.83, 0.62, and 0.44. For $m_{1j} \sim m_{12j}$, the distribution of error variance proportions can be 5%, 3%, 12%, 6%, 10%, 2%, 15%, 9%, 8%, 11%, 7%, 12%.

In summary, the tentative study design will be a $2 \times 2 \times 3 \times 3$ study with two effect sizes of the interaction effect, two distributions of indicators, three sample sizes, and three reliability levels between the latent predictor and the latent moderator. The Monte Carlo Simulation will be structured and conducted using the R package `SimDesign` (Chalmers & Adkins, 2020), and the tentative number of replication will be 2,000 for each condition. Across all the models, the standardized point estimate of the interaction effect (i.e., $\hat{\gamma}_{xm}$) with standard error (i.e., $\hat{SE}[\hat{\gamma}_{xm}]$) will be compared.

Evaluation Criteria

For each method, I will compute convergence rate, standardized bias, relative standard error (SE) bias, root mean squared error (RMSE), empirical Type I error, and empirical statistical power, and compare these indices to examine the performance of each method on estimating latent interaction effect.

Convergence Rate. For each replication, the program may or may not produce an error, such as non positive definite variance-variance matrix or negative variance estimates, depending on the random simulated sample. The convergence rate will be calculated as the proportion of replications that do not generate any error messages out of all replications. Sometimes extreme parameter values and standard errors will appear especially in small sample size (i.e., $N = 100$) even though no error messages are generated, and robust versions of bias, relative SE bias, and RMSE values will be used.

Standardized Bias. The standardized bias will be used to evaluate how far an estimate is from its true value in standard error units. It is defined using the raw bias and standard error of a point estimate:

$$B(\gamma_{xm}) = R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm}), \tag{28}$$

$$SB = \frac{B(\gamma_{xm})}{SE_{\gamma_{xm}}}, \tag{29}$$

where R will be the total number of replications for $r = 1, 2, \ldots, 2{,}000$. $\hat{\gamma}_{xm_r}$ is the estimated interaction effect in each replication r and γ_{xm} is the population parameter set at 0.3. $B(\hat{\gamma}_{xm})$ is the averaged deviation $\hat{\gamma}_{xm}$ from the population parameter, and $SE_{\hat{\gamma}_{xm}}$ is the empirical standard error of $\hat{\gamma}_{xm}$ across replications. An absolute value of $SB \leq 0.40$ will be considered acceptable for each replication condition (Collins et al., 2001).

Robust Relative Standard Error (SE) Bias. The robust relative SE bias will be computed as:

$$Robust\ Relative\ SE\ Bias = \frac{MDN(\widehat{SE_r}) - MAD}{MAD}, \tag{30}$$

where MDN will be the median of the estimated SE values and MAD will be the empirical median-absolute-deviation of SE values. An absolute value of robust relative SE bias within 10% range will be considered acceptable (HOOGLAND & BOOMSMA, 1998).

Root Mean Squared Erorr (RMSE). The RMSE is defined as the squared root of the sum of squared bias:

$$RMSE = \sqrt{R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm})^2}. \tag{31}$$

RMSE measures the average difference between calculated interaction estimates and their true value, which can account for both bias, the systematic deviation from the true value, and variability, the spread of estimates across replications. In a 2,000 replication simulation, lower RMSE indicates greater accuracy in estimating $\hat{\gamma}_{xm}$. RMSE provides the most informative comparison across methodologies when key factors, including sample size, model complexity, and disturbance level, are held constant in the simulation.

Empirical Type I Error and Statistical Power. The empirical Type I error rate will be computed as the proportion of replications in which the Wald test rejects the true null hypothesis $H_0 : \gamma_{xm} = 0$ at the significance level $\alpha = .05$ for the condition $\hat{\gamma}_{xm} = 0$. The empirical power will be computed similarly for the condition $\gamma_{xm} \neq 0$.

Study 3: Two-Stage Path Analysis with Corrected Standard Error

The two-stage path analysis (2S-PA) approach, as a review, separates the model specification and estimation into two steps (Lai & Hsiao, 2022). At the first stage, factor scores are obtained using any appropriate psychometric methods across observations, and the corresponding estimated standard errors of measurement are computed for each observation. Different types of factor scores can be used at the first stage, such as expected-a-posterior (EAP) scores, regression scores, and composite scores. At the second stage, the factor scores from the first stage are analyzed in a path model while incorporating the measurement error to correct for biases. Below I will review the statistical model of 2S-PA with two types of factor scores.

Assume that $\tilde{x}_j$ and $\tilde{m}_j$ are estimated factor scores obtained at stage 1 for the observation j from two sets of observed indicators $x_{1j} \sim x_{3j}$ and $m_{1j} \sim m_{3j}$. At stage 2, the structural is composed of two parts:

$$\begin{cases} \tilde{\boldsymbol{\xi}}_{\boldsymbol{j}} = \boldsymbol{\Lambda_j}^* \boldsymbol{\xi_j} + \boldsymbol{\varepsilon_j}^* \\ \boldsymbol{\xi_j} = \boldsymbol{\alpha} + \boldsymbol{\Gamma} \boldsymbol{\xi_j} + \boldsymbol{\zeta_j}, \end{cases} \tag{32}$$

where $\tilde{\boldsymbol{\xi}}_{\boldsymbol{j}}$ is a $q \times 1$ vector of factor scores (i.e., $\tilde{\boldsymbol{\xi}}_{\boldsymbol{j}} \;=\; [\tilde{x}_j,\; \tilde{m}_j]^T$), and $\boldsymbol{\xi_j}$ is the corresponding $q \times 1$ vector of latent constructs indicated by their factor scores for each observation. $\boldsymbol{\Lambda_j}^*$ and $\boldsymbol{\varepsilon_j}^*$ are factor loadings and error terms of the factor score indicators. Note that $\boldsymbol{\varepsilon_j}^*$ follows normal distribution, $\boldsymbol{\varepsilon_j}^* \sim N(0,\; \boldsymbol{\Theta_j}^*)$, where $\boldsymbol{\Theta_j}^*$ is a $q \times q$ covariance matrix of measurement error for the factor score indicators. $\boldsymbol{\zeta_j}$ is the disturbance term of the endogenous latent variable in the structural model with a distribution of $\boldsymbol{\zeta_j} \sim N(0,\; \boldsymbol{\psi})$ where ψ is the variance of disturbance. Under the conditions that each factor scores are computed by separate unidimensional factor models or by using Bartlett methods (Bartlett, 1937), $\boldsymbol{\Lambda_j}^*$ will be a diagonal matrix. For example,

$$\boldsymbol{\Lambda_j}^* = \begin{bmatrix} \lambda^*_{x_j} & 0 \\ 0 & \lambda^*_{m_j} \end{bmatrix}. \tag{33}$$

As mentioned before, factor scores have different types. The first type is obtained using the regression method (Thurstone, 1935). Taking the group of x indicators as an example to demonstrate the computation of factor score, the estimation of observation-specific factor scores based on observed indicators can be represented as $\tilde{\boldsymbol{\xi}}_{\boldsymbol{x_j}} = \boldsymbol{A}_{\boldsymbol{x_j}} \boldsymbol{x}_j$ where $\boldsymbol{A}_{\boldsymbol{x_j}}$ is the factor score weighting matrix using the regression method and $\boldsymbol{x_j} = [x_{1j},\ x_{2j},\ x_{3j}]^T$. From existing literature (e.g., Devlieger et al., 2016), the formula of the factor score weighting matrix is $\boldsymbol{A}_{\boldsymbol{x_j}} = Var(\xi_{x_j}) \boldsymbol{\lambda}_{\boldsymbol{x_j}}{}^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{x_j}}$, wherein $Var(\xi_{x_j})$ is the variance of latent variable ξ_{x_j}, $\boldsymbol{\lambda}_{\boldsymbol{x_j}}{}^T$ is the transposed vector of factor loadings implied by a CFA model for observed indicators such that $\boldsymbol{x_j} = \boldsymbol{\lambda}_{\boldsymbol{x_j}} \xi_{x_j} + \boldsymbol{\varepsilon}_{\boldsymbol{x_j}}$, and $\Sigma^{-1}_{\boldsymbol{x_j}}$ is an inverse matrix of variance-covariance of observed indicators. Thus, the estimated factor scores can be represented using the weighting matrix, such that $\tilde{\boldsymbol{\xi}}_{\boldsymbol{x_j}} = \boldsymbol{A}_{\boldsymbol{x_j}} \boldsymbol{\lambda}_{\boldsymbol{x_j}} \xi_{x_j} + \boldsymbol{A}_{\boldsymbol{x_j}} \boldsymbol{\varepsilon}_{\boldsymbol{x_j}}$. Further more, the reliability measure of the estimated factor scores can be computed as the proportion of true score variance over the total variance, such that $\rho_{\tilde{\boldsymbol{\xi}}_{\boldsymbol{x_j}}} = (\boldsymbol{A}_{\boldsymbol{x_j}} \boldsymbol{\lambda}_{\boldsymbol{x_j}})^2 Var(\xi_{x_j}) / [(\boldsymbol{A}_{\boldsymbol{x_j}} \boldsymbol{\lambda}_{\boldsymbol{x_j}})^2 Var(\xi_{x_j}) + \boldsymbol{A}_{\boldsymbol{x_j}} \boldsymbol{\Theta}_{\boldsymbol{x_j}} \boldsymbol{A}_{\boldsymbol{x_j}}{}^T]$ with $\boldsymbol{\Theta}_{\boldsymbol{x_j}}$ being the unique factor covariance matrix (Lai & Hsiao, 2023).

As for Bartlett factor scores (Bartlett, 1937), the factor loadings of estimated factor scores on the corresponding latent variables are constrained to 1 (e.g., $\lambda^*_{x_j} = 1$). Accordingly, the score matrix for Bartlett scores now changes to $\boldsymbol{A}_{\boldsymbol{x_j}} = (\boldsymbol{\lambda}_{\boldsymbol{x_j}}{}^T \boldsymbol{\Theta}_{\boldsymbol{x_j}}{}^{-1} \boldsymbol{\lambda}_{\boldsymbol{x_j}})^{-1} \boldsymbol{\lambda}_{\boldsymbol{x_j}}{}^T \boldsymbol{\Theta}_{\boldsymbol{x_j}}{}^{-1}$ and the error variance-covariance matrix of estimated factor scores becomes $(\boldsymbol{\lambda}_{\boldsymbol{x_j}}{}^T \boldsymbol{\Theta}_{\boldsymbol{x_j}}{}^{-1} \boldsymbol{\lambda}_{\boldsymbol{x_j}})^{-1}$.

The Issue of Using Reliability as Known

Currently, 2S-PA treats the standard error of measurement of factor scores from stage 1 as known and uses the standard error of measurement as error-constraints on the structural coefficients in the stage 2 estimation. However, the factor loading matrix and error variance-covariance matrix of the factor scores obtained in the stage 1 can have uncertainty and should be accounted for in the second stage.

The second part in the equation (1) is the structural model used to estimate the structural parameters that describe the relations between the latent constructs, where $\boldsymbol{\alpha}$ is the constant intercept, $\boldsymbol{\Gamma}$ is the structural coefficients, and $\boldsymbol{\zeta_j}$ is the disturbance term for the endogenous latent construct. However, currently the estimation of structural parameters in 2S-PA assumes that the measurement error (reliability) from stage 1 is known. Alternatively speaking, the error constraints on $\boldsymbol{\Lambda_j}^*$ and $\boldsymbol{\Theta_j}^*$ are fixed values, which does not account for uncertainty in the estimation of measurement model at stage 1. Meijer et al. (2021) points out that treating the estimates as the true values (i.e, population values) may result in underestimated standard errors of the structural parameter estimators. Rosseel and Loh (2022) also mentioned that the uncertainty from the first stage could not be ignored and should be reflected in the structural model.

To account for the uncertainty at stage 1, we proposed to obtain corrected standard errors using a method similar to the one discussed in Meijer et al. (2021), as can be incorporated in the estimation of parameters in the structural model at stage 2:

$$\hat{V}_{\gamma,\ c} = \hat{V}_{\gamma} + J_{\gamma}(\hat{\boldsymbol{\omega}})\hat{V}_{\hat{\boldsymbol{\omega}}}J_{\gamma}(\hat{\boldsymbol{\omega}})^{T}, \tag{34}$$

where $\hat{V}_{\gamma}$ is a variance-covariance matrix of structural parameters, in which $\boldsymbol{\gamma}$ is a vector of individual structural parameters in the matrix, in the structural model that ignores uncertainty in the stage 1. In a simple regression model (e.g., ξ_m is predicted by ξ_x), $\boldsymbol{\gamma}$ has only one parameter γ_{xm}. $\hat{V}_{\hat{\boldsymbol{\omega}}}$ is a variance-covariance matrix of the measurement parameters of the factor scores. Assuming factor scores of x indicators and m indicators have been computed as $\tilde{\xi}_{x_j}$ and $\tilde{\xi}_{m_j}$, the vector of measurement parameters will be $\hat{\boldsymbol{\omega}} = (\lambda^*_{x_j},\ \lambda^*_{m_j},\ \theta^*_{x_j},\ \theta^*_{m_j})$ for regression factor scores. As for Bartlett scores,

$\hat{\boldsymbol{\omega}} = (\theta^*_{x_j},\ \theta^*_{m_j})$ since the factor loadings have been constrained to 1.

$\hat{V}_{\hat{\boldsymbol{\omega}}}$ is the variance-covariance matrix of estimated measurement error from stage 1. $J_\gamma(\hat{\boldsymbol{\omega}})$ is the Jacobian matrix of $\boldsymbol{\gamma}$ with respect to $\hat{\boldsymbol{\omega}}$, which maps changes in the parameter $\hat{\boldsymbol{\omega}}$ to changes in $\boldsymbol{\gamma}$ and hence accounts for the uncertainty in $\hat{\boldsymbol{\omega}}$. Specifically, the Jacobian matrix can be represented as:

$$J_\gamma(\hat{\boldsymbol{\omega}}) = \begin{bmatrix} \frac{\partial\gamma}{\partial\hat{\omega}_1} & \ldots & \frac{\partial\gamma}{\partial\hat{\omega}_k} \end{bmatrix}, \tag{35}$$

where ∂ is the first-order partial derivative. Mathematically, it represents the differential of $\boldsymbol{\gamma}$ at every point of measurement parameters $\hat{\omega}$ for $\hat{\omega}_1,\ \hat{\omega}_2,\ \ldots\ , \hat{\omega}_k$ where $\boldsymbol{\gamma}$ is differentiable.

In the case of simple regression example where $\boldsymbol{\gamma} = \gamma_{xm}$, the formula of $J_\gamma(\hat{\boldsymbol{\omega}})$ for regression factor scores can be expanded as:

$$J_\gamma(\hat{\boldsymbol{\omega}}) = \begin{bmatrix} \frac{\partial\gamma}{\partial\lambda^*_{x_j}} & \frac{\partial\gamma}{\partial\lambda^*_{m_j}} & \frac{\partial\gamma}{\partial\theta^*_{x_j}} & \frac{\partial\gamma}{\partial\theta^*_{m_j}} \end{bmatrix}, \tag{36}$$

and

$$J_\gamma(\hat{\boldsymbol{\omega}}) = \begin{bmatrix} \frac{\partial\gamma}{\partial\theta^*_{x_j}} & \frac{\partial\gamma}{\partial\theta^*_{m_j}} \end{bmatrix} \tag{37}$$

for Bartlett scores. For more complicated structural models where $\boldsymbol{\gamma}$ has multiple elements of structural parameters, $J_\gamma(\hat{\boldsymbol{\omega}})$ will have multiple rows where each row representing one parameter.

Each element in the Jacobian matrix $J_\gamma(\hat{\boldsymbol{\omega}})$ represents the partial derivatives of each γ with respect to each $\hat{\omega}$ and reflects how changes in the measurement parameters of factor scores in $\hat{\omega}_1 \sim \hat{\omega}_k$ will affect the estimation of structural parameters. Thus, the term $J_\gamma(\hat{\omega})\hat{V}_\omega J_\gamma(\hat{\omega})^T$ in general captures the contribution of the uncertainty of measurement parameters at stage 1 to the uncertainty in the estimator of the structural model at stage 2. $\hat{V}_{\gamma,\ c}$ represents the estimated structural parameters adjusting for the uncertainty in $\hat{\boldsymbol{\omega}}$, namely corrected estimators.

Rosseel and Loh (2022) introduced another new two-step method of estimating SEM models, called "structural-after-measurement", in which the measurement model is estimated first and then the structural model. The two-step corrected standard errors are available in SAM models, and it was shown that SAM is advantageous in generating unbiased parameter estimates with reasonable standard errors with robustness against misspecifications and small sample sizes. Similar to 2S-PA, the mechanism of two-step estimation largely avoids convergence issue and reduces model complexity. Their SAM method has two variants, local SAM and global SAM. Given the promising performance of SAM, I plan to compare the performance of 2S-PA with corrected standard errors with SAM models. Below I will briefly review local and global SAM, while more technical details are available in Rosseel and Loh (2022).

General Idea of SAM Framework

The general idea of SAM framework is using estimated parameters from measurement models (step 1), and then use these parameters as fixed values or obtained information to estimate the structural parameters in step 2. To keep consistency of notation using in Rosseel and Loh (2022), $\boldsymbol{\omega_1}$ will be used to represent parameters from measurement models and $\boldsymbol{\omega_2}$ is for structural parameters.

$$\boldsymbol{\omega_1} = (\boldsymbol{\nu}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}), \tag{38}$$

where $\boldsymbol{\nu}$ is a vector of the constant intercepts, $\boldsymbol{\Lambda}$ is a vector of the factor loadings, and $\boldsymbol{\Theta}$ is a variance-covariance matrix of measurement error, for observed indicators.

$$\boldsymbol{\omega_2} = (\boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{\Psi}), \tag{39}$$

where $\boldsymbol{\alpha}$ is a vector of the constant intercepts, $\boldsymbol{B}$ is a vector of the regression/path coefficients, and $\boldsymbol{\Psi}$ is a variance-covariance matrix of disturbances, for latent variables.

As Rosseel and Loh (2022) mentioned, the goal of using two-step estimation is to use the estimators $\hat{\boldsymbol{\omega}}_{\mathbf{1}}$ as a mere mean to obtain reliable estimates for $\hat{\boldsymbol{\omega}}_{\mathbf{2}}$ with reasonable standard errors.

Local SAM. Local SAM involves computing the measurement parameters, such as factor loadings and measurement errors, and using them to derive and estimate the mean vector (i.e., $E(\xi)$) and covariance matrix (i.e., $Cov(\xi)$) of the latent variables. The derivation utilizes summary statistics of observed indicators and transforms them through a mapping matrix M, which effectively bridges the observed data with the latent construct space. This process benefits from reducing biases in parameter estimates due to model misspecification and helps to alleviate issues with model convergence, especially in smaller samples or more complex models.

Specifically, the observed indicators can be estimated using a CFA model, such that

$$\boldsymbol{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\varepsilon} \tag{40}$$

, and the relations between latent variables $\boldsymbol{\xi}$ and observed indicators $\boldsymbol{y}$ can be connected by the mapping matrix, where

$$\boldsymbol{\xi} = \boldsymbol{M}(\boldsymbol{y} - \boldsymbol{\nu} - \boldsymbol{\varepsilon}) \tag{41}$$

. The dimension of $\boldsymbol{M}$ is the number of latent variables by the number of observed indicators. The choices of $\boldsymbol{M}$ can be the maximum-likehlihood (ML) discrepancy function, weighted least squares (WLS) function, or unweighted least squared (ULS) function.

Details of rationales of choosing the mapping matrix are available in Rosseel and Loh (2022).

Hence, the estimation of the mean vector and the variance-covariance matrix of latent variables can be expresses as:

$$\widehat{\boldsymbol{E}(\boldsymbol{\xi})} = \widehat{\boldsymbol{M}}(\bar{\boldsymbol{y}} - \hat{\boldsymbol{\nu}}) \tag{42}$$

, and

$$\widehat{\boldsymbol{Var}(\boldsymbol{\xi})} = \widehat{\boldsymbol{M}}(\boldsymbol{S} - \hat{\boldsymbol{\Theta}})\widehat{\boldsymbol{M}}^T \tag{43}$$

, where $\boldsymbol{S}$ is the sample variance-covariance matrix.

Then the estimates $\widehat{\boldsymbol{E}(\boldsymbol{\xi})}$ and $\widehat{\boldsymbol{Var}(\boldsymbol{\xi})}$ estimated from step 1 will be used to estimate parameters in $\boldsymbol{\omega_2}$.

Global SAM. Global SAM starts by estimating the measurement model parameters as listed in the equation (7). Once these parameters are established, they are kept constant throughout the subsequent analysis. This is in contrast to Local SAM, where the estimation of structural parameters is based on derived statistics (i.e., $\widehat{\boldsymbol{E}(\boldsymbol{\xi})}$ and $\widehat{\boldsymbol{Var}(\boldsymbol{\xi})}$) from the measurement model. In Global SAM, after fixing the measurement model parameters, the entire structural model, including these fixed parameters, is estimated. This approach does not require the intermediate calculation of the latent variables' mean vector and covariance matrix as in Local SAM. Alternatively speaking, the calculations of $\widehat{\boldsymbol{E}(\boldsymbol{\xi})}$ and $\widehat{\boldsymbol{Var}(\boldsymbol{\xi})}$ are implicitly estimated (@ Rosseel & Loh, 2022). Instead, the parameters of the structural model in the equation (8) are directly estimated while treating the measurement part parameters as known constants. Global SAM is particularly advantageous in complex SEM frameworks characterized by intricate variable dependencies, offering a thorough analysis by utilizing the stability of the pre-estimated measurement model. Compared to local SAM, global SAM is more generalized as it can be fitted to any models that can be fitted with traditional SEM models.

Current Study

In this study, I plan to evaluate 2S-PA with corrected standard errors by accounting for measurement uncertainty from stage 1 estimation, and compare the performance of traditional structural equation modeling (SEM), local and global structural-after-measurement (SAM) models, and 2S-PA without correcting standard errors through the Monte Carlo simulation experiments.

Methods and Study Design

I will use `R 4.3.2` (R Core Team, 2022) to conduct the simulation studies in this project. Specifically, the SEM and 2S-PA models will be implemented through the R package `lavaan` (Rosseel, 2012) and user-defined functions, while SAM models will be implemented using `sam` function in `lavaan`. The simulation design will be structured and conducted using the R package `SimDesign` (Chalmers & Adkins, 2020).

Model 1: Simple Regression Model. The first model I intend to assess is a simple regression model with two latent variables, ξ_x and ξ_m, where ξ_m is predicted by ξ_x:

$$\xi_{m_j} = \gamma_0 + \gamma_1 \xi_{x_j} + \zeta_j, \tag{44}$$

where ξ_{x_j} is simulated with standard normal distribution, and ζ_j is simulated with a normal distribution following $N \sim (0, 1 - \gamma_1^2)$. The variance of ξ_{m_j} can be computed from pre-defined variance of ζ_j, as discussed later. The indicators are then generated by the confirmatory factor analysis (CFA) model:

$$\begin{gathered} x_{ij} = \tau_{x_{ij}} + \lambda_{x_{ij}} \xi_{x_j} + \delta_{x_{ij}}; \\ m_{ij} = \tau_{m_{ij}} + \lambda_{m_{ij}} \xi_{m_j} + \delta_{m_{ij}}, \end{gathered} \tag{45}$$

where each observation-specific observed indicator is represented by x_{ij} and m_{ij} that are simulated as continuous variables with normally distributed error. $\delta_{x_{ij}}$ and $\delta_{m_{ij}}$ are observation-specific error term for each observed indicator. $\delta_{x_{ij}}$, $\delta_{m_{ij}}$ and ζ_j are assumed to have multivariate normal distributions and be mutually independent. $\tau_{x_{ij}}$ and $\tau_{m_{ij}}$ are their corresponding constant intercepts and assumed to be 0. As there are two methods of computing factor scores, I plan to use both methods in this study, and hence the 2S-PA model will have two variants to be included in the simulation study, one for the regression scores (2S-PA-reg) and one for the Bartlett scores (2S-PA-Bar).

Reliability in factor scores is crucial because it impacts the accuracy of measuring latent constructs, which affects the validity and interpretability of SEM results (Brown, 2015). High reliability in factor scores ensures consistent and stable measurements, reducing the impact of measurement error and enhancing the statistical power of the analysis. As discussed earlier, reliability of estimated factor scores is a function of factor loadings and measurement errors in observed indicators. Therefore, I plan to manipulate the level of error variances of observed indicators (and hence reliability of observed indicators) to explore how reliability of observed indicators will affect the estimation of structural parameters across methods. The error variances will be varied across items to achieve three different levels of reliability of observed indicators: $\rho = 0.7, 0.8, 0.9$ where ρ is the reliability measure. For each reliability level, the error variances will be systematically varied to have different proportions. Suppose that three observed indicators will be generated for each latent variable (i.e., $x_{1j} \sim x_{3j}$ for ξ_{x_j}; $m_{1j} \sim m_{3j}$ for ξ_{m_j}) and the factor loadings of two sets of indicators are set to [.8, .7, .6] for $x_{1j} \sim x_{3j}$ and [.75, .65, .50] for $m_{1j} \sim m_{3j}$, the proportions can be manipulated as: 44% of the total error variance for the first indicator, 33% for the second, and 23% for the third, for x and m indicators. Any appropriate reliability measure can be used as an estimator of reliability, and I will use composite reliability in this study, where $\rho = \Sigma(\lambda)^2/(\Sigma[\lambda]^2 + \theta)$. Assuming that $\rho = .70$, the total error variance for $x_{1j} \sim x_{3j}$ will be 1.89 and for $m_{1j} \sim m_{3j}$ will be 1.55 using the formula of composite reliability. After adjusting the proportions of items, the the error variances of three indicators would be manipulated as $[0.83, 0.62, 0.44]$ for $x_{1j} \sim x_{3j}$, and

[0.68, 0.51, 0.36] for $m_{1j} \sim m_{3j}$. The sample sizes will be manipulated to have 100, 250, and 500 observations to represent small, medium, and large sample size.

Furthermore, sample size is a prevalent experiment condition in most SEM research (Fan et al., 1999; Wolf et al., 2013), since sample size is related to multiple important statistical properties of SEM, such as statistical power and convergence issue (Kline, 2016). Past research has proposed some rule-of-thumbs of suggested sample size, such as "sample size $\geq$ 100" for simple path model in general recommendation (Boomsma, 1982). However, Wolf et al. (2013) pointed out that there is not a "gold rule" for a single number of the best sample size, and the determination of sample size should take into accout the number of indicators, number of factors, and complexity or types of SEM models. Given that Lai and Hsiao (2022) adopted the design of "sample size per indicator", I plan to continue to use this design to simultaneously explore the effect of sample size and number of indicators on the performance of the 2S-PA models. The ratio of sample size per indicator will be set to $N/i = 6,\ 12,\ 100$, and the tentative numbers of indicators will be $i = 6,\ 12,\ 24$. Specifically, the number of indicators will be split to 3 and 3 for two latent variables in this model (i.e., $i = 3$ for ξ_{x_j}). Then, the range of sample sizes will be 36 $\sim$ 2,400.

The intercept γ_0 will be set to 0 and γ_1 will be set to 0 or .39 as null effect or medium effect, for the structural parameters.

In summary, the current study will have a $3 \times 3 \times 2$ design with three levels of the ratio of sample sizes per observed indicator, three reliability levels of observed indicators, and two levels of the effect size of the path coefficient. The Monte Carlo simulation will tentatively have 2,000 replications for each condition.

Model 2: Mediation Model. Given that median analysis is popular in psychological research (Agler & De Boeck, 2017; Rucker et al., 2011), I plan to evaluate the performance of 2S-PA with corrected standard errors in an expanded model from the equation (13) by including a mediator ξ_{z_j}, where:

$$\xi_{z_j} = \gamma'_0 + \gamma'_1\xi_{x_j} + \zeta'_j;$$
$$\xi_{m_j} = \gamma''_0 + \gamma''_1\xi_{x_j} + \gamma''_2\xi_{z_j} + \zeta''_j, \tag{46}$$

where ξ_{z_j} mediates the path from ξ_{x_j} to ξ_{m_j}. γ'_0, γ'_1, ζ'_j are the structural parameters of the path from ξ_{x_j} to ξ_{z_j}, where as γ''_0, γ''_1, γ''_2, ζ''_j are those from ξ_{x_j} and ξ_{m_j} to ξ_{m_j}. The indirect effect is defined as the product of the path coefficients $\gamma'_1 \times \gamma''_2$. The error variances will be manipulated using the same way for ξ_{x_j} and ξ_{m_j}'s indicators.

Based on the study design in Lai and Hsiao (2022), γ'_0 and γ''_0 are set 0. γ''_1 is fixed to .15 as a small direct effect of ξ_{x_j} on ξ_{m_j}, while each of γ'_1 and γ''_2 is set to either 0 or .39 as null effect and medium effect. The indirect effect will be either 0 or .1521. Hence the possible configurations of path coefficients (i.e., γ'_1, γ''_2, γ''_1, $\gamma'_1 \times \gamma''_2$) will be [0, 0, .15, 0], [0, .39, .15, 0], [.39, 0, .15, 0], and [.39, .39, .15, .1521].

The other designs will be similar to the study of model 1. In summary, the study of mediation model will have a $3 \times 3 \times 4$ design with three levels of the ratio of sample sizes per observed indicator, three reliability levels of observed indicators, and four levels of the effect sizes of the path coefficients.

Evaluation Criteria

Across all the models, the standardized point estimate of the path coefficients with standard error will be compared. For study 1, the point estimate of $\hat{\gamma}_1$ with its standard error $\hat{SE}(\hat{\gamma}_1)$ will be obtained for model evaluation and comparisons. As for study 2, the point estimates of all the four path coefficients with standard errors will be obtained and compared.

The main evaluation criteria will include convergence rate, empirical Type I error, and empirical statistical power, which are used to examine the performance of each method on estimating path coefficients. The specific details of evaluation criteria have been described in Study 2.

References

Agler, R., & De Boeck, P. (2017). On the Interpretation and Use of Mediation: Multiple Perspectives on Mediation Analysis. *Front. Psychol.*, *8*. https://doi.org/10.3389/fpsyg.2017.01984

Agresti, A., & Kateri, M. (2011). Categorical Data Analysis. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 206–208). Springer. https://doi.org/10.1007/978-3-642-04898-2_161

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect Size and Power in Assessing Moderating Effects of Categorical Variables Using Multiple Regression: A 30-Year Review. *Journal of Applied Psychology*, *90*(1), 94–107. https://doi.org/10.1037/0021-9010.90.1.94

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions* (pp. xi, 212). Sage Publications, Inc.

Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, *63*(3), 271–300. https://doi.org/10.1007/BF02294856

Asparouhov, T., & Muthén, B. (2021). Bayesian estimation of single and multilevel models with latent variable interactions. *Structural Equation Modeling*, *28*(2), 314–328. https://doi.org/10.1080/10705511.2020.1761808

Aytürk, E., Cham, H., Jennings, P. A., & Brown, J. L. (2020). Latent Variable Interactions With Ordered-Categorical Indicators: Comparisons of Unconstrained Product Indicator and Latent Moderated Structural Equations Approaches. *Educ Psychol Meas*, *80*(2), 262–292. https://doi.org/10.1177/0013164419865017

Aytürk, E., Cham, H., Jennings, P. A., & Brown, J. L. (2021). Exploring the Performance of Latent Moderated Structural Equations Approach for Ordered-Categorical Items. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(3), 410–422. https://doi.org/10.1080/10705511.2020.1810047

Azen, R., & Walker, C. M. (2021). *Categorical Data Analysis for the Behavioral and Social Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780429330308

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section*, *28*(1), 97–104. https://doi.org/10.1111/j.2044-8295.1937.tb00863.x

Boker, S. M., Neale, M. C., Maes, H. H., Spiegel, M., Brick, T. R., Estabrook, R., Bates, T. C., Gore, R. J., Hunter, M. D., Pritikin, J. N., Zahery, M., & Kirkpatrick, R. M. (2023). *OpenMx: Extended structural equation modelling.* https://CRAN.R-project.org/package=OpenMx

Bollen, K. A. (1989). *Structural equations with latent variables* (pp. xiv, 514). John Wiley & Sons. https://doi.org/10.1002/9781118619179

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. *Part I*, *1*.

Boomsma, A. (2013). Reporting Monte Carlo Studies in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(3), 518–540. https://doi.org/10.1080/10705511.2013.797839

Brandt, H., Kelava, A., & Klein, A. (2014). A Simulation Study Comparing Recent Approaches for the Estimation of Nonlinear Effects in SEM Under the Condition of Nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 181–195. https://doi.org/10.1080/10705511.2014.882660

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, 2nd ed* (pp. xvii, 462). The Guilford Press.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Stat Med*, *25*(24), 4279–4292. https://doi.org/10.1002/sim.2673

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, *93*(3), 549–562.

https://doi.org/10.1037/0033-2909.93.3.549

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective, second edition* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781420010138

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed). Thomson Learning.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, R. P., & Adkins, M. C. (2020). Writing effective and reliable Monte Carlo simulations with the SimDesign package. *The Quantitative Methods for Psychology*, *16*(4), 248–280. https://doi.org/10.20982/tqmp.16.4.p248

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012a). Estimating latent variable interactions with non-normal observed data: A comparison of four approaches. *Multivariate Behav Res*, *47*(6), 840–876. https://doi.org/10.1080/00273171.2012.732901

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012b). Estimating Latent Variable Interactions With Non-Normal Observed Data: A Comparison of Four Approaches. *Multivariate Behav Res*, *47*(6), 840–876. https://doi.org/10.1080/00273171.2012.732901

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2021). Testing Moderation in Business and Psychological Studies with Latent Moderated Structural Equations. *J Bus Psychol*, *36*(6), 1009–1033. https://doi.org/10.1007/s10869-020-09717-0

Cheung, G. W., & Lau, R. S. (2017). Accuracy of Parameter Estimates and Confidence Intervals in Moderated Mediation Models: A Comparison of Regression and Latent Moderated Structural Equations. *Organizational Research Methods*, *20*(4), 746–769. https://doi.org/10.1177/1094428115595869

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*, *6*(4), 330–351.

Cortina, J. M., Markell-Goldstein, H. M., Green, J. P., & Chang, Y. (2021). How are we testing interactions in latent variable models? Surging forward or fighting shy? *Organizational Research Methods*, *24*(1), 26–54. https://doi.org/10.1177/1094428119872531

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-Type Scales in the Social Sciences. *Journal of Adult Education*, *40*(1), 19–22.

Dawson, J. F. (2014). Moderation in Management Research: What, Why, When, and How. *J Bus Psychol*, *29*(1), 1–19. https://doi.org/10.1007/s10869-013-9308-7

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38. https://www.jstor.org/stable/2984875

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression. *Educ Psychol Meas*, *76*(5), 741–770. https://doi.org/10.1177/0013164415607618

Ettner, S. L. (1996). New evidence on the relationship between income and health. *J Health Econ*, *15*(1), 67–85. https://doi.org/10.1016/0167-6296(95)00032-1

Fairchild, A. J., & McQuillin, S. D. (2010). Evaluating mediation and moderation effects in school psychology: A presentation of methods and review of current practice. *Journal of School Psychology*, *48*(1), 53–84. https://doi.org/10.1016/j.jsp.2009.09.001

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 56–83. https://doi.org/10.1080/10705519909540119

Feinberg, R. A., & Rubright, J. D. (2016). Conducting Simulation Studies in Psychometrics. *Educational Measurement: Issues and Practice*, *35*(2), 36–49. https://doi.org/10.1111/emip.12111

Fisicaro, S. A., & Lautenschlager, G. J. (1992). Power and Reliability: The Case of Homogeneous True Score Regression Across Treatments. *Educational and Psychological Measurement*, *52*(3), 505–511. https://doi.org/10.1177/0013164492052003001

Flora, D. B. (2020). Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Green, S. B., & Yang, Y. (2009). Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. *Psychometrika*, *74*(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

Hallgren, K. A. (2013). Conducting Simulation Studies in the R Programming Environment. *Tutor Quant Methods Psychol*, *9*(2), 43–60.

Hinds, A. M., Sajobi, T. T., Sebille, V., Sawatzky, R., & Lix, L. M. (2018). A systematic review of the quality of reporting of simulation studies about methods for the analysis of complex longitudinal patient-reported outcomes data. *Qual Life Res*, *27*(10), 2507–2516. https://doi.org/10.1007/s11136-018-1861-0

Hoaglin, D. C., & Andrews, D. F. (1975). The Reporting of Computation-Based Results in Statistics. *The American Statistician*, *29*(3), 122–126. https://doi.org/10.1080/00031305.1975.10477393

HOOGLAND, J. J., & BOOMSMA, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*(3), 329–367. https://doi.org/10.1177/0049124198026003003

Hoyle, R. H. (2022). *Handbook of Structural Equation Modeling*. Guilford Publications.

Hsiao, Y.-Y., Kwok, O.-M., & Lai, M. H. C. (2018). Evaluation of two methods for modeling measurement errors when testing interaction effects with observed composite scores. *Educ Psychol Meas*, *78*(2), 181–202. https://doi.org/10.1177/0013164416679877

Hsiao, Y.-Y., Kwok, O.-M., & Lai, M. H. C. (2021). Modeling measurement errors of the exogenous composites from congeneric measures in interaction models. *Struct Equ Modeling*, *28*(2), 250–260. https://doi.org/10.1080/10705511.2020.1782206

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. https://doi.org/10.1007/BF02291393

Jöreskog, K. G., & Yang, F. (1996). *Nonlinear structural equation models: The Kenny-Judd model with Interaction effects.*

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *`semTools`: Useful tools for structural equation modeling.* https://CRAN.R-project.org/package=semTools

Kelava, A., & Brandt, H. (2019). A Nonlinear Dynamic Latent Class Structural Equation Model. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 509–528. https://doi.org/10.1080/10705511.2018.1555692

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*(1), 201–210. https://doi.org/10.1037/0033-2909.96.1.201

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*(4), 647–673. https://doi.org/10.1080/00273170701710205

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*(4), 457–474. https://doi.org/10.1007/BF02296338

Kline, R. B. (2016). *Principles and practice of structural equation modeling, 4th ed* (pp. xvii, 534). Guilford Press.

Lai, M. H. C., & Hsiao, Y.-Y. (2022). Two-stage path analysis with definition variables: An alternative framework to account for measurement error. *Psychol Methods*, *27*(4), 568–588. https://doi.org/10.1037/met0000410

Lee, S.-Y., Song, X.-Y., & Tang, N.-S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling*, *14*(3), 404–434. https://doi.org/10.1080/10705510701301511

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychol Methods*, *9*(3), 275–300. https://doi.org/10.1037/1082-989X.9.3.275

Maslowsky, J., Jager, J., & Hemken, D. (2015). Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations method. *Int J Behav Dev*, *39*(1), 87–96. https://doi.org/10.1177/0165025414552301

Meijer, E., Oczkowski, E., & Wansbeek, T. (2021). How measurement error affects inference in linear regression. *Empir Econ*, *60*(1), 131–155. https://doi.org/10.1007/s00181-020-01942-z

Mooijaart, A., & Bentler, P. M. (2010). An alternative approach for nonlinear latent variable models. *Structural Equation Modeling*, *17*(3), 357–373. https://doi.org/10.1080/10705511.2010.488997

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Muthen, L. K., Muthen, B., & Angeles)., M. &. M. (Los. (2017). *Mplus version 8 user's guide*. Muthen & Muthen.

Muthén, B., Asparouhov, T., & Muthén, M. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus. *Mplus Web Notes*, *6*.

Newsom, C. R., Archer, R. P., Trumbetta, S., & Gottesman, I. I. (2003). Changes in adolescent response patterns on the MMPI/MMPI-A across four decades. *Journal of*

Personality Assessment, *81*(1), 74–84. https://doi.org/10.1207/S15327752JPA8101_07

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ Theory Pract*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research, and Evaluation*, *8*(1). https://doi.org/10.7275/r222-hv23

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Pieters, C., Pieters, R., & Lemmens, A. (2022). Six Methods for Latent Moderation Analysis in Marketing Research: A Comparison and Guidelines. *Journal of Marketing Research*, *59*(5), 941–962. https://doi.org/10.1177/00222437221077266

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychol Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Rogers, W. M., & Schmitt, N. (2004). Parameter Recovery and Model Fit Using Multidimensional Composites: A Comparison of Four Empirical Parceling Algorithms. *Multivariate Behavioral Research*, *39*(3), 379–412. https://doi.org/10.1207/S15327906MBR3903_1

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., & Loh, W. W. (2022). A structural after measurement approach to structural equation modeling. *Psychol Methods*. https://doi.org/10.1037/met0000503

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation Analysis in Social Psychology: Current Practices and New Recommendations. *Social and Personality Psychology Compass*, *5*(6), 359–371. https://doi.org/10.1111/j.1751-9004.2011.00355.x

Sainani, K. L. (2015). What is Computer Simulation? *PM R*, *7*(12), 1290–1293. https://doi.org/10.1016/j.pmrj.2015.10.010

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*(4, Pt. 2), 100–100.

Schoemann, A. M., & Jorgensen, T. D. (2021). Testing and interpreting latent variable interactions using the semTools package. *Psych*, *3*(3), 322–335. https://doi.org/10.3390/psych3030024

Siepe, B. S., Bartoš, F., Pawel, S., & Boulesteix, A.-L. (2023). *Simulation Studies for Methodological Research in Psychology*.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychol Sci*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits* (pp. xi, 274). University of Chicago Press. https://doi.org/10.1037/10018-000

Tomitaka, S., Kawasaki, Y., Ide, K., Akutagawa, M., Ono, Y., & Furukawa, T. A. (2018). Distribution of item responses and total item scores for the Center for Epidemiologic Studies Depression Scale (CES-D): Data from the Irish Longitudinal Study on Ageing (TILDA). *PLOS ONE*, *13*(8), e0202607. https://doi.org/10.1371/journal.pone.0202607

Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, *26*(1), 1–30. https://doi.org/10.3102/10769986026001001

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*(1), 47–63. https://doi.org/10.1348/000711003321645331

Whisman, M. A., & McClelland, G. H. (2005). Designing, Testing, and Interpreting Interactions and Moderator Effects in Family Research. *Journal of Family Psychology*, *19*(1), 111–120. https://doi.org/10.1037/0893-3200.19.1.111

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237

Wu, Y., Wen, Z., Marsh, H., & Hau, K.-T. (2013). A comparison of strategies for forming product indicators for unequal numbers of items in structural equation models of latent interactions. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 551–567. https://doi.org/10.1080/10705511.2013.824772