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

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

Modeling interaction effects within the latent variable modeling framework has become increasingly popular in psychological research as it facilitates the exploration of in-depth theory and complex data structure. Comprared to the extensitvely used regression-based approaches assuming error-free variables, the latent interaction approach is able to account for measurement error and produce estimates with less bias and more accurate standard error. In this study, we investigated three product indicator methods based on structural equation modeling (SEM): Matched-pair Unconstrained Product Indicator (UPI), Reliability-Adjusted Product Indicator (RAPI), and an extended model based on the two-stage path analysis (2S-PA) framework, namely 2S-PA-Int, by conducting a simulation study with 2,000 replications. The results showed that 2S-PA-Int produced consistently less standardized bias, acceptable relative SE bias and coverage rates, and lower RMSE values than matched-pair UPI and RAPI, particularly under the conditions of small sample size and low reliability. Generally, 2S-PA-Int showed promising statistical properties and simpler model specification, indicating that it could serve as a competitive alternative of existing methods. Future research directions of 2S-PA-Int were 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 effects) rather than simple bivariate relationships, as the real world is rarely simple and straightforward (Carte & Russell, 2003; Cunningham & Ahn, 2019; MacKinnon & Luecken, 2008). For example, previous research has demonstrated that exercise can help people lose weight. However, there is growing interest in understanding the specifics of this relationship, including the mechanisms, timing, target groups, and conditions under which exercise is most effective for weight loss. Research into moderation (or interaction) provides insights into these questions by exploring how third variables, or a set of additional variables, influence the dynamics between the variables of interest.

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

$$Y = b_0 + b_1X + b_2Z + b_3XZ + \epsilon, \tag{1}$$

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 error 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 when measurement errors are present in empirical research (Bollen, 1989; Carroll et al., 2006; Cohen et al., 2003), and the bias may be more seriously inflated for interaction effects (Anderson et al., 1996). 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). A growing body of research has shown that moderation models based on SEM provide reliably true relationships among latent constructs (Cham et al., 2012; Maslowsky et al., 2015; Mueller, 1997; Steinmetz et al., 2011).

The Two-Stage Path Analysis (2S-PA; Lai & Hsiao, 2022) method models pathway relationships among latent variables through the use of factor scores. Simulation studies have shown its ability to yield parameter estimates with reduced standard error bias, enhanced convergence rates, and improved management of Type I error, particularly in small sample contexts (Lai et al., 2023; Lai & Hsiao, 2022). Given its promising statistical property, simpler model specification, and easier implementation in widely used softwares, 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 using the product indicator method, 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) initially introduced a foundational structural model for estimating latent interaction effects, specifically addressing a scenario with two latent predictors and their 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 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 an observed variable or a latent construct.

The measurement model for the first-order latent predictors, for instance ξ_x, is 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); $\boldsymbol{\tau_x}$ is a $p \times 1$ vector of constant intercepts; $\boldsymbol{\lambda}_x$ is a $p \times 1$ 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 of 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 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) refined Jöreskog and Yang's (1996) model by using 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 model by improving interpretability of parameter estimates, facilitating model convergence rate, and reducing bias of estimation (Algina & Moulder, 2001; Marsh et al., 2004; Moulder & Algina, 2002). Furthermore, mean-centering first-order indicators helps mitigate the issue of multicollinearity, thereby more clearly distinguishing the contributions of the first-order latent variables and their interactions, as noted by Schoemann and Jorgensen (2021).

Unconstrained Product Indicator (UPI)

While Algina and Moulder (2001) significantly improved the model, their approach required complicated nonlinear constraints on parameters of PIs and the interaction term. Constraints in SEM are predefined conditions or restrictions applied to model parameters to ensure model identifiability, theoretical consistency, and interpretability (Kline, 2016). Suppose that x_2 and m_2 are two first-order indicators of their corresponding latent predictors ξ_x and ξ_m, and their formed PI is x_2m_2. Then x_2m_2 can be decomposed using the measurement model of x_2 and m_2:

$$x_2m_2 = (\lambda_{x_2}\xi_x + \delta_{x_2})(\lambda_{m_2}\xi_m + \delta_{m_2}), \tag{4}$$

where λ is the factor loading, ξ is the first-order latent variable, and δ is the error term of first-order indicators. After expanding the equation, it can be shown that the factor loading of this formed PI is a function of first-order indicators' factor loadings, such that $\lambda_{x_2m_2} = \lambda_{x_2}\lambda_{m_2}$. Similarly, the error term can be derived as a function of parameters from first-order indicators: $\delta_{x_2m_2} = \lambda_{x_2}\xi_x\delta_{m_2} + \lambda_{m_2}\xi_m\delta_{x_2} + \delta_{x_2}\delta_{m_2}$. As the number of first-order indicators increases, the model specification becomes overwhelmingly cumbersome due to resulting nonlinear constraints, which may lead to convergence issue.

Marsh et al. (2004) investigated ways to eliminate complex constraints and introduced the innovative Unconstrained Product Indicator (UPI) method, which simplifies model specification and reduces the likelihood of 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:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1} \\ \tau_{x_2} \\ \tau_{x_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1} \\ \lambda_{x_2} \\ \lambda_{x_3} \end{bmatrix} \begin{bmatrix} \xi_x \end{bmatrix} + \begin{bmatrix} \delta_{x_1} \\ \delta_{x_2} \\ \delta_{x_3} \end{bmatrix}, \tag{5}$$

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} \tau_{m_1} \\ \tau_{m_2} \\ \tau_{m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{m_1} \\ \lambda_{m_2} \\ \lambda_{m_3} \end{bmatrix} \begin{bmatrix} \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{m_1} \\ \delta_{m_2} \\ \delta_{m_3} \end{bmatrix} \tag{6}$$

Marsh et al. (2004) presented two methods for specifying UPI: the all-pair UPI and the matched-pair UPI. In the all-pair UPI model, the latent interaction term is represented by all possible pairings of the first-order indicators of ξ_x and ξ_m:

$$\begin{bmatrix} x_1m_1 \\ x_1m_2 \\ x_1m_3 \\ x_2m_1 \\ ... \\ x_3m_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1m_1} \\ \tau_{x_1m_2} \\ \tau_{x_1m_3} \\ \tau_{x_2m_1} \\ ... \\ \tau_{x_3m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1m_1} \\ \lambda_{x_1m_2} \\ \lambda_{x_1m_3} \\ \lambda_{x_2m_1} \\ ... \\ \lambda_{x_3m_3} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1m_1} \\ \delta_{x_1m_2} \\ \delta_{x_1m_3} \\ \delta_{x_2m_1} \\ ... \\ \delta_{x_3m_3} \end{bmatrix}, \tag{7}$$

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 PIs are formed ($3 \times 3 = 9$).

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

$$\begin{bmatrix} x_1m_1 \\ x_2m_2 \\ x_3m_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1m_1} \\ \tau_{x_2m_2} \\ \tau_{x_3m_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1m_1} \\ \lambda_{x_2m_2} \\ \lambda_{x_3m_3} \end{bmatrix} \begin{bmatrix} \xi_x\xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1m_1} \\ \delta_{x_2m_2} \\ \delta_{x_3m_3} \end{bmatrix} \tag{8}$$

This alternative formulation results in a significantly reduced number of PIs due to its straightforwardness. Marsh et al. (2004) suggested that the matched-pair UPI is more favorable according to two criteria: (1) It uses all available information by utilizing every first-order indicator; (2) It avoids redundancy by ensuring that no first-order indicators are used more than once. This method is thus recommended for simplicity and effectiveness. Moreover, they showed that the matched-pair UPI approach performs as well as the all-pair model by demonstrating low bias and robustness to non-normal data, whereas the matched-pair model is more preferable in terms of simplicity and efficiency.

Since the mean of $\xi_x\xi_m$ may not equal to 0 even though ξ_x and ξ_m are assumed to have 0 means, Marsh et al. (2004) included a mean structure in their UPI model: $\kappa = (0,\ 0,\ Cov[\xi_x, \xi_m])^T$, where κ should be the means of the three latent variables (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) further simplified the model by proposing a Double Mean Centering (DMC) strategy, wherein PIs composed of paired mean-centered first-order indicators are mean-centered again (e.g., $x_i m_i - \mu_{x_i m_i}$. DMC eliminates the necessity of including a mean structure in the UPI model and demonstrates good performance of parameter estimation under the violation of normality assumption. Therefore we used the UPI method with DMC in this study.

Although UPI with DMC has simpler model specification and better performance of parameter estimation compared to the classical model, an arbitrariness-complexity dilemma between the all-pair and the matched-pair methods has not been well resolved (Foldnes & Hagtvet, 2014). Consider a model with two complex psychological constructs as latent predictors that each may need to involve over 10 indicators to achieve sufficient coverage of theory. The all-pair UPI method may potentially lead to a latent interaction term indicated by hundreds of PIs. Although a plenty of items can improve the representation of latent constructs and theoretically increase statistical power for detecting nuanced effects, they likely build up a cumbersome model that negatively impacts interpretability, escalates computational demands, and overfits the sample. The matched-pair UPI strategy effectively simplifies the model by reducing the number of necessary PIs, but also introduces a challenge of PI selection when substantive researchers need to deal with unbalanced numbers of first-order indicators. For unbalanced indicators, researchers need to make a decision on how to properly form PIs as there are multiple solutions. They 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. To address this issue, Wu et al. (2013) investigated two solutions in which researchers could form PIs by using highly reliable first-order indicators (i.e., items with higher factor loadings) while ignoring those with low reliability, or by matching parcels of the larger group of first-order indicators with indicators of the smaller group. They recommended to form PIs according to the order of item reliability.

Reliability Adjusted Product Indicator (RAPI)

The RAPI method 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 (SIs) to indicate first-order latent variables, and forms PIs by multiplying the SIs to indicate the latent interaction term. Accordingly, the formed PI is a SI 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 SIs, thus ensuring that parameter estimates are less biased. The model is succinctly represented as follows:

$$\begin{bmatrix} x_{comp} \\ m_{comp} \\ x_{comp} \cdot m_{comp} \end{bmatrix} = \begin{bmatrix} \tau_{x_{comp}} \\ \tau_{m_{comp}} \\ \tau_{x_{comp} \cdot m_{comp}} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_x \\ \xi_m \\ \xi_x \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_{comp}} \\ \delta_{m_{comp}} \\ \delta_{x_{comp} \cdot m_{comp}} \end{bmatrix}, \tag{9}$$

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 SIs for their respective latent variables, with factor loadings uniformly constrained to 1 for model identification. The measurement errors are represented by δs.

A key characteristic of the RAPI method is its ability to accommodate measurement error in first-order indicators through the incorporation of error-variance constraints, which are calculated using composite reliability. Although technically composite reliability estimates as part of error-variance constraints can be obtained by any available methods, Hsiao et al. (2018) 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 (Ten Berge & Sočan, 2004), and Coefficient H (Hancock &

Mueller, 2011). 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:

$$\hat{\sigma}^2_{\delta_x} = (1 - \rho_{xx'})\hat{\sigma}^2_x, \tag{10}$$

where $\hat{\sigma}^2_{\delta_x}$ represents the estimated error variance and $\hat{\sigma}^2_x$ represents the 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 $\hat{\sigma}^2_x = \hat{\sigma}^2_{\xi_x} + \hat{\sigma}^2_{\delta_x}$ according classical test theory (Lord et al., 1968). 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:

$$\begin{aligned}\hat{\sigma}^2_{\delta_{xm}} =& \rho_{xx'}\hat{\sigma}^2_x(1 - \rho_{mm'}\hat{\sigma}^2_m) + \\ & \rho_{mm'}\hat{\sigma}^2_m(1 - \rho_{xx'})\hat{\sigma}^2_x + \\ & (1 - \rho_{xx'})\hat{\sigma}^2_x(1 - \rho_{mm'})\hat{\sigma}^2_m.\end{aligned} \tag{11}$$

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

The utilization of composite scores as SIs 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 (2022), introduces an alternative 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 SIs 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 et al., 2023; Lai & Hsiao, 2022). 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, 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 obtain factor scores 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 $\xi_{x_j}\xi_{m_j}$:

$$\begin{bmatrix} \tilde{x}_j \\ \tilde{m}_j \\ \widetilde{xm}_j \end{bmatrix} = \begin{bmatrix} \tau_{\tilde{x}_j} \\ \tau_{\tilde{m}_j} \\ \tau_{\widetilde{xm}_j} \end{bmatrix} + \begin{bmatrix} \lambda_{\tilde{x}_j} & 0 & 0 \\ 0 & \lambda_{\tilde{m}_j} & 0 \\ 0 & 0 & \lambda_{\widetilde{xm}_j} \end{bmatrix} \begin{bmatrix} \xi_{x_j} \\ \xi_{m_j} \\ \xi_{x_j}\xi_{m_j} \end{bmatrix} + \begin{bmatrix} \delta_{\tilde{x}_j} \\ \delta_{\tilde{m}_j} \\ \delta_{\widetilde{xm}_j} \end{bmatrix}, \tag{12}$$

wherein the factor scores $\tilde{x}_j$, $\tilde{m}_j$ and the PI $\widetilde{xm}_j$ are SIs of the respective latent variables. The intercepts, factor loadings, and error variances are all model parameters to be freely estimated.

Researchers can use multiple ways to calculate factor scores (e.g., regression factor scores, expected-a-posterior factor scores), as reviewed in Estabrook and Neale (2013). We used Bartlett factor scores that are adjusted to have the same units as latent variables and constrained their factor loadings to 1 for model identification (i.e., $\lambda_{\tilde{x}_j} = \lambda_{\tilde{m}_j} = \lambda_{\widetilde{xm}_j} = 1$), as shown in Devlieger et al. (2016) and Lai et al. (2023).

Given that the focus of the current study is on continuous variables and the first-order indicators of ξ_{x_j} and ξ_{m_j} are assumed to be normally distributed, the corresponding error variances are constant for all observations. The error variance constraints for factor scores are $\hat{\sigma}^2_{\tilde{x}_j}$ where $\hat{\sigma}_{\tilde{x}_j}$ is the estimated standard error of measurement of the factor score $\tilde{x}$ for the person j. The error-variance constraint for the interaction term is defined similarly as equation (11). Alternatively speaking, the RAPI method is a special case of 2SPA where the composite scores are used for continuous items (Lai & Hsiao, 2022).

In this paper, we investigate whether the 2S-PA-Int approach is a good alternative to existing methods of estimating latent interaction effects, for its simplicity in model complexity and clarity in model specification. Lai and Hsiao (2022) 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 the performance of UPI, RAPI and 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 pre-defined parameter values:

$$
\begin{aligned}
x_i &= \tau_{x_i} + \lambda_{x_i}\xi_x + \delta_{x_i}; \\
m_i &= \tau_{m_i} + \lambda_{m_i}\xi_m + \delta_{m_i}; \\
y &= \tau_y + \gamma_x\xi_x + \gamma_m\xi_m + \gamma_{xm}\xi_x\xi_m + \zeta,
\end{aligned} \tag{13}
$$

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 the three methods.

Drawing from the concepts of Jöreskog (1971), 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 for 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., $\lambda_{x_1} = \lambda_{m_1} = 1.0$, $\lambda_{x_2} = \lambda_{m_2} = 0.9$, $\lambda_{x_3} = \lambda_{m_3} = 0.75$). According to equation (11), 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, and hence we varied reliability 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, We systematically manipulated error variance proportions for each indicator, consistent with Hsiao et al. (2021).: 44% of the total error variance for the first indicator, 33% for the second, and 23% for the third. Assuming that an appropriate reliability estimate has been obtained, the total error variance would be $[3.01,\ 1.76,\ 0.78]$ when $\lambda_{x_1},\ \lambda_{x_2},\ \lambda_{x_3} = \lambda_{m_1},\ \lambda_{m_2},\ \lambda_{m_3} = [1,\ 0.9,\ 0.75]$, as the reliability was varied at .70, .80, and .90 respectively. Taking the condition of $\rho = .70$ as an example, the error variances of three indicators would be manipulated as $1.32,\ 0.99,\ 0.69$.

Following the suggestion of using matched-pair UPI by Marsh et al. (2004), we included the model in which $\xi_x\xi_m$ was indicated by three pairs of PIs: 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 methodological literature on latent interaction models exhibited a range of researcher-selected sample sizes from 20 to 5,000 (Cham et al., 2012; Chin et al., 2003; Lin et al., 2010), with common selections ranging from 100 to 500. Consequently, we chose N = 100, 250, and 500 to represent small, medium, and large sample sizes, respectively. As for the correlation between first-order latent predictors, we followed the study design in Hsiao et al. (2021) and pre-specified three population correlations $Corr[\xi_x, \xi_m]$ (0, 0.3, 0.6) as zero to large correlation. Given that the variances of y (i.e., σ^2_y), $\sigma^2_{\xi_x}$, and $\sigma^2_{\xi_x}$ were all set to 1, ψ could be computed as $1 - R^2$ in which $R^2 = \gamma^2_x + \gamma^2_m + 2\gamma_x\gamma_m Corr[\xi_x, \xi_m] + \gamma^2_{xm}(1 + Corr[\xi_x, \xi_m]^2)$. For instance, $\psi = 1 - (0.3^2 + 0.3^2 + 2 \times 0.3 \times 0.3 \times 0 + 0.3^2 \times (1 + 0)^2) = 0.73$ for $Corr[\xi_x, \xi_m] = 0$. Similarly, $\psi = 0.668$ and 0.590 for $Corr[\xi_x, \xi_m] = 0.3$ 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 across 2,000 replications to evaluate the accuracy and precision of the interaction effect estimates (γ_{xm}) of the three methods. To facilitate the interpretation of path coefficients, we obtained and evaluated standardized estimates of γ_x, γ_m and γ_{xm}.

Raw Bias and Standardized Bias. Standardized bias (SB) was used to evaluate the 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 raw Bias (B):

$$SB = \frac{B(\gamma_{xm})}{SE_{\gamma_{xm}}}, \tag{14}$$

$$B(\gamma_{xm}) = R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm}), \tag{15}$$

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 $\hat{\gamma}_{xm}$ from the population parameter, 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 ≤ 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:

$$Relative\ SE\ Bias = \frac{R^{-1}\Sigma_{r=1}^{R}(\widehat{SE}_r - SD)}{SD}, \tag{16}$$

where $\widehat{SE}_r$ was the estimated standard error of $\hat{\gamma}_{xm}$ in a single replication cycle r and SD was 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). 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 the study design, a robust version of relative SE bias was calculated as an alternative to the regular one:

$$Robust\ Relative\ SE\ Bias = \frac{MDN(\widehat{SE}_r) - MAD}{MAD}, \tag{17}$$

where MDN represented the median value of the estimated SE values and MAD was the empirical median-absolute-deviation of SE values. The MAD was defined by the median of absolute deviations from the median of sample, such that

$MAD = b * MDN(|\widehat{SE}_r - MDN(SE)|)$ where b is a scale factor set to 1.4826 to match the standard deviation of a normal distribution so that MAD can be a consistent estimator for SD (Huber, 2011; Rousseeuw & Croux, 1993). 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 substitute 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; Dekking et al., 2005) method was included as a supplemental information of SE estimates:

$$O_a \notin (Q_1 - 1.5 \times IQR,\ Q_3 + 1.5 \times IQR), \tag{18}$$

where O_a was an observation of outlier for a = 1, 2, . . . , 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 (Dekking et al., 2005).

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:

$$RMSE = \sqrt{R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm})^2}. \tag{19}$$

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.

Empirical Type I Error Rate and Statistical Power. The empirical type I error informed the probability of incorrectly rejecting the null hypothesis that the latent interaction effect is not significant (i.e., $H_0 : \gamma_{xm} = 0$) at the significance level ($\alpha = .05$). The type I error rate was computed across 2,000 replications by determining the proportion of instances where a Type I error occurred. Statistical power measures a method's ability to detect an effect if one truly exists. In this study, it was defined as the proportion of correctly rejecting the null hypothesis when the interaction effect truly exists (i.e., $H_a : \gamma_{xm} = 0.3$).

Results

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

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 B and SB associated with the estimate of γ_{xm} using the three methods consistently remained within 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., $\rho = 0.7$) and showed consistently decreasing magnitude of SB and B as ρ increased from 0.7 to 0.9, across conditions of sample size and $Corr(\xi_x, \xi_m)$. For instance, holding other conditions constant, the magnitude of SB(B) decreased from .14(.08) to .08(.01) for RAPI, .13(.05) to .04(0) for matched pair UPI, and .14(.04) to .02(0) for 2S-PA-Int, implying that the estimation of γ_{xm} gradually became more accurate as the amount of measurement error within first-order indicators reduced.

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 reduced sampling errors. Nevertheless this pattern was not exactly consistent with the absolute SB because the empirical standard error 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 B 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 error, and the comparability of B 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 under high reliability. 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 under high reliability, indicating that the absolute SB values were more randomly distributed in that condition.

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 magnitude of robust relative SE bias were all within the -10% ~ 10% range for RAPI, matched-pair UPI, and 2S-PA-Int across 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 unbiased. Compared to 2S-PA-Int, matched-pair UPI produced two relative SE values outside the acceptable range under 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 appeared in various conditions under low reliability ($\rho = .7$), meaning that RAPI may generate the interaction effect with more uncertainty as measurement error is high. The outlier proportions of SEs identified by the IQR method showed declining trends for all the methods as sample size increased and reliability levels improved, meaning that the estimation of γ_{xm} became more accurate and stable with less extreme values. For instance, holding $Corr(\xi_x, \xi_m) = 0$, the proportion of outliers for RAPI decreased from 10.90% to 5.55% and to 2.60% as N increased for $\rho = .70$, while within the condition of $N = 100$ and the proportion decreased from 10.90% to 5.40% and to 1.90% as ρ.

However, the robust relative SE bias did not demonstrate a clear pattern associated with ρ and N. It was found that even though under the condition of large sample size and high reliability, the robust relative SE bias could be higher then those under the worse conditions. 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 within the methods in terms of sample size, population reliability level, and correlation between first-order latent variables. 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 with 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.

Empirical Type I Error Rate and Statistical Power

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 (2022) to support latent interaction estimation, namely 2S-PA-Int. The major difference between matched-pair UPI, RAPI, and 2S-PA-Int is on the formation of the latent interaction term. Specifically, matched-pair UPI forms the latent interaction term by using multiple PIs 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 SIs to the latent interaction term, respectively. Our findings indicated that all the three methods were capable of generating unbiased estimates of interaction effects by accounting for measurement errors, with the magnitude of SB and B 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 in our study, which implied that it is not as robust as RAPI and 2S-PA-Int and not recommended to use under this condition. Overall, the results of coverage rate also implied that RAPI and 2S-PA-Int have the potential of adequate statistical power to accurately identify the presence and magnitude of interaction effects across conditions. Our results were consistent 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 reliability level significantly impacted the estimation of non-zero interaction effects. The absolute values of SB and B were sensitive to low sample size and high amount of measurement error reflected by the estimated reliability, and they generally became smaller with increased sample size and decreased error for all the methods. It implies 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 level of sample size, higher reliability levels of first-order indicators generally result in more unbiased estimated for all the methods in most cases. The relative SE biases showed similar patterns, such that the magnitude of relative SE bias became smaller with increasing sample size and reliability level. Particularly RAPI generally exhibited larger relative SE biases than matched-pair UPI and 2S-PA-Int especially under small sample size and low reliability level, indicating that RAPI is more inclined to generate unstable interaction estimates under such conditions. Overall, although the 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 apparently affected by both sample size and reliability level for all the methods. As sample size and reliability individually or jointly increase, the RMSE values demonstrated declining trends in magnitude, meaning that the interaction estimates showed more accuracy and variability. Despite the consistent trends, the 2S-PA-Int method produced estimates with less RMSE than RAPI and matched-pair UPI particularly under small sample size and low reliability level, implying that 2S-PA-Int can generate more accurate and stable interaction effects under these conditions. Taking all the evaluation criteria into account, 2S-PA-Int showed ample potential to serve as a good alternative to RAPI and matched-UPI for latent interaction estimation by demonstrating robustness to extreme conditions.

Revisiting Marsh's criteria of a good latent interaction model, 2S-PA-Int is practically preferable in terms of simple model specification as a single-indicator method, 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 Directions

Marsh et al. (2004) demonstrated that all-pair UPI was not preferred because it did not have substantial improvement on latent interaction estimation but has slightly more complicated model compared to matched-pair UPI. We explored the performance of all-pair UPI on congeneric items in our preliminary study, as Marsh et al. (2004) only explored parallel items without varied conditions of factor loadings and error variances. We found that the standardized biases of interaction estimates produced by all-pair UPI were negligible for parallel items but not for congeneric factor items (i.e., items with varied factor loadings but same error variances) and congeneric items. A reasonable explanation of increasing standard biases is that raw biases were very small for three types of items, while the corresponding standard errors became systematically smaller with sample size increased. Thus, the results implied that all-pair UPI did not have as good performance as matched-pair UPI and hence we did not included it in our main study (Results of preliminary simulation study of all-pair UPI are available at: https://github.com/Gengrui-Zhang/2S-PA-Int/blob/main/Qual_1_Supplemental_Material/Supplemental-Material.pdf).

Regarding the study design, we have not included the condition when the effect of $\xi_x\xi_m$ is 0, which means that the evaluation of Type I error rate and power cannot be conducted. We will include this condition as a factor of study design in the future. Besides, some alternative methods using distribution analytic approach, such as the latent moderated structural equation (LMS; Klein & Moosbrugger, 2000) method, were not evaluated in this study, given that the focus of this study was on product indicator methods. Past literature has shown that LMS tended to produce unbiased estimates of the latent interaction effect with acceptable statistical power on congeneric items with normal distribution (Hsiao, et al., 2021; Cham et al., 2012). Therefore, in future investigations, we plan to incorporate widely used alternative methods for comparison with the 2S-PA-Int approach. Additionally, 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 is increasingly used in educational, counseling, and organizational research (e.g., students nested in classrooms, patients nested in clinics, employees nested in companies), and it is worth exploring the potential of 2S-PA-Int with complicated data types and structures with varied sample sizes and reliability levels.

As for other limitations, 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). When factor scores with standard errors are not available in some secondary dataset, researchers may not be able to compute factors scores and thus apply 2S-PA-Int 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. Nevertheless, 2S-PA-Int is expected to show better performance on estimating latent interaction effects since theoretically it can incorporate differential standard error of measurement for each observation.

References

Algina, J., & Moulder, B. C. (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction using LISREL 8.3. *Structural Equation Modeling*, *8*(1), 40–52. https://doi.org/10.1207/S15328007SEM0801_3

Anderson, L. E., Stone-Romero, E. F., & Tisak, J. (1996). A Comparison of bias and mean squared error in parameter estimates of interaction effects: Moderated multiple regression versus error-in-variables regression. *Multivariate Behavioral Research*, *31*(1), 69–94. https://doi.org/10.1207/s15327906mbr3101_5

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

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*(1), 605–634. https://doi.org/10.1146/annurev.psych.53.100901.135239

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models* (p. 320). Sage Publications, Inc.

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

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge. https://doi.org/10.4324/9781315757421

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

Carte, T. A., & Russell, C. J. (2003). In pursuit of moderation: Nine common errors and their solutions. *MIS Quarterly*, *27*(3), 479–501. https://doi.org/10.2307/30036541

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). 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

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research*, *14*(2), 189–217. https://doi.org/10.1287/isre.14.2.189.16018

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences, 3rd ed* (pp. xxviii, 703). Lawrence Erlbaum Associates Publishers.

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.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Cunningham, G. B., & Ahn, N. Y. (2019). Moderation in sport management research: Room for growth. *Measurement in Physical Education and Exercise Science*, *23*(4), 301–313. https://doi.org/10.1080/1091367X.2018.1472095

Daszykowski, M., Kaczmarek, K., Vander Heyden, Y., & Walczak, B. (2007). Robust statistics in data analysis — A review. *Chemometrics and Intelligent Laboratory Systems*, *85*(2), 203–219. https://doi.org/10.1016/j.chemolab.2006.06.016

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics*. Springer. https://doi.org/10.1007/1-84628-168-7

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

Estabrook, R., & Neale, M. (2013). A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to nicotine dependence.

Multivariate Behav Res, *48*(1), 1–27. https://doi.org/10.1080/00273171.2012.730072

Foldnes, N., & Hagtvet, K. A. (2014). The choice of product indicators in latent variable interaction models: Post hoc analyses. *Psychological Methods*, *19*(3), 444–457. https://doi.org/10.1037/a0035728

Hancock, G. R., & Mueller, R. O. (2011). The reliability aradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement*, *71*(2), 306–324. https://doi.org/10.1177/0013164410384856

Harwell, M. (2019). A strategy for using bias and RMSE as outcomes in Monte Carlo studies in statistics. *J. Mod. Appl. Stat. Methods*, *17*(2), jmasm.eP2938. https://doi.org/10.22237/jmasm/1551907966

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

Huber, P. J. (2011). Robust Statistics. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1248–1251). Springer. https://doi.org/10.1007/978-3-642-04898-2_594

Jackman, G.-A., Leite, W., & Cochrane, D. (2011). Estimating latent variable interactions with the unconstrained approach: A comparison of methods to form product indicators for large, unequal numbers of items. *Structural Equation Modeling*, *18*, 274–288. https://doi.org/10.1080/10705511.2011.557342

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*.

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., & 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.

Kyriazos, T. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *09*, 2207–2230. https://doi.org/10.4236/psych.2018.98126

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

Lai, M. H. C., Tse, W. W.-Y., Zhang, G., Li, Y., & Hsiao, Y.-Y. (2023). Correcting for unreliability and partial invariance: A two-stage path analysis approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(2), 258–271. https://doi.org/10.1080/10705511.2022.2125397

Lin, G.-C., Wen, Z., Marsh, H. W., & Lin, H.-S. (2010). Structural equation models of latent interactions: Clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(3), 374–391. https://doi.org/10.1080/10705511.2010.488999

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

MacKinnon, D. P., & Luecken, L. J. (2008). How and for whom? Mediation and moderation in health psychology. *Health Psychol*, *27*(2S), S99–S100. https://doi.org/10.1037/0278-6133.27.2(Suppl.).S99

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

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x

Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, *9*(1), 1–19. https://doi.org/10.1207/S15328007SEM0901_1

Mueller, R. O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling*, *4*(4), 353–369. https://doi.org/10.1080/10705519709540081

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

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

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283. https://doi.org/10.2307/2291267

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *WIREs*

Data Mining and Knowledge Discovery, *1*(1), 73–79. https://doi.org/10.1002/widm.2

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

Steinmetz, H., Davidov, E., & Schmidt, P. (2011). Three approaches to estimate latent interaction effects: Intention and perceived behavioral control in the theory of planned behavior. *Methodological Innovations Online*, *6*(1), 95–110. https://doi.org/10.4256/mio.2010.0030

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613–625. https://doi.org/10.1007/BF02289858

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

Table 1

Standardized Bias (Raw Bias) for $\gamma_{xm}(= 0.3)$ over 2,000 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	0.16 (0.08)	0.18 (0.03)	0.04 (0)	0.09 (0.03)	0.11 (0.02)	0 (0)	0.11 (0.03)	0.05 (0.01)	-0.02 (0)
	0.3	0.08 (0.06)	0.18 (0.03)	0.01 (0)	0.12 (0.04)	0.12 (0.02)	-0.01 (0)	0.11 (0.03)	0.06 (0.01)	-0.04 (0)
	0.6	0.19 (0.06)	0.19 (0.02)	0.06 (0.01)	0.12 (0.03)	0.14 (0.02)	0 (0)	0.16 (0.02)	0.1 (0.01)	0 (0)
250	0	0.21 (0.02)	0.11 (0.01)	0.09 (0.01)	0.06 (0.01)	0.06 (0.01)	0.06 (0)	0.08 (0.01)	0.02 (0)	0.03 (0)
	0.3	0.24 (0.03)	0.11 (0.01)	0.07 (0)	0.11 (0.01)	0.08 (0.01)	0.03 (0)	0.12 (0.01)	0.03 (0)	0.02 (0)
	0.6	0.21 (0.02)	0.13 (0.01)	0.05 (0)	0.14 (0.02)	0.11 (0.01)	0.02 (0)	0.13 (0.01)	0.07 (0)	0.02 (0)
500	0	0.15 (0.01)	0.06 (0)	0.06 (0)	0.08 (0.01)	0.03 (0)	0.02 (0)	0.06 (0)	-0.02 (0)	0 (0)
	0.3	0.2 (0.01)	0.1 (0.01)	0.04 (0)	0.13 (0.01)	0.06 (0)	0 (0)	0.1 (0.01)	0.03 (0)	-0.02 (0)
	0.6	0.14 (0.01)	0.12 (0.01)	0.02 (0)	0.09 (0.01)	0.06 (0)	-0.01 (0)	0.07 (0)	0.08 (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. Raw biases 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 2,000 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	**14.22 (10.55)**	4.92 (5.60)	1.69 (3.30)	**-12.77 (8.60)**	-8.22 (5.15)	-2.67 (3.05)	-1.93 (8.25)	-6.07 (3.85)	-3.27 (1.75)
	0.3	9.44 (10.75)	5.54 (4.80)	2.96 (2.20)	**-11.41 (7.10)**	**-10.29 (4.85)**	-5.36 (3.40)	-6.28 (7.50)	-4.49 (3.00)	-4.18 (1.90)
	0.6	**15.19 (8.05)**	5.08 (3.90)	0.53 (1.45)	-7.37 (6.70)	-7.29 (3.90)	-7.17 (2.30)	-3.03 (5.65)	-4.24 (3.25)	-4.25 (1.10)
250	0	7.42 (5.55)	**11.93 (2.50)**	2.43 (1.65)	-4.95 (3.75)	-2.15 (4.20)	-2.84 (1.95)	-5.61 (3.25)	2.15 (1.65)	-1.85 (1.35)
	0.3	5.74 (5.45)	6.36 (2.80)	6.47 (1.30)	-6.2 (3.65)	-5.98 (3.10)	-1.2 (1.75)	**-10.24 (4.10)**	-4.16 (2.00)	1.84 (1.20)
	0.6	8.4 (4.55)	9.4 (2.50)	4.9 (0.95)	-5.88 (4.60)	1.36 (3.40)	-2.09 (1.65)	-8.3 (3.55)	-1.68 (1.70)	-1.82 (0.55)
500	0	**12.46 (3.85)**	3.87 (1.70)	-0.89 (1.15)	-0.65 (3.15)	-3.85 (2.50)	-4.06 (1.55)	-1.54 (3.60)	-5.68 (1.40)	-4.99 (0.85)
	0.3	**13.39 (3.10)**	7.97 (2.10)	5.3 (0.85)	3.06 (3.40)	-7.51 (3.00)	-0.19 (0.65)	-2.26 (2.05)	-4.14 (1.75)	0.23 (0.65)
	0.6	5.34 (2.85)	**10.14 (1.85)**	3.34 (0.60)	-3.61 (3.20)	0.19 (2.00)	0 (0.70)	**-11.12 (2.30)**	-0.87 (1.50)	-2.61 (0.70)

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 2,000 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	97.05	97.45	96.05	**87.05**	92.20	93.90	94.85	95.20	94.70
	0.3	96.90	96.80	95.55	**88.8**	92.35	94.40	95.30	94.95	93.85
	0.6	97.20	96.90	96.25	91.05	92.70	94.55	95.70	94.60	94.10
250	0	96.65	97.05	96.40	91.55	93.50	95.20	93.45	94.20	95.50
	0.3	97.40	97.15	95.55	92.35	93.80	94.40	94.45	94.50	94.45
	0.6	96.80	97.15	95.15	93.2	95.20	93.70	92.75	95.75	94.05
500	0	96.95	96.20	96.25	93.65	94.05	95.00	93.80	94.40	94.75
	0.3	97.60	96.40	96.20	94.45	93.70	95.25	94.25	94.50	95.00
	0.6	96.85	97.15	95.90	92.65	94.95	94.70	92.45	95.05	94.85

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 2,000 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	0.52	0.40	0.25	0.16	0.20	0.14	0.11	0.12	0.11
	0.3	0.74	0.31	0.26	0.14	0.19	0.13	0.11	0.11	0.10
	0.6	0.30	0.24	0.16	0.12	0.15	0.11	0.09	0.10	0.09
250	0	0.12	0.14	0.11	0.08	0.10	0.08	0.07	0.07	0.07
	0.3	0.11	0.13	0.10	0.08	0.10	0.08	0.06	0.07	0.06
	0.6	0.09	0.12	0.09	0.07	0.08	0.06	0.06	0.06	0.06
500	0	0.08	0.10	0.07	0.06	0.07	0.06	0.05	0.05	0.05
	0.3	0.07	0.09	0.07	0.05	0.07	0.05	0.04	0.05	0.04
	0.6	0.06	0.08	0.06	0.05	0.05	0.05	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.

Table 5

Empirical Type I Error Rate for $\gamma_{xm}(=0)$ *over 2,000 Replications.*

		$\rho = .70$			$\rho = .80$			$\rho = .90$		
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	0.01	0.03	0.04	0.00	0.03	0.05	0.03	0.06	0.05
	0.3	0.01	0.03	0.04	0.01	0.03	0.04	0.04	0.05	0.05
	0.6	0.01	0.03	0.04	0.01	0.03	0.04	0.04	0.06	0.05
250	0	0.02	0.05	0.05	0.02	0.05	0.05	0.04	0.06	0.05
	0.3	0.03	0.04	0.05	0.02	0.04	0.05	0.04	0.05	0.05
	0.6	0.03	0.05	0.05	0.03	0.05	0.05	0.05	0.05	0.05
500	0	0.04	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05
	0.3	0.04	0.04	0.05	0.04	0.04	0.05	0.05	0.05	0.05
	0.6	0.04	0.05	0.05	0.04	0.05	0.05	0.05	0.06	0.05

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 the empirical Type I error rate under the same conditions.

Table 6

Statistical Power for $\gamma_{xm}(= 0.3)$ *over 2,000 Replications.*

		$\rho = .70$			$\rho = .80$			$\rho = .90$		
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	0.21	0.56	0.80	0.14	0.49	0.80	0.45	0.66	0.82
	0.3	0.28	0.66	0.84	0.19	0.60	0.82	0.53	0.75	0.85
	0.6	0.43	0.79	0.92	0.35	0.75	0.92	0.70	0.86	0.93
250	0	0.83	0.97	0.99	0.73	0.95	0.99	0.90	0.98	0.99
	0.3	0.90	0.99	1.00	0.81	0.97	1.00	0.95	0.99	1.00
	0.6	0.97	1.00	1.00	0.93	1.00	1.00	0.98	1.00	1.00
500	0	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
	0.3	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	0.6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

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 statistical power under the same conditions.