Two-Stage Path Analysis with Interaction: A Good Alternative to Current Methods of Modeling Latent Interaction

Jimmy Zhang[1]

[1] University of Southhern California

Author Note

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

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

Abstract

Interaction effects between latent variables are becoming increasingly popular to support theory in depth and complex data structure in psychology research. Comprared to widely used methods of modeling latent interactions, matched-pair Unconstrained Product Indicator (UPI) and Reliability-Adjusted Product Indicator (RAPI), an extended model based on the two-stage path analysis (2S-PA) framework, namely 2S-PA-Int, was evaluated and demonstrated good performance. Based on a simulation study with 2000 replications, 2S-PA-Int showed consistently less standardized bias, acceptable relative SE bias and coverage rates, and lower RMSE values than matched-UPI and RAPI, particularly under the conditions of sample size and high amount of measurement error. General advantages and future research directions of 2S-PA-Int are discussed.

Keywords: Latent interaction, UPI, RAPI, 2S-PA

Two-Stage Path Analysis with Interaction: A Good Alternative to Current Methods of Modeling Latent Interaction

Social science research increasingly focuses on complex effects (e.g., nonlinear effects, moderation) rather than simple bivariate relationships as the real world is rarely simple and straightforward (Carte and Russell (2003); MacKinnon & Luecken, 2008; Cunningham & Ahn, 2019). Research demonstrates that exercising may help people lose weight, but people may be further interested in how, when, for whom, and under what conditions that exercising can do for losing weight. Moderation (or interaction) research can answer such questions by investigating how a third variable (or a group of additional variables) modifies relations among variables of interest.

One widespread way to model moderation is through regression model, specifically incorporating an interaction term XZ:

$$Y = b_0 + b_1 X + b_2 Z + b_3 XZ + \epsilon, \tag{1}$$

where b_0 is the intercept, b_1 and b_2 are the regression coefficients for X and Z, b_3 is the coefficient for the interaction term XZ, and ϵ is the residual term. To maintain consistency with the naming conventions used by Marsh et al. (2004), we refer to main effects (i.e., non-interaction effects) as "first-order effects". Hence X and Z are first-order variables and b_1 and b_2 are first-order effects in this case. Classical regression assumes variables are measured without error, which may lead to biased parameter estimates (especially for the interaction) when measurement errors are not uncommonly present in empirical research (Bollen, 1989; Cohen et al., 2003; Caroll et al., 2006). To address this problem, researchers use latent variables that are inferred and measured by a set of observed indicators in the structural equation modeling (SEM) framework, which can control and accommodate measurement errors in these observed indicators (Bollen, 2002). For example, depression is widely tested and measured by the Center for Epidemiologic Studies Depression (CES-D) scale consisting of 20 items (Radloff, 1977). Moderation models based on SEM provide reliably true relationships among latent constructs (Mueller, 1997; Steinmetz et al., 2011; Cham et al., 2012; Maslowsky et al., 2015).

The Two-Stage Path Analysis (2S-PA) is a method of modeling latent variables based SEM and it is demonstrate to show the capability of producing parameter estimates with less standard error bias, higher convergence rates, and better handling of Type I error in small samples (Lai & Hsiao, 2021). Given its good statistical property, we extended the 2S-PA method to incorporate latent interaction estimation in this study, and named it 2S-PA-Int. We reviewed two widely used latent interaction models, Unconstrained Product Indicator (UPI; Marsh et al., 2004) and Reliability-Adjusted Product Indicator (RAPI; Hsiao et al., 2018), and conducted a Monte Carlo simulation study to compare their performance with 2S-PA-Int. To proceed, we first introduced a classical model of latent interaction and then presented UPI, RAPI, and 2S-PA-Int with technical details.

A Classical Model of Latent Interaction

Kenny and Judd (1984) first proposed a classical structural model that provided a seminal idea of estimating latent interaction effecs, focusing on a basic scenario involving two latent predictors and one latent interaction term:

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

where α is the constant intercept, ξ_x and ξ_m denote the first-order latent predictors, and the product $\xi_x \xi_m$ constitutes the interaction term. Note that ξ_x and ξ_m are allowed to correlate with each other. As for other parameters, ζ is the model's disturbance term assumed to follow a normal distribution $\zeta \sim N(0, \psi)$ where ψ is a scalar representing the variance of ζ that captures unobserved factors influencing the dependent variable. The coefficients γ_x and γ_m indicate first-order effects of the latent predictors, whereas γ_{xm} quantifies the latent interaction effect. The dependent variable y can be either directly observed or a latent construct.

The measurement model for the first-order latent predictors, for instance ξ_x, are described by the following confirmatory factor analysis (CFA) framework:

$$\mathbf{x} = \boldsymbol{\tau_x} + \boldsymbol{\lambda_x}\xi_x + \boldsymbol{\delta_x}, \tag{3}$$

wherein, for each indicator $i = [1,\ 2,\ ...,\ p]$ associated with the latent predictor ξ_x, $\mathbf{x}$ denotes a $p \times 1$ vector of observed first-order indicators (i.e., indicators of ξ_x); ξ_x is a 1×1 scalar representing the latent variable; $\boldsymbol{\tau_x}$ is a $p \times 1$ vector of constant intercepts; $\boldsymbol{\lambda}_x$ is a $p \times p$ vector of factor loadings, and $\boldsymbol{\delta_x}$ is a $p \times 1$ vector of indicator-level measurement errors. Each measurement error δ_{x_i} is normally distributed with a mean of zero and a variance θ_{x_i}. Assuming local independence (i.e., first-order indicators are uncorrelated with each other when indicating the same latent variable), the variance-covariance matrix of all indicators' measurement errors is a diagonal matrix $\boldsymbol{\Theta}_{\delta_{\mathbf{x}}} = diag(\theta_{x_1}, \theta_{x_2}, ..., \theta_{x_p})$. This measurement model and its associated parameters similarly apply to ξ_m.

Kenny and Judd's original formulation of their model omitted the intercept α, a point later corrected by Jöreskog and Yang (1996) who revised the model under a set of assumptions. The revised latent interaction model is grounded in three primary assumptions related to multivariate normal distribution and independence: (1) The measurement errors of first-order indicators, the first-order latent predictors, and the disturbance term in the structural model are multivariate normal, uncorrelated, and independent to each other (i.e., $Corr[\delta, \xi] = 0$; $Corr[\zeta, \xi] = 0$; $Corr[\delta, \zeta] = 0$ where $Corr$ denotes the correlation index); (2) All measurement errors are mutually independent and uncorrelated to each other (i.e., $Corr[\delta_i, \delta_{i'}] = 0$ for $i \neq i'$); (3) The correlation between first-order latent predictors (i.e., $Corr[\xi_x, \xi_m]$) is assumed to be non-zero and freely estimated since $\xi_x\xi_m$ may have a non-normal distribution even though ξ_x and ξ_m are normally distributed with means of 0 (Jöreskog & Yang, 1996).

Algina and Moulder (2001) advanced Jöreskog and Yang's (1996) model by incorporating a mean-centering technique. They used mean-centered first-order indicators (e.g., $x_i - \mu_{x_i}$ where μ_{x_i} is the mean of x_i) to form product indicators (PI) that indicate the latent interaction term, which enhances the modeling approach by improving interpretability of parameter estimates, facilitating model convergence rate, and reducing bias of estimating the interaction effect (Algina & Moulder, 2001; Moulder & Algina, 2002; Marsh et al., 2004). Bsides, mean-centering first-order indicators diminishes problem of multicollinearity, clarifying the distinct contributions of first-order latent variable and their interaction (Schoemann & Jorgensen, 2021).

Unconstrained Product Indicator (UPI)

Although Algina and Moulder's modification improved the statistical properties of parameter estimation, their model necessitated implementation of complicated nonlinear constraints in their model. Constraints in tructural equation modeling (SEM) are predefined conditions or restrictions applied to model parameters to ensure model identifiability, theoretical consistency, and interpretability (Kline, 2016). Usual constraints include equality constraints (i.e., equal values constrained on two or more parameters), fixed-value constraints (i.e., specific values constrained on parameters), and nonlinear constraints (i.e., specific relationship between first-order latent variables and their interaction term as constraints). Nonlinear constraints are required in Algina and Moulder's model. Specifically, parameters are constrained that relate PIs to the interaction term (e.g., factor loadings, variances and covariances of error among PIs) and relate PIs to firsr-order indicators (e.g., error covariances between PIs and first-order indicators). Additionally, since the model is based on the assumption that first-order latent predictors have multivariate normal distribution, constraints on variances and covariances between them and their interaction term are required as well.

Marsh et al. (2004) explored the possibility of removing complex constraints and introduced the pivotal Unconstrained Product Indicator (UPI) method to simplify model specifications and reduce risks of erroneous specification and convergence issue. The structural model of UPI is the same as equation (2) except for omitting the intercept α. To illustrate, consider a measurement model in which the latent variables ξ_x and ξ_m are each associated with three indicators:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1} \\ \tau_{x_2} \\ \tau_{x_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1} \\ \lambda_{x_2} \\ \lambda_{x_3} \end{bmatrix} \begin{bmatrix} \xi_x \end{bmatrix} + \begin{bmatrix} \delta_{x_1} \\ \delta_{x_2} \\ \delta_{x_3} \end{bmatrix}, \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} \tau_{m_1} \\ \tau_{m_2} \\ \tau_{m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{m_1} \\ \lambda_{m_2} \\ \lambda_{m_3} \end{bmatrix} \begin{bmatrix} \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{m_1} \\ \delta_{m_2} \\ \delta_{m_3} \end{bmatrix} \tag{4}$$

Marsh et al. (2004) proposed two strategies of specifying the UPI model: the all-pair UPI and the matched-pair UPI. In the all-pair UPI, the latent interaction term is indicated by all possible configurations of pairs formed by the first-order indicators of ξ_x and ξ_m:

$$\begin{bmatrix} x_1m_1 \\ x_1m_2 \\ x_1m_3 \\ x_2m_1 \\ \ldots \\ x_3m_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1m_1} \\ \tau_{x_1m_2} \\ \tau_{x_1m_3} \\ \tau_{x_2m_1} \\ \ldots \\ \tau_{x_3m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1m_1} \\ \lambda_{x_1m_2} \\ \lambda_{x_1m_3} \\ \lambda_{x_2m_1} \\ \ldots \\ \lambda_{x_3m_3} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1m_1} \\ \delta_{x_1m_2} \\ \delta_{x_1m_3} \\ \delta_{x_2m_1} \\ \ldots \\ \delta_{x_3m_3} \end{bmatrix}, \tag{5}$$

where each PI is derived from multiplying two corresponding mean-centered first-order indicators, one from ξ_x and the other from ξ_m (e.g., the PI x_1m_1 is formed by the product of x_1 and m_1. The coefficients $\tau_{x_im_i}$, $\lambda_{x_im_i}$ and $\delta_{x_im_i}$ are estimated freely as intercepts, factor loadings and measurement errors, respectively. The total number of PIs are the multiplicative product of the number of first-order indicators for each latent predictor. In this case, nine unique configurations are generated ($3 \times 3 = 9$).

Regarding the matched-pair UPI, the indicators are matched to create PIs:

$$\begin{bmatrix} x_1m_1 \\ x_2m_2 \\ x_3m_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1m_1} \\ \tau_{x_2m_2} \\ \tau_{x_3m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1m_1} \\ \lambda_{x_2m_2} \\ \lambda_{x_3m_3} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1m_1} \\ \delta_{x_2m_2} \\ \delta_{x_3m_3}, \end{bmatrix} \tag{6}$$

This alternative formulation of UPI results in a significantly reduced number of PIs due to the straightforward configuration strategy. Marsh et al. (2004) suggested that the matched-pair UPI is more favorable based on two criteria: (1) It uses all available information by utilizing every first-order indicator; (2) It avoids redundancy by ensuring that no firs-order indicators are used more than once. This method is thus recommended for simplicity and effectiveness. Furthermore, they demonstrated that the matched-pair UPI excels by exhibiting lower bias and increased robustness in estimating the interaction effect, particularly under the violation of normality assumptions.

Since the mean of $\xi_x\xi_m$ is not equal to 0 even though ξ_x and ξ_m are assumed to have 0 means with non-zero correlation, Marsh et al. (2004) included a mean structure in their UPI model: $\kappa = (0,\ 0,\ Cov[\xi_x, \xi_m])^T$, where κ represents the model's mean structure. In this structure, the means of ξ_x and ξ_m are presumed to be 0, while the mean of the latent interaction term, denoted $Cov[\xi_x, \xi_m]$, is constrained as the covariance between ξ_x and ξ_m (see Algina & Boulder [2001] for more details). This adjustment ensures that the model accurately reflects the statistical relations between the first-order latent variables and their interaction term.

Lin et al. (2010) recently proposed a more refined method, the Double Mean Centering (DMC) strategy, and showed its advantages in eliminating the necessity of including a mean structure, simplifying the procedure of model specification, and demonstrating outstanding performance of parameter estimation under the violation of normality assumption. This method begins with mean-centering first-order indicators, and continues to mean-center PIs of the interaction term (e.g., $x_im_i - \mu_{x_im_i}$. Therefore we used the UPI method with DMC in this study.

Although UPI has more simplicity and better performance of parameter estimation compared to the classical model, a arbitrariness-complexity dilemma between the all-pair and the matched-pair methods is not well resolved (Foldness & Hadtvet, 2014). Consider a model with two complex psychological constructs that each may involve over 10 items to achieve sufficient coverage and depth of theory. The all-pair UPI method may potentially lead to a latent interaction term indicated by hundreds of PIs. More items can improve the representation of latent constructs and theoretically increase statistical power for detecting nuanced effects, but also result in a cumbersome model that negatively impacts interpretability, escalates computational demands, and overfits the sample. The matched-pair UPI strategy effectively simplifies model complexity by reducing the number of necessary PIs but introduces a challenge of indicator selection. Researchers may aggregate multiple observed indicators into fewer parcels (Jackman et al., 2011) or prioritize items with higher reliability for PI formation (Wu et al., 2013). However, there is not a consensus on the best strategy to form matched pairs, and the considerable arbitrariness across various alternative approaches introduces uncertainty in selecting the optimal strategy and complicates the decision-making process in model specification. Marsh et al. (2004) simplifies this process by applying the matched-pair UPI to the model with equal number of first-order indicators, but it is very likely that substantive researchers may need to deal with unbalanced numbers of first-order indicators.

Reliability Adjusted Product Indicator (RAPI)

The RAPI introduced by Hsiao et al. (2018) also forms PI, but it uses composite scores (sum or mean scores) of multiple first-order items. Specifically, it combines all first-order indicators into single indicators to indicate first-order latent variables, and forms PIs by multiplying the single indicators to indicate the latent interaction term. Accordingly, the formed PI is a single indicator as well. This method effectively circumvents the issue of arbitrariness in indicator selection while using all information without redundancy. RAPI adjusts for measurement error in composite scores by constraining error variances of single indicators, thus ensuring that parameter estimates are less biased. The model is succinctly represented as follows:

$$\begin{bmatrix} x_{comp} \\ m_{comp} \\ x_{comp} \cdot m_{comp} \end{bmatrix} = \begin{bmatrix} \tau_{x_{comp}} \\ \tau_{m_{comp}} \\ \tau_{x_{comp} \cdot m_{comp}} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_x \\ \xi_m \\ \xi_x \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_{comp}} \\ \delta_{m_{comp}} \\ \delta_{x_{comp} \cdot m_{comp}} \end{bmatrix}, \tag{7}$$

where x_{comp} and m_{comp} are the composite scores formed by their corresponding first-order indicators, and $x_{comp} \cdot m_{comp}$ is the formed PI indicating the latent interaction term. These composite scores serve as single indicators for their respective latent variables, with factor loadings uniformly constrained to 1. The measurement errors are represented by δ.

An important feature of the RAPI method is that it can account for measurement error within first-order indicators by including error-variance constraints computed using composite reliability. Although technically composite reliability estimates as part of error-variance contraints can be obtained by any existing methods, Hsiao et al. (2021) summarized and compared four normally used estimators for composite reliability: Cronbach's α (Cronbach, 1951), ω (McDonald, 1970; Raykov, 1997), the greatest lower bound reliability (GLB; Berge & Sočan, 2004), and Coefficient H (Hancock & Mueller, 2001). Suppose that $\rho_{xx'}$ denotes the estimated reliability index, the error variance of ξ_x can be shown as a function of the reliability index:

$$\hat{\sigma}^2_{\delta_x} = (1 - \rho_{xx'})\hat{\sigma}^2_x, \tag{8}$$

where $\hat{\sigma}^2_{\delta_x}$ represents the sample-estimated error variance and $\hat{\sigma}^2_x$ represents the sample-estimated variance of the indicator. The formula can be converted by linear transformation to show the relations between variances of the error and the latent predictor in terms of reliability: $\hat{\sigma}^2_{\delta_x} = [(1 - \rho_{xx'})/\rho_{xx'}]\hat{\sigma}^2_{\xi_x}$, where $\hat{\sigma}^2_{\xi_x}$ represents the estimated variance of ξ_x and $\hat{\sigma}^2_x = \hat{\sigma}^2_{\xi_x} + \hat{\sigma}^2_{\delta_x}$. Hence, under the assumption of independently and identically distributed measurement error, the equation for the error-variance constraint of the interaction term $\xi_x\xi_m$ can be derived:

$$\begin{aligned}\hat{\sigma}^2_{\delta_{xm}} =& \rho_{xx'}\hat{\sigma}^2_x(1-\rho_{mm'}\hat{\sigma}^2_m)+ \\ & \rho_{mm'}\hat{\sigma}^2_m(1-\rho_{xx'})\hat{\sigma}^2_x+ \\ & (1-\rho_{xx'})\hat{\sigma}^2_x(1-\rho_{mm'})\hat{\sigma}^2_m.\end{aligned} \tag{9}$$

More technical details are available in Appendix A of Hsiao et al. (2018).

The utilization of composite scores as single indicators significantly simplifies model specification, as the total number of PIs directly corresponds to the number of interaction terms. By accounting for measurement error, RAPI is expected to produce less biased estimates of interaction effects and exhibit enhanced statistical power. However, the method's effectiveness is contingent upon accurate estimation of reliability measures since inaccurate reliability estimates which serve as the basis for error constraints can lead to biased results. Despite its acceptable model complexity and approachable implementation, Hsiao et al. (2021) showed that RAPI may lead to non-positive definite matrices due to negative error variance and inflated interaction effect estimates, under conditions of low reliability (e.g., r = .70) and small sample size (e.g., N = 100). This suggests that RAPI may generate unstable interaction estimates under such conditions.

Two-stage Path Analysis with Interaction (2S-PA-Int)

The 2S-PA method, as proposed by Lai and Hsiao in 2021, introduces a refined approach to addressing measurement error within the context of multiple congeneric items by incorporating reliability adjustment. It is similar to RAPI but uses factor scores as single indicators to latent predictors. A key advancement of the 2S-PA approach is its capacity to assign observation-specific estimated reliability, thereby extending its applicability to ordered categorical items and accommodating distributions that deviate from normality (Lai & Hsiao, 2021; Lai et al., 2023). Besides, conventional SEM models typically estimate measurement and structural models simultaneously, which necessitates a considerable sample size to achieve satisfactory convergence rates (Kline, 2016; Kyriazos, 2018). To address this potential issue, the 2S-PA separates the step of specifying the measurement model from estimating the structural model, therefore alleviating computational burden and improving stability of parameter estimation.

At the first stage of 2SPA, researchers calculate factor scores ($\hat{F}$) using first-order indicators for each participant j for $j = 1, 2, ..., n$. Next, parallel to RAPI, the factor scores of latent predictors are multiplied to construct a PI for the interaction term $\xi_{x_j}\xi_{m_j}$:

$$\begin{bmatrix} \widehat{F}_{x_{ij}} \\ \widehat{F}_{m_{ij}} \\ \widehat{F}_{xm_{ij}} \end{bmatrix} = \begin{bmatrix} \tau_{x_{ij}} \\ \tau_{m_{ij}} \\ \tau_{xm_{ij}} \end{bmatrix} + \begin{bmatrix} \lambda_{x_{ij}} & 0 & 0 \\ 0 & \lambda_{m_{ij}} & 0 \\ 0 & 0 & \lambda_{xm_{ij}} \end{bmatrix} \begin{bmatrix} \xi_{x_j} \\ \xi_{m_j} \\ \xi_{x_j}\xi_{m_j} \end{bmatrix} + \begin{bmatrix} \delta_{x_{ij}} \\ \delta_{m_{ij}} \\ \delta_{xm_{ij}} \end{bmatrix}, \tag{10}$$

wherein the factor scores $\hat{F}_{x_j}$, $\hat{F}_{m_j}$ and the PI $\hat{F}_{xm_{ij}}$ are single indicators of the respective latent variables. The intercepts, factor loadings, and error variances are all model parameters to be freely estimated.

Although there are multiple ways of calculating factor scores (e.g., regression factor scores, expected-a-posterior factor scores; Devlieger et al., 2016; Estabrook & Neale, 2013), we used Bartlett factor scores because they align with the strengths of the 2S-PA framework (Bartlett, 1937). Bartlett scores produce unbiased estimates of true factor scores, enhancing the accuracy of our representation of the latent constructs and minimizing distortions that could arise from measurement errors (Hershberger, 2005). As a maximum likelihood-based procedure, it aligns with the estimation methods often used within SEM and with 2S-PA itself. This consistency further strengthens the reliability and interpretability of our analysis. In the case of using Bartlett scores as the factor scores, the Bartlett scores are adjusted to have the same units as latent variables, which means that the factor laodings relating the single indicators to latent variables are constrained to 1 (i.e., $\lambda_{x_{ij}} = \lambda_{m_{ij}} = \lambda_{xm_{ij}} = 1$).

Given that the focus of the current study is continuous variable, first-order indicators of ξ_x and ξ_m are continuous variables assumed to be normally distributed, and hence the corresponding error variances are constant for all observations. Accordingly the observation-specific subscript j from the above equations can be dropped in this study. The error variance constraint for the first-order latent predictors is $\sigma^2_{F_i}$ as σ_{F_i} is the estimated standard error of measurement. The error-variance constraint for the interaction term is defined similarly as equation (9). Alternatively speaking, the RAPI method is a special case of 2SPA where the composite scores are used for continuous items (Lai & Hsiao, 2021).

We argue that the 2S-PA-Int approach is a good alternative to existing methods of estiamting latent interaction effects for its simplicity in model complexity and clarity in model specification. Lai and Hsiao (2021) demonstrated that 2S-PA provides robust and precise estimates with less SE bias, lower Type I error rate, and higher convergence rates in small sample size and low reliability conditions. Hence we expect the 2S-PA-Int method to inherit the advantages and demonstrate desirable performance in latent interaction estimation.

Method

Simulation Design

Adapted from Hsiao et al. (2021), the current simulation study aimed to compare performance of UPI and RAPI with that of 2S-PA-Int on estimating latent interaction effects for continuous congeneric items. We investigated the bias and variance of interaction estimates generated by the three methods over various levels of sample size, reliability, and correlation between first-order latent variables. The generated population data was based on the model below with predefined parameter values:

$$
\begin{aligned}
x_i &= \tau_{x_i} + \lambda_{x_i}\xi_x + \delta_{x_i}; \\
m_i &= \tau_{m_i} + \lambda_{m_i}\xi_m + \delta_{m_i}; \\
y &= \tau_y + \gamma_x\xi_x + \gamma_m\xi_m + \gamma_{xm}\xi_x\xi_m + \zeta,
\end{aligned} \tag{11}
$$

where the path coefficients of two latent predictors (i.e., γ_x and γ_m) and their interaction term (i.e., γ_{xm}) were all set to 0.3 for the structural model. The first-order latent predictors ξ_x and ξ_m were simulated from standard normal distributions with means of 0 and variances fixed at 1, each indicated by three items (i.e., ξ_x indicated by $[x_1, x_2, x_3]$; ξ_m indicated by $[m_1, m_2, m_3]$). The first-order indicators and the dependent variable y were all observed continuous variables with normally distributed error. Accordingly, δ_{x_i}, δ_{m_i} and ζ were assumed to have multivariate normal distributions and be mutually independent. τ_{x_i}, τ_{m_i}, and τ_y were their corresponding constant intercepts and assumed to be 0. The first-order indicators were mean-centered for UPI, RAPI and 2S-PA-Int at the sample level.

Drawing from Jöreskog's (1971) concept, congeneric tests were defined as a set of observed items measuring a latent construct with different factor loadings and unique error terms. These error terms were assumed to be uncorrelated with each other and with the latent construct, reflecting random measurement error unique to each item. To align with this concept, we manipulated the factor loadings and error variances of first-order indicators to create sets of congeneric items in the measurement model. Specifically, the first, second, and third indicators were set to fixed values of 1.0, 0.9, and 0.75 for both first-order latent variables (i.e., $\lambda_{x_1} = \lambda_{m_1} = 1.0$, $\lambda_{x_2} = \lambda_{m_2} = 0.9$, $\lambda_{x_3} = \lambda_{m_3} = 0.75$). We involved reliability estimates to manipulate error variances since equation (9) demonstrates that the error variance of the interaction term was a function of first-order indicators' reliability, implying that the interaction effect could be impacted by the amount of measurement error. Hence we included reliability as a varying condition to explore how each method performed under three reliability conditions: .70, .80, and .90, which resulted in three levels of error variances. For each level of error variance, we systematically manipulated proportions of error variances each first-order indicator occupied. The proportions were set to maintain consistency with the design in Hsiao et al. (2021): 44% of the total error variance for the first indicator, 33% for the second, and 23% for the third. Then we obtained the manipulated error variances according to equation (8). For instance, $\theta_{x1},\ \theta_{x2},\ \theta_{x3}$ and $\theta_{m1},\ \theta_{m2},\ \theta_{m3}$ were $[3.01,\ 1.76,\ 0.78]$ when $\lambda_{x1},\ \lambda_{x2},\ \lambda_{x3} = \lambda_{m1},\ \lambda_{m2},\ \lambda_{m3} = [1,\ 0.9,\ 0.75]$, as the reliability was varied at .70, .80, and .90 respectively.

Following the recommendation by Marsh et al. (2004), $\xi_x\xi_m$ was represented through a matched-pair configuration of indicators in the UPI method, namely x_1m_1, x_2m_2, and x_3m_3. For the RAPI and 2SPA methods, $\xi_x\xi_m$ was loaded by single PIs. Specifically, for RAPI the interaction term's PI was the mean scores of first-order indicators, while for 2S-PA-Int was pre-computed Bartlett factor scores. To reduce the problem of multicollinearity between first-order latent predictors and the interaction term, the DMC strategy was applied to all the methods.

The literature on latent interaction methods showed a range of researcher-selected sample sizes from 20 to 5,000 (Chin, Marcolin, & Newsted, 2003; Lin et al., 2010; Cham et al., 2012), with common selections ranging from 100 to 500. Consequently, we selected N = 100, 250, and 500 to represent small, medium, and large sample sizes, respectively.

Based on the study design in Hsiao et al. (2021), we pre-specified three population correlations between latent predictors ($Corr[\xi_x, \xi_m]$): 0, 0.3, 0.6 as zero to large correlation. Given that the variance of y (i.e., σ^2_y), $\sigma^2_{\xi_x}$, and $\sigma^2_{\xi_x}$ was set to 1, ψ could be computed as $1 - R^2$ in which $R^2 = \gamma^2_x + \gamma^2_m + 2\gamma_x\gamma_m Corr[\xi_x, \xi_m] + \gamma^2_{xm}(1 + Corr[\xi_x, \xi_m]^2)$. Take $Corr[\xi_x, \xi_m] = 0$ as an example, $\psi = 1 - (0.3^2 + 0.3^2 + 2 \times 0.3 \times 0.3 \times 0 + 0.3^2 \times (1 + 0)^2) = 0.73$. Similarly, ψ was determined to be 0.668 and 0.590 for $Corr[\xi_x, \xi_m]$ equal to 0.3 and 0.6, respectively.

In summary, our study implemented a $3 \times 3 \times 3$ factorial design, accommodating variations across three sample sizes, three levels of correlation between first-order latent predictors, and three levels of reliability.

Evaluation Criteria

We chose widely used evaluation criteria that were summarized from 2000 replications to assess the accuracy and precision of interaction effect estimates (γ_{xm}) of the three methods.

Averaged Raw Bias and Standardized Bias. Standardized bias (SB) was used to evaluate averaged raw bias and accuracy of parameter estimates. It provided a normalized measure that allowed for comparing bias across different scales or units of measurement, and reflected how far an estimate was from its true value in standard error units. Hence SB was useful in comparisons where models often contained a variety of parameter types (e.g., factor loadings, path coefficients).

The Standardized Bias (SB) was defined through the averaged raw Bias (B):

$$SB = \frac{B(\gamma_{xm})}{SE_{\gamma_{xm}}}, \tag{12}$$

$$B(\gamma_{xm}) = R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm}), \tag{13}$$

where R was the total number of replication cycles that were counted from 1 to 2,000. $\hat{\gamma}_{xm_r}$ was the estimated interaction effect in each replication cycle r and γ_{xm} was the population parameter set at 0.3. $B(\gamma_{xm})$ was the averaged deviation of estimates, $\hat{\gamma}_{xm}$, from 0.3, and $SE_{\gamma_{xm}}$ represented the empirical standard error of $\hat{\gamma}_{xm}$ across replications. Collins et al. (2001) suggested that an absolute value of SB $\leq$ 0.40 would be considered acceptable for each replication condition.

Coverage Rate. The coverage rate with a 95% confidence interval (CI) served as a critical metric for evaluating the reliability and accuracy of simulation results. It was defined as the percentage of replications in which the Wald confidence interval captured the true interaction effect γ_{xm}. Low coverage rates meant that the proportion of times that γ_{xm} fell within the CI across replications was low, indicating that the model might have issues of misspecification, inappropriate estimation methods, small sample sizes, or violations of statistical assumptions. A coverage rate larger than 91% was considered acceptable (Muthén & Muthén, 2002).

Robust Relative Standard Error Bias and Outlier Proportion of SE. The relative standard error (SE) bias was used to evaluate the precision of $\hat{\gamma}_{xm}$. This criterion compared the empirical standard deviation of $\hat{\gamma}_{xm}$ with the sample-estimated standard error across replications:

$$Relative\ SE\ Bias = \frac{R^{-1}\Sigma_{r=1}^{R}(\widehat{SE}_r - SD)}{SD}, \tag{14}$$

where $\widehat{SE}_r$ was the sample-estimated standard error of $\hat{\gamma}_{xm}$ in a single replication cycle r and SD is the empirical standard deviation obtained from all replications. With SD being used as a reference variability measure of $\hat{\gamma}_{xm}$, smaller relative SE bias meant the estimated standard errors were closer to the referenced variability, and the uncertainty of $\hat{\gamma}_{xm}$ across replications was more accurately measured in each simulation condition. Absolute values of relative SE bias $\leq 10\%$ were considered acceptable and indicated that the standard errors were reasonably unbiased (Hoogland & Boomsma, 1998). SEM typically required a relatively large sample size to obtain sufficient information to reliably estimate model parameters. Insufficient sample sizes might result in largely biased SEs due to increased uncertainty around the parameter estimates (Bollen & Long, 1993; Byrne, 2016). Given that the conditions of small sample size (N = 100) and high amount of measurement error ($\rho = 0.7$) were included in this study design, a robust version of relative SE bias was calculated as an alternative to the regular one:

$$Robust\ Relative\ SE\ Bias = \frac{\widehat{MDN}(SE_r) - MAD}{MAD}, \tag{15}$$

where $\widehat{MDN}(SE_r)$ represented the median value of the estimated SE values and MAD is the empirical median-absolute-deviation of SE values. In the context of biased SEs, we did not assume a specific distribution of SEs (e.g., normal distribution) and hence we used the median due to its robustness to non-normal distributions with skewed data and outliers (Rousseeuw & Hubert, 2011). In addition, MAD measured variability around the median and could serve as a robust alternative to standard deviation that could be inflated by outliers or non-normality (Daszykowski et al., 2007). Besides, an outlier detection using the interquartile range (IQR) method was included as a supplemental information of SE estimates:

$$O_a \notin (Q_1 - 1.5 \times IQR,\ Q_3 + 1.5 \times IQR), \tag{16}$$

where O_a was an observation of outlier for $a = 1, 2, \ldots,$ b. IQR captured the spread of the middle 50% of the sample SEs by $IQR = Q_3\ -\ Q_1$, where Q_1 and Q_3 were the 25th percentile and the 75th percentile of the sample. The outlier proportion was then calculated by b/R where b represented the total count of identified outliers. Like the robust relative SE bias, the IQR method did not rely on the assumption of normal distribution, thus making it versatile across any distribution.

Root Mean Squre Error. The last criterion was the root mean square error (RMSE), calculated by taking the squared root of the sum of squared bias:

$$RMSE = \sqrt{R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm})^2}. \tag{17}$$

It quantified the average magnitude of the difference between the interaction estimates and the true value, reflecting both the bias and variability of the estimates across replications. Under one condition across 2,000 replication, a smaller RMSE value of a method indicated that it had relatively more accuracy than the other two methods in estimating $\hat{\gamma}_{xm}$ (Harwell, 2019). RMSE was most informative when comparing across methods under the same simulated conditions by isolating factors of sample size, model complexity, and the amount of disturbance.

Results

The results of the interaction effect estimated by RAPI, matched-pair UPI, and 2S-PA-Int were summarized and compared in terms of the average raw bias, the standardized bias, the relative standard error (SE) bias with outlier proportions, the 95% CI coverage rate of the interaction effect, and the root mean square error (RMSE) over 2,000 replications. Detailed statistics are displayed in Table 1, 2, 3, and 4, respectively. For all simulation conditions, the matched-pair UPI and 2S-PA-Int methods successfully converged without producing any inadmissible results. Models with the RAPI method did not fully converge under 8 of the 27 conditions, particularly those with low reliability (rho = 0.7) and small sample size (N = 100), and had a range of 1% ~ 12% non-convergence rate. Subsequent analyses did not include the inadmissible solutions generated by the RAPI method.

Average Raw Bias and Standardized Bias for γ_{xm}

As delineated in Table 1, an examination of all simulation conditions revealed that the absolute values of both the average raw bias (B) and the standardized biases (SB) associated with the interaction effect estimate (γ_{xm}) using the three methods consistently remained below the predetermined acceptable threshold of .40 (B = .00 ~ .08; SB = -.04 ~ .25). A discernible pattern in the impact of the correlation between the two first-order latent predictors on γ_{xm} was not identified. Regarding the influence of population reliability levels, all the methods demonstrated robustness to conditions of low reliability (i.e., $\rho = 0.7$). Notably, with an increase in population reliability levels, both the absolute SB and B exhibited declining trends across all the conditions with medium to high sample sizes (i.e., $N = 250$ and $N = 500$). For instance, when $N = 250$ and $Corr(\xi_x, \xi_m) = 0$, the absolute SB and B for the RAPI method decreased from .21(.03) to .03(.00) as ρ increased from .70 to .90. Similar decreasing trends were observable in the matched-pair UPI and 2S-PA-Int methods, where their absolute SB and B decreased from .08(.01) to .02(.00), and from 0.10 (.01) to .03(.00), respectively. As for $N = 100$, the same trends were still observed in 2S-PA-Int while two exceptions appeared in RAPI and matched-pair UPI respectively. The absolute SB and B for RAPI first increased from .14(.08) to .18(.03) and then decreased to .08(.01) as $Corr[\xi_x, \xi_{xm}] = 0$, and those for matched-pair UPI first increased from .10(.03) to .11(.02) and then decreased to .03(.00) as $Corr[\xi_x, \xi_{xm}] = 0.6$.

The B values generally became smaller as sample size increased for the three methods, which aligned with the statistical property of SEM models such that larger sample sizes tend to provide more accurate and reliable parameter estimates and reduce sampling errors. Nevertheless this pattern was not exactly consistent with the absolute SB because the empirical standard deviation of B decreased as the sample size increased, which might amplify the absolute SB. For instance, when $\rho = .70$ and $Corr[\xi_x, \xi_m] = 0$, the magnitude of raw average biases decreased from .08 to .01 for RAPI while the absolute SB first increased from .14 to .21 and then decreased to .19. The above findings revealed that the pattern displayed through B values might be masked by the corresponding empirical standard deviation, and the comparability of raw average biases need to be cautiously considered in standard units.

It was found that the absolute SB of RAPI and matched-pair UPI were almost positive while some of the 2S-PA-Int estimates were negative across simulation conditions. The results were consistent with previous findings for RAPI and matched-pair UPI such that they tended to provide overestimated interaction estimates with high correlations between first-order latent predictors and low reliability (Marsh et al., 2004; Hsiao et al., 2018). 2S-PA-Int did not show a clear sign of over or underestimation, indicating that the absolute SB values were more randomly distributed. Nevertheless, all the methods yielded comparably low standardized biases across simulation conditions, which was acceptable for practical use.

Relative SE Bias of γ_{xm}

Table 2 showed the robust relative standard error (SE) bias ratio with outlier proportions of SE when $\gamma_{xm} = 0.3$. All the values outside the -10% ~ 10% range were bolded. Generally, the values of robust relative SE bias were all below 10% for RAPI, matched-pair UPI, and 2S-PA-Int across the conditions of medium to high reliability level. The ranges were from .56%(1.55%) to 8.33%(1.65%) for RAPI, .09%(1.55%) to -8.96%(5.85%) for matched-pair UPI, and -.57%(1.40%) to -7.39%(1.30%) for 2S-PA-Int, which implied that the estimated SE values of $\hat{\gamma}_{xm}$ were not biased and $\hat{\gamma}_{xm}$ estimated by the three methods under medium to high reliability showed less variability across other conditions. Compared to 2S-PA-Int, matched-pair UPI produced two relative SE values outside the acceptable range under the conditions of small sample size ($N = 100$) and low reliability ($\rho = .70$): -11.52%(8.15%) and -14.14%(8.40%), meaning that the SE values were negatively biased. As for RAPI, unacceptable relative SE biases were generated across various conditions under low reliability ($\rho = .7$). The outlier proportions of SEs identified by the IQR method showed declining trends as sample size increased and reliability levels improved for all the methods, meaning that the estimation of γ_{xm} became more accurate and stable with less extreme values. For instance, the proportion of outliers for RAPI decreased from 10.90% to 5.55% and to 2.60% under $N = 100$ and $\rho = .70$, while within the condition of $N = 100$ the proportion decreased from 10.90 to 5.40 to 1.90 as ρ increased.

However, the robust relative SE bias did not demonstrate a clear pattern associated with the population reliability value and sample size for the three methods. It was found that even though under the condition of large sample size and high reliability, the relative SE bias could be higher then those under worse conditions. For example, when $N = 500$ and $\rho = .90$, the relative SE bias produced by 2S-PA-Int was -2.41%, while the value was -1.02 under $N = 100$ and $\rho = .90$. It implied that the estimated SEs were unstably deviated from the true reference value (i.e., empirical standard deviation of γ_{xm}). Additionally, the overall values of the relative SE bias were negative for matched-pair UPI and 2S-PA-Int, and almost positive for RAPI, which indicated that the SEs were systematically underestimated for matched-pair UPI and 2S-PA-Int but overestimated for RAPI.

Coverage Rate of 95% CI of γ_{xm}

As shown in Table 3, the coverage rates of 95% CI were adequately within the acceptable range (91 - 98%) for RAPI and 2S-PA-Int across all the simulation conditions, with a range from 95.50% to 97.75% for RAPI and 93.10% to 95.50% for 2S-PA-Int. For UPI, three values that occurred under the condition of small sample size ($N = 100$) and low reliability level ($\rho = .70$) were beyond the acceptable range: 87.9%, 88.75%, and 89.65%; Nevertheless, the lowest coverage only showed a 2.1% gap to 91%. No clear trends of coverage rate were observed in terms of the sample size, the population reliability level, and the correlation between first-order latent variables within the methods. However, across the methods, it was observed that generally RAPI demonstrated the highest coverage rate, followed by 2S-PA-Int with the second highest, and UPI exhibiting the lowest coverage rate. This order revealed that the RAPI method had the highest chance of capturing the true interaction effect with 2S-PA-Int and UPI followed, when the true interaction effect existed.

RMSE of γ_{xm}

Table 4 showed that the RMSE values for γ_{xm} decreased as the sample size increased and the reliability level increased. Comparing RMSE across methods, 2S-PA-Int showed the least (or equally least) RMSE values across all the simulation conditions, indicating that 2S-PA-Int had a closer fit of to the data and more accurate estimation of the true γ_{xm}. For example, under the small sample size and low reliability, the RMSE values of 2S-PA-Int ranged from .20 to .32 while those of RAPI and matched-pair UPI ranged from .25 to .61 and .34 to .39 respectively. However, note that the differences on RMSE across the methods became less obvious under the condition of high reliability ($\rho = .90$), meaning that all the methods tended to produce more accurate and less unstable estimations of the interaction effect.

Discussion

Applied researchers often focus on complex relationships between variables, such as interactions. However, classical regression models, which assume variables are free of measurement error, have been shown to produce biased estimates. Consequently, latent variables approaches with the SEM framework are increasingly being considered. In this study, we reviewed and compared the performance of matched-pair UPI and RAPI with 2S-PA-Int in estimating interaction effects on congeneric items with varying factor loadings and errors.

We extended the 2S-PA model by Lai and Hsiao (2021) to support the latent interaction estimation, namely 2S-PA-Int. The major difference between the three methods is on the formation of the latent interaction term. Specifically, matched-pair UPI forms the latent interaction term by using multiple product indicators generated by first-order indicators, and thus it is a multiple-indicator method. Instead, RAPI and 2S-PA use composite scores and factor scores as single indicators to the latent interaction term, respectively. Our findings indicated that all three methods were capable of generating unbiased estimates of interaction effects by accounting for measurement errors, with all the absolute SB and B values estimates falling below the .40 threshold. Notably, RAPI and UPI exhibited substantially positive SB values, suggesting a tendency to overestimate interaction effects when true effects are present. These observations align with the results from Marsh et al. (2004) using items with congeneric factors (i.e., only factor loadings were varied), Hsiao et al. (2018) using tau-equivalent items (i.e., only error variances were varied), and Hsiao et al. (2021) using congeneric items, where matched-pair UPI and RAPI slightly overestimated interaction coefficients when true interaction effects were nonzero, albeit to an acceptable degree. Our results echoed that RAPI and matched-pair UPI should be used with caution when researchers prefer to be more conservative with estimated effects.

Higher coverage rates with 95% CI for RAPI around 95% ~ 97% were observed in our results, implying that RAPI has higher chance and accuracy in capturing true interaction effects within the 95% confidence intervals, compared to matched-pair UPI and 2S-PA-Int. 2S-PA-Int estimated interaction effects with acceptable coverage rates as well, though slightly lower than those estimated by RAPI, implying that 2S-PA-Int is able to capture the true effects with high likelihood. Matched-pair UPI was affected mostly by small sample size and low reliability level, which implied that it was not as robust as RAPI and 2S-PA-Int and not recommended to use under this condition. Our results showed consistency with past research mentioned above; however, Marsh et al. (2004) did not test matched-pair UPI on fully congeneric items and it may imply that matched-pair UPI has less chance of capturing true effects with varied error variances within first-order indicators.

Sample size and the level of reliability significantly influenced the estimation of non-zero interaction effects. The SB and B values were sensitive to low sample size and high amount of measurement error, reflected by the sample-estimated reliability, such that they generally became smaller with increased sample size and decreased error for all the methods. It means that RAPI, matched-pair UPI, and 2S-PA-Int tend to have better performance in estimating interaction effects with larger sample sizes. Within the same sample size, higher reliability levels of first-order items generally result in more unbiased estimated for all the methods in most cases. The relative SE biases showed similar patterns with sample size and reliability level for the three methods, while RAPI generally exhibited larger biases than matched-pair UPI and 2S-PA-Int especially under small sample size and low reliability level. Thus RAPI is more inclined to generate unstable interaction estimates under such conditions. Overall, although three methods had at least one case of relative SE bias outside the acceptable range, 2S-PA-Int was slightly more stable under most of conditions. In terms of RMSE, it was obviously affected by both sample size and reliability level. For the three methods, as sample size and reliability individually or jointly increase, the RMSE values demonstrated declining trends, meaning that the interaction estimates showed more accuracy and variability. Despite the consistent trends, the 2S-PA-Int method produced estimates with less RMSE particularly under small sample size and low reliability level and thus 2S-PA-Int is more robust to these conditions. Taking all the evaluation criteria into account, 2S-PA-Int shows ample potential to serve as a good alternative to RAPI and matched-UPI for latent interaction estimation.

Revisiting Marsh's criteria of a good model of estimating latent interaction effects, 2S-PA-Int is practically preferable in terms of simple model specification as a single-indicator model and comprehensive usage of information by using factors scores based on all first-order indicators. Specifically, models overloaded with indicators may have difficulties in reaching convergence due to the intricate covariance structures to be estimated, potentially resulting in non-identifiable models (Bollen, 1989). Furthermore, Byrne (2016) highlights that excessive indicators can introduce redundancy, complicating the model unnecessarily and increasing the likelihood of estimation problems. Thus, 2S-PA-Int should be a safer alternative to matched-pair UPI especially with small sample size and low reliability level. Compared to RAPI, 2S-PA-Int is more advantageous in terms of stability and accuracy of interaction estimates.

Limitations and Future Direction

A few limitations in the current study are discussed below. First, Hsiao et al. (2018) mentioned that RAPI may be more approachable when researchers do not have the access of original data and have to analyze secondary data since composite scores are usually reported with reliability index (e.g., usually Cronbach's α). Since reporting factor scores with standard errors is still not a commonly applied practice, some secondary dataset may not contain computed factors scores and thus 2S-PA-Int is not applicable in this case. Second, currently the congeneric items in this study design are all continuous with normal distributions. Given that categorical data is frequently used in psychology research to capture the qualitative aspects of human behavior, attitudes, and characteristics (Brown, 2015; Kline, 2016), 2S-PA-Int has not been evaluated and should be studied with categorical items in the future. Third, the study designs in the past methodological paper on latent interaction effects were almost simply structured with two latent predictors and one interaction term, which could be insufficient to accommodate more complicated real-world scenarios, such as multiple interaction terms. Besides, multilevel design rencently is increasingly used in educational, counseling, and organizational research (e.g., students nested in classrooms, patients nested in clinics, employees nested in companies). Thus, it is worth exploring the potential of 2S-PA-Int with complicated data types and structures with varied sample sizes and reliability levels.

References

Carte, T., & Russell, C. J. (2003). *In pursuit of moderation: Nine common errors and thir solutions.*

Table 1

Standardized Bias (Average Raw Bias) for $\gamma_{xm}(= 0.3)$ over 2000 Replications.

		RAPI			Matched-Pair UPI			2S-PA-Int		
N	$Corr(\xi_x, \xi_m)$	$\rho = .70$	$\rho = .80$	$\rho = .90$	$\rho = .70$	$\rho = .80$	$\rho = .90$	$\rho = .70$	$\rho = .80$	$\rho = .90$
100	0	0.14 (0.08)	0.18 (0.03)	0.08 (0.01)	0.13 (0.05)	0.13 (0.03)	0.04 (0)	0.14 (0.04)	0.06 (0.01)	0.02 (0)
	0.3	0.19 (0.08)	0.13 (0.02)	0.04 (0)	0.13 (0.05)	0.11 (0.02)	0.01 (0)	0.16 (0.03)	0.07 (0.01)	-0.02 (0)
	0.6	0.25 (0.06)	0.16 (0.02)	0.06 (0.01)	0.1 (0.03)	0.11 (0.02)	0.03 (0)	0.16 (0.03)	0.06 (0.01)	0.02 (0)
250	0	0.21 (0.03)	0.12 (0.01)	0.03 (0)	0.08 (0.01)	0.09 (0.01)	-0.02 (0)	0.1 (0.01)	0.04 (0)	-0.03 (0)
	0.3	0.23 (0.03)	0.14 (0.01)	0.05 (0)	0.13 (0.02)	0.11 (0.01)	0 (0)	0.11 (0.01)	0.06 (0)	0 (0)
	0.6	0.2 (0.02)	0.09 (0.01)	0.03 (0)	0.15 (0.02)	0.05 (0)	-0.01 (0)	0.12 (0.01)	0.03 (0)	-0.01 (0)
500	0	0.19 (0.01)	0.1 (0.01)	0.09 (0)	0.09 (0.01)	0.05 (0)	0.04 (0)	0.09 (0.01)	0.03 (0)	0.03 (0)
	0.3	0.16 (0.01)	0.06 (0)	0.02 (0)	0.12 (0.01)	0.02 (0)	-0.02 (0)	0.07 (0)	-0.01 (0)	-0.04 (0)
	0.6	0.16 (0.01)	0.09 (0)	0.05 (0)	0.12 (0.01)	0.05 (0)	0.01 (0)	0.09 (0)	0.03 (0)	0.01 (0)

Note. N = sample size; $Corr(\xi_x, \xi_m)$ = correlation between ξ_x and ξ_m; ρ = reliability level; RAPI = reliability-adjusted product indicator method; Matched-Pair UPI = matched-pair product unconstrained indicator method; 2S-PA-Int = two-stage path analysis with interaction method. Average raw bias are shown in pararenthese. Note that numerical values have been rounded to two decimal places for consistency, which means that some values, while very close to 0 but not exactly 0, are displayed as 0.

Table 2

Robust Relative Standard Error (SE) Bias Ratio (Outlier Proportion of SE; %) for $\gamma_{xm}(=0.3)$ over 2000 Replications.

		RAPI			Matched-Pair UPI			2S-PA-Int		
N	$Corr(\xi_x, \xi_m)$	$\rho = .70$	$\rho = .80$	$\rho = .90$	$\rho = .70$	$\rho = .80$	$\rho = .90$	$\rho = .70$	$\rho = .80$	$\rho = .90$
100	0	7.38 (10.90)	3.77 (5.40)	-1.38 (1.90)	**-11.52 (8.15)**	-8.96 (5.85)	-7.67 (3.45)	-8.07 (8.05)	-5.7 (3.55)	-7.39 (1.30)
	0.3	**10.68 (9.10)**	6.27 (6.15)	2.84 (1.65)	**-14.14 (8.40)**	-4.44 (4.85)	-0.73 (2.10)	-5.34 (6.85)	-4.64 (4.20)	-1.02 (1.25)
	0.6	**13.3 (9.20)**	7.6 (4.70)	2.74 (1.90)	-9.21 (6.05)	-8.79 (4.85)	-0.84 (2.85)	-7.67 (6.40)	-4.48 (3.50)	-1.81 (1.70)
250	0	9.29 (5.55)	1.86 (2.80)	0.16 (1.10)	-4.22 (3.85)	-5.25 (3.70)	-1.49 (2.00)	-9.44 (4.25)	-6.73 (1.85)	-3.71 (0.80)
	0.3	3.46 (5.35)	7.96 (2.20)	2.74 (1.55)	-8.04 (3.30)	0.41 (3.45)	0.58 (2.25)	**-10.8 (3.85)**	-0.57 (1.40)	-3.24 (1.15)
	0.6	**12.84 (4.35)**	3.84 (2.70)	2.34 (1.40)	-8.23 (3.70)	-6.85 (3.35)	-0.8 (1.85)	-5.98 (3.35)	-4.7 (1.50)	-3.43 (1.25)
500	0	**13.74 (2.60)**	4.65 (1.65)	0.56 (1.20)	-2.21 (2.05)	-0.66 (2.80)	-2.54 (1.50)	-3.47 (2.15)	-3.32 (1.05)	-3.62 (1.05)
	0.3	**10.34 (3.85)**	4.88 (1.60)	2.25 (0.95)	-5.23 (3.15)	-2.59 (2.70)	0.09 (1.55)	-6.67 (2.90)	-4.81 (1.60)	-2.41 (0.90)
	0.6	**13.7 (3.40)**	8.33 (1.65)	2.23 (0.65)	-1.89 (3.30)	-4.03 (1.75)	-1.88 (0.60)	-2.47 (2.30)	-1.71 (1.85)	-2.97 (0.65)

Note. N = sample size; $Corr(\xi_x, \xi_m)$ = correlation between ξ_x and ξ_m; ρ = reliability level; RAPI = reliability-adjusted product indicator method; Matched-Pair UPI = matched-pair product unconstrained indicator method; 2S-PA-Int = two-stage path analysis with interaction method. Outlier proportions of SE are shown in parenthese and all the numbers were percentages. Note that relative SE bias values outside the acceptable range of [-10%, 10%] are bolded.

Table 3

95 % Confidence Interval (CI) Coverage Rate for $\gamma_{xm}(=0.3)$ over 2000 Replications.

		RAPI			Matched-Pair UPI			2S-PA-Int		
N	$Corr(\xi_x, \xi_m)$	$\rho = .70$	$\rho = .80$	$\rho = .90$	$\rho = .70$	$\rho = .80$	$\rho = .90$	$\rho = .70$	$\rho = .80$	$\rho = .90$
100	0	96.75	97.15	96.45	**87.9**	91.30	94.55	94.80	94.55	94.80
	0.3	97.05	96.65	96.15	**88.75**	92.30	94.20	95.50	94.30	94.80
	0.6	97.50	96.85	95.50	**89.65**	92.35	93.65	95.20	94.45	94.60
250	0	96.85	96.55	95.70	91.05	93.85	94.00	94.20	94.50	94.50
	0.3	97.15	97.75	95.70	91.65	95.00	94.70	94.55	95.20	94.35
	0.6	96.75	96.00	96.05	93.5	93.30	94.90	93.50	93.75	94.60
500	0	97.20	96.20	96.20	93.65	94.40	94.85	94.15	94.20	95.15
	0.3	96.30	96.60	95.50	93.7	94.60	94.10	93.10	94.10	94.70
	0.6	97.45	96.50	96.15	93.6	94.55	95.20	93.85	94.25	95.40

Note. N = sample size; $Corr(\xi_x, \xi_m)$ = correlation between ξ_x and ξ_m; ρ = reliability level; RAPI = reliability-adjusted product indicator method; Matched-Pair UPI = matched-pair product unconstrained indicator method; 2S-PA-Int = two-stage path analysis with interaction method. Coverage rates not reaching the acceptable threshold of 91% are bolded.

Table 4

Root Mean Square Error (RMSE) for $\gamma_{xm}(= 0.3)$ over 2000 Replications.

		$\rho = .70$			$\rho = .80$			$\rho = .90$		
N	$Corr(\xi_x, \xi_m)$	RAPI	Matched-Pair UPI	2S-PA-Int	RAPI	Matched-Pair UPI	2S-PA-Int	RAPI	Matched-Pair UPI	2S-PA-Int
100	0	0.61	0.39	0.32	0.16	0.20	0.15	0.12	0.13	0.11
	0.3	0.42	0.40	0.20	0.17	0.18	0.13	0.11	0.11	0.10
	0.6	0.25	0.34	0.20	0.12	0.15	0.11	0.09	0.10	0.09
250	0	0.12	0.14	0.11	0.09	0.11	0.08	0.07	0.07	0.07
	0.3	0.11	0.14	0.10	0.08	0.09	0.07	0.06	0.07	0.06
	0.6	0.09	0.12	0.08	0.07	0.08	0.07	0.06	0.06	0.05
500	0	0.07	0.09	0.07	0.06	0.07	0.06	0.05	0.05	0.05
	0.3	0.07	0.09	0.07	0.05	0.06	0.05	0.04	0.05	0.04
	0.6	0.06	0.08	0.05	0.05	0.05	0.04	0.04	0.04	0.04

Note. N = sample size; $Corr(\xi_x, \xi_m)$ = correlation between ξ_x and ξ_m; ρ = reliability level; RAPI = reliability-adjusted product indicator method; Matched-Pair UPI = matched-pair product unconstrained indicator method; 2S-PA-Int = two-stage path analysis with interaction method. Note that the methods are grouped in the second-order header for comparing RMSE under the same conditions.