Two-Stage Path Analysis with Interaction: the Method of Processing Categorical Data

Gengrui (Jimmy) Zhang

Two-Stage Path Analysis with Interaction: the Method of Processing Categorical Data

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

Interaction effects, wherein the effect of predictors depends on a third variable, are

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

Unfortunately, researchers often encounter non-normal ordered categorical data in

psychology research, which poses unique challenges for latent interaction analysis. Ordered categorical data often arise from survey responses, behavioral classifications, or diagnostic categories, where the data can take on a limited number of categories but do not follow a normal distribution (Agresti, 2013). This deviation from normality is especially common in assessments and questionnaires, where Likert scales and similar ordinal measures are used to capture attitudes, symptoms, or behaviors (Liddell & Kruschke, 2018). For example, an impactful study examining responses of the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) within a Japanese national survey highlighted a common pattern among depressive symptom items, such that a distinctive intersection point was found between "rarely" and "some of the time" categories (Tomitaka et al., 2018). It demonstrated that the item responses of the CES-D scale were mostly heavily right-skewed with most responses at the left extreme category. In the field of psychology research, numerous scales with ordered categorical items, including but not limited to the CES-D, are popularly utilized. It is therefore important to explore reliable methods for modeling latent interaction effects within highly non-normal categorical data for observed first-order indicators.

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

Below, I will introduce the three methods with technical details, and briefly discuss

84   tentative procedures of conducting a simulation study.

85   **A General Model of Latent Interaction**

86   Kenny and Judd (1984) laid the groundwork for a structural model designed to

87   estimate latent interaction effects, including two first-order latent predictors and their

88   interaction term:

$$y = \alpha + \gamma_x \xi_x + \gamma_m \xi_m + \gamma_{xm} \xi_x \xi_m + \zeta. \tag{1}$$

89   In the model, α is the intercept, ξ_x and ξ_m are the first-order latent predictors with

90   their product, $\xi_x \xi_m$, defining the interaction effect. Notably, ξ_x and ξ_m can correlate. The

91   disturbance term, ζ, follows a normal distribution $\zeta \sim N(0, \psi)$, where ψ is its variance

92   reflecting unexplained variance in the dependent variable. The coefficients γ_x, γ_m, and γ_{xm}

93   quantify the effects of the predictors and their interaction on the dependent variable, which

94   could be either observed or latent.

95   The measurement model for the first-order latent predictors, such as ξ_x, is framed

96   within the confirmatory factor analysis (CFA) approach:

$$\mathbf{x} = \boldsymbol{\tau_x} + \boldsymbol{\lambda_x} \xi_x + \boldsymbol{\delta_x}. \tag{2}$$

97   For each indicator i ranging from 1 to p associated with ξ_x, $\mathbf{x}$ represents a vector of

98   observed indicators of dimensions $p \times 1$. The vector $\boldsymbol{\tau_x}$ consists of constant intercepts, $\boldsymbol{\lambda_x}$

99   encompasses factor loadings, and $\boldsymbol{\delta_x}$ includes measurement errors, all of dimension $p \times 1$.

100   Each error δ_{x_i} follows a normal distribution with a mean of zero and a variance θ_{x_i}. Under

101   the assumption of local independence, the error variance-covariance matrix, $\boldsymbol{\Theta}_{\delta_{\mathbf{x}}}$, is diagonal,

102   containing variances $\theta_{x_1}, \theta_{x_2}, ..., \theta_{x_p}$. The same measurement framework is applicable to ξ_m.

103   **Unconstrained Product Indicator**

104   The unconstrained product indicator (UPI) method operates by constructing product

105   indicators from observed first-order indicators that relate to their corresponding latent

106 constructs. Unlike the constrained approaches that impose nonlinear constraints on

107 parameters related to the interaction term (Jöreskog & Yang, 1996), the UPI method does

108 not require those constraints and thus simplify the model specification. UPI facilitates the

109 estimation of interaction effects between latent variables directly within any commercial

110 SEM software (e.g., the R package `lavaan`), enhancing accessibility for applied researchers.

111   The structural model of UPI is the same as equation (1). Suppose ξ_x is indicated by

112 three items (i.e., $x_1 \sim x_3$) and ξ_m is indicated by six items (i.e., $m_1 \sim m_6$), the measurement

113 models are presented as:

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \tau_{x_1} \\ \tau_{x_2} \\ \tau_{x_3} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1} \\ \lambda_{x_2} \\ \lambda_{x_3} \end{bmatrix} \begin{bmatrix} \xi_x \end{bmatrix} + \begin{bmatrix} \delta_{x_1} \\ \delta_{x_2} \\ \delta_{x_3} \end{bmatrix},
\tag{3}
$$

$$
\begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix} = \begin{bmatrix} \tau_{m_1} \\ \tau_{m_2} \\ \tau_{m_3} \\ \tau_{m_4} \\ \tau_{m_5} \\ \tau_{m_6} \end{bmatrix} + \begin{bmatrix} \lambda_{m_1} \\ \lambda_{m_2} \\ \lambda_{m_3} \\ \lambda_{m_4} \\ \lambda_{m_5} \\ \lambda_{m_6} \end{bmatrix} \begin{bmatrix} \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{m_1} \\ \delta_{m_2} \\ \delta_{m_3} \\ \delta_{m_4} \\ \delta_{m_5} \\ \delta_{m_6} \end{bmatrix},
\tag{4}
$$

114 where τs are the intercepts, λs are the factor loadings, and δs are the error terms. In this

115 study, we will explore the performance of two UPI models with the matching (matched-pair

116 UPI) and parceling (parceling UPI) strategies.

Matched-pair UPI

118   The matching strategy involves selecting indicators from the first-order indicators of

119 ξ_x and ξ_m, and pairing them up to form PIs. If the numbers of first-order indicators of ξ_x

120 and ξ_m are equal, all the indicators will be used; Otherwise, Wu et al. (2013) recommended

121 to select first-order indicators in the order of reliability from the latent variable with more

122   indicators and discard the rest of indicators. For example, suppose that the sequences of $x_1 \sim$

123   x_3 and $m_1 \sim m_6$ are sorted decreasingly by their reliability (i.e., x_1 has the highest reliability

124   and x_3 has the lowest for ξ_x; m_1 has the highest reliability and m_6 has the lowest for ξ_m),

125   the formed PIs for the interaction term will be x_1m_1, x_2m_2, x_3m_3. The items $m_4 \sim m_6$

126   should be discarded due to low reliability. The reliability of any single indicator (e.g., m_1) is:

$$\frac{\lambda_{m_1}^2 \sigma_{\xi_m}^2}{\lambda_{m_1}^2 \sigma_{\xi_m}^2 + \theta_{m_1}}, \tag{5}$$

127   where $\sigma_{\xi_m}^2$ is the variance of ξ_m and θ_{m_1} is the error variance of m_1 (Aytürk et al., 2020).

128   Then, the model for matched-pair UPI will be represented as:

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

129   Due to the nature of discrete data for first-order indicators, the indicators do not

130   follow multivariate normal assumptions and their error variances are not homogeneous.

131   However, UPI is robust to violations of multivariate normal distributions of the first-

132   order latent variables, the disturbance term, and the measurement errors of first-order indicators,

133   since these parameters are freely estimated in UPI. The mean of the latent interaction term is

134   equal to the covariance between the first-order latent variables (i.e., $E[\xi_x\xi_m] = Corr[\xi_x, \xi_m]$

135   where $Corr[\cdot]$ is the correlation term; Jöreskog & Yang, 1996). Although UPI is theoretically

136   less statistically powerful than CPI in the condition of non-normal first-order indicators,

137   Marsh et al. (2004) showed that the performance of UPI on simulated non-normal data was

138   acceptable with unbiased estimates of interaction effects and reasonable standard errors.

139   ***Parceling UPI***

140   According to the original idea of Wu et al. (2013), another strategy of dealing with

141   unequal numbers of first-order indicators is to take average of two or more indicators from

the same source latent construct and form parcels. The number of parcels will be equal to

the number of indicators from the latent variable with fewer indicators. Parcels can be

formed based on considerations such as thematic consistency among indicators, item-to-total

correlations, or factor loadings. In this study, we follow the factorial algorithm (Rogers &

Schmitt, 2004) in the study design of Aytürk et al. (2020) by taking the average of the

indicators with the highest and lowest reliability until all indicators have been used for

parceling. In the example of $m_1 \sim m_6$, three parcels will be created to match three items of

ξ_x and the measurement model will be represented as:

$$
\begin{bmatrix} p_{m_1 m_6} \\ p_{m_2 m_5} \\ p_{m_3 m_4} \end{bmatrix} = \begin{bmatrix} \tau_{p_{m_1 m_6}} \\ \tau_{p_{m_2 m_5}} \\ \tau_{p_{m_3 m_4}} \end{bmatrix} + \begin{bmatrix} \lambda_{p_{m_1 m_6}} \\ \lambda_{p_{m_2 m_5}} \\ \lambda_{p_{m_3 m_4}} \end{bmatrix} \begin{bmatrix} \xi_x \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{p_{m_1 m_6}} \\ \delta_{p_{m_2 m_5}} \\ \delta_{p_{m_3 m_4}} \end{bmatrix}, \tag{7}
$$

where $p_{m_1 m_6}$, $p_{m_2 m_5}$, $p_{m_3 m_4}$ are three formed parcels according to the factorial algorithm, in

which m_1 has the highest reliability and m_6 has the lowest reliability, m_2 has the second

highest reliability and m_5 has the second lowest reliability, and m_3 has the third highest

reliability and m_4 has the third lowest reliability. Then, the parcels are matched to the

indicators of another first-order latent variable according to the order of reliability. For

instance, $p_{m_1 m_6}$ will be matched to x_1, and so on. The model is represented as:

$$
\begin{bmatrix} x_1 p_{m_1 m_6} \\ x_2 p_{m_2 m_5} \\ x_3 p_{m_3 m_4} \end{bmatrix} = \begin{bmatrix} \tau_{x_1 p_{m_1 m_6}} \\ \tau_{x_2 p_{m_2 m_5}} \\ \tau_{x_3 p_{m_3 m_4}} \end{bmatrix} + \begin{bmatrix} \lambda_{x_1 p_{m_1 m_6}} \\ \lambda_{x_2 p_{m_2 m_5}} \\ \lambda_{x_3 p_{m_3 m_4}} \end{bmatrix} \begin{bmatrix} \xi_x \xi_m \end{bmatrix} + \begin{bmatrix} \delta_{x_1 p_{m_1 m_6}} \\ \delta_{x_2 p_{m_2 m_5}} \\ \delta_{x_3 p_{m_3 m_4}} \end{bmatrix}, \tag{8}
$$

For UPI with matching and parceling strategies, the first-order indicators will be

simulated as continuous variables and standardized. Additionally, they will be mean-centered

before forming PIs to enhance model convergence rates and reduce model complexity (Marsh

et al., 2004). Alternatively speaking, the formed PIs are treated and estimated as continuous

variables, which are liberal to multivariate normal distributions (Cham et al., 2012)

Latent Moderated Structural Equation

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

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

$$\xi = \mathbf{A}\mathbf{z}, \tag{9}$$

where

$$\xi = \begin{pmatrix} \xi_x \\ \xi_m \end{pmatrix}, \tag{10}$$

$$\mathbf{A} = \begin{pmatrix} a_{xx} & 0 \\ a_{xm} & a_{mm} \end{pmatrix}, \tag{11}$$

and

$$\mathbf{z} = \begin{pmatrix} z_x \\ z_m \end{pmatrix}. \tag{12}$$

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

$$y = \alpha + \mathbf{\Gamma}\xi + \xi^T \mathbf{\Omega} \xi + \zeta, \tag{13}$$

179 where $\mathbf{\Gamma}$ is a $(1 \times p)$ vector of first-order path coefficients of ξ, and $\mathbf{\Omega}$ is a $(p \times p)$

180 upper-diagnol matrix of the interaction effects' coefficients. In the example of $\xi = (\xi_x, \ \xi_m)^T$

181 (T represents transpose), the expanded structural equation can be shown as:

$$y = \alpha + (\gamma_x \ \gamma_m) \begin{pmatrix} \xi_x \\ \xi_m \end{pmatrix} + (\xi_x \ \xi_m) \begin{pmatrix} 0 & \gamma_{xm} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_x \\ \xi_m \end{pmatrix} + \zeta. \tag{14}$$

182 By substituting the ξ vector into the structural model of LMS, the structrual model

183 can be further expanded to:

$$\eta = \gamma_0 + \mathbf{\Gamma A z} + \mathbf{z}^T \mathbf{A}^T \mathbf{\Omega A z} + \zeta$$

$$= \gamma_0 + \mathbf{\Gamma A} \begin{pmatrix} z_1 \\ 0 \end{pmatrix} + \mathbf{\Gamma A} \begin{pmatrix} 0 \\ z_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ 0 \end{pmatrix}^T \mathbf{A}^T \mathbf{\Omega A} \begin{pmatrix} z_1 \\ 0 \end{pmatrix} + \begin{pmatrix} z_1 \\ 0 \end{pmatrix}^T \mathbf{A}^T \mathbf{\Omega A} \begin{pmatrix} 0 \\ z_2 \end{pmatrix} + \zeta$$

$$= \left(\gamma_0 + \mathbf{\Gamma A} \begin{pmatrix} z_1 \\ 0 \end{pmatrix} + \begin{pmatrix} z_1 \\ 0 \end{pmatrix}^T \mathbf{A}^T \mathbf{\Omega A} \begin{pmatrix} z_1 \\ 0 \end{pmatrix} \right) + \left(\mathbf{\Gamma A} + \begin{pmatrix} z_1 \\ 0 \end{pmatrix}^T \mathbf{A}^T \mathbf{\Omega A} \right) \begin{pmatrix} 0 \\ z_2 \end{pmatrix} + \zeta, \tag{15}$$

184 wherein the interaction effect demonstrates that the dependent measure η is linearly related

185 to z_2 but non-linearly related to z_1, and hence η is not normally distributed. LMS then

186 estimates the model by finding the maximum likelihood solution of the mixture distribution

187 of the first-order indicators and η using the EM algorithm with numerical integration (Klein

188 & Moosbrugger, 2000).

189 The assumptions of LMS in Klein and Moosbrugger's work include: (1) the first-order

190 indicators are multivariate normal; (2) The errors of first-order indicators are normally and

191 independently distributed with means of 0 (which is violated for order categorical items); (3)

192 The disturbance term (i.e., ζ) in the structural equation is normally distributed with a mean

193 of 0, and independent of latent variables and The errors of first-order indicators. Since LMS

194 is estimated using maximum likelihood method that is based on normal distribution, it has

195 been reported that LMS could generate biased estimates of the interaction effect when the

196 first-order items did not follow normal distributions, for either continuous or ordered

categorical indicators (Marsh et al., 2004; Cham et al., 2012; Maslowsky et al., 2015; Cheung & Lau, 2017; Aytürk et al., 2020; Cheung et al., 2021). Aytürk et al. (2021) have shown that LMS with categorical variables in `Mplus` is able to generate unbiased estimates for both first-order and interaction effects across sample size, interaction effect size, missing data conditions, and numbers of item response category, but could produce highly biased measurement parameters such as factor loadings and item category thresholds.

Two-Stage Path Analysis with Interaction

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

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

$$
\begin{bmatrix} \tilde{x}_j \\ \tilde{m}_j \\ \widetilde{xm}_j \end{bmatrix} = \begin{bmatrix} \tau_{\tilde{x}_j} \\ \tau_{\tilde{m}_j} \\ \tau_{\widetilde{xm}_j} \end{bmatrix} + \begin{bmatrix} \lambda_{\tilde{x}_j} & 0 & 0 \\ 0 & \lambda_{\tilde{m}_j} & 0 \\ 0 & 0 & \lambda_{\widetilde{xm}_j} \end{bmatrix} \begin{bmatrix} \xi_{x_j} \\ \xi_{m_j} \\ \xi_{x_j}\xi_{m_j} \end{bmatrix} + \begin{bmatrix} \delta_{\tilde{x}_j} \\ \delta_{\tilde{m}_j} \\ \delta_{\widetilde{xm}_j} \end{bmatrix} , \tag{16}
$$

where for each observation j from $j = 1, 2, ..., n$, $\tilde{x}_j$, $\tilde{m}_j$ are factor scores that are SIs of ξ_x and ξ_m. The PI $\widetilde{xm}_j$ is the SI of the latent interaction term created by multiplying the SIs of

ξ_x and ξ_m. The structural model of 2S-PA-Int is the same as equation (1).

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

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

246 error of measurement.

247   Since latent variables do not have meaningful units in nature, it is suggested to scale

248 the variance of latent variables to unit when fitting the measurement model (i.e., $\sigma^2_{\xi_x} = 1$.

249 In the example that factor scores using the EAP method from item response models, let $\hat{\sigma}_{\tilde{x}_j}$

250 be the estimates standard error of $\tilde{x}_j$ for the individual j. The estimated variance of the true

251 score, also the estimated observation-specific reliability, can be computed as

252 $\hat{\rho}_{\tilde{x}_j} = \sigma^2_{\xi_x} - \hat{\sigma}^2_{\tilde{x}_j} = 1 - \hat{\sigma}^2_{\tilde{x}_j}$, where $\hat{\rho}_{\tilde{x}_j}$ represents the reliability. This reliability is used as a

253 constraint on the factor loading of $\tilde{x}_j$ (i.e., $\lambda_{\tilde{x}_j} = \hat{\rho}_{\tilde{x}_j} = 1 - \hat{\sigma}^2_{\tilde{x}_j}$), and the error variance

254 constraint is set as $\theta_{\tilde{x}_j} = \hat{\sigma}^2_{\tilde{x}_j} \hat{\rho}_{\tilde{x}_j} = \hat{\sigma}^2_{\tilde{x}_j}(1 - \hat{\sigma}^2_{\tilde{x}_j})$ accordingly. The parameter constraints for

255 the factor score of ξ_m can be obtained in the similar way. Assuming that the error variance

256 of ξ_m has been calculated as $\theta_{\tilde{m}_j}$, the error variance constraint for the PI of the interaction

257 term can be calculated as the equation (6) ~ (8) in Hsiao et al. (2018):

$$\theta_{\widetilde{xm}_j} = \hat{\rho}_{\tilde{x}_j}\theta_{\tilde{m}_j} + \hat{\rho}_{\tilde{m}_j}\theta_{\tilde{x}_j} + \theta_{\tilde{m}_j}\theta_{\tilde{x}_j}. \tag{17}$$

258 **Method and Simulation Design**

259   Based on the study design of Aytürk et al. (2020) and Hsiao et al. (2021), I intend to

260 compare the performance of UPI with two strategies of forming PIs, LMS for categorical

261 items, and 2S-PA-Int, on estimating latent interaction effects. The observed first-order

262 indicators (e.g., x_{ij}) will be generated from a graded response model (Samejima, 1969) with

263 differential factor loadings:

$$x^*_{ij} = \lambda_{x_i}\xi_{x_j} + \delta_{x_{ij}}, \tag{18}$$

264 where x^*_{ij} is the score of underlying latent continuous variable for each observed categorical

265 item i. $\delta_{x_{ij}}$ is the individual-specific error term for each observed indicator i and follows

266 standard logistic distribution, assuming the item factor model is estimated with a cumulative

267 logit link (Wirth & Edwards, 2007). Given x^*_{ij}, the observed categorical item x_{ij} can be

268 created with multiple categories with ordinal categories:

$$x_{ij} = \begin{cases} 0 & \text{if } x_{ij}^* < \beta_{x_{i1}} \\ k & \text{if } \beta_{x_{ik}} \le x_{ij}^* < \beta_{x_{i(k+1)}} \\ K-1 & \text{if } \beta_{x_{i(K-1)}} \le x_{ij}^* \end{cases}, \tag{19}$$

269 where β_{ik} is the threshold parameter between the kth and $(k+1)$th category for

270 $k = 1, 2, ..., K$. The framework for observed categorical items of ξ_{m_j} is similar to ξ_{x_j}.

271   I will use R 4.3.2 (R Core Team, 2023) to conduct the simulation studies in this

272 project. To simulate the population dataset, the first-order latent variables ξ_{x_j} and ξ_{m_j} will

273 be first simulated with standard normal distributions (i.e., $N[0, 1]$). The population

274 structural model is the same as the equation (1), where the path coefficients are set to fixed

275 values (i.e., $\gamma_x = 0.3$, $\gamma_m = 0.3$ and $\gamma_{xm} = 0.3$). The disturbance term ζ will be simulated

276 with normal distribution and a mean of 0. When the correlation between ξ_{x_j} and ξ_{m_j} is 0,

277 the variance of disturbance (i.e., ψ) will be 0.73. The interaction term will be computed as

278 the product of ξ_{x_j} and ξ_{m_j}, and the exogenous variable y will be simulated with a

279 distribution of $N(0, \psi)$.

280   Next, the observed ordered categorical indicators will be generated according to the

281 equations (18) ~ (19). Three items will be simulated for ξ_{x_j} (i.e., $x_{1j} \sim x_{3j}$) and 12 items for

282 ξ_{m_j} (i.e., $m_{1j} \sim m_{12j}$). Regarding factor loadings, I plan to adapt the design in Aytürk et

283 al. (2020) in which the unstandardized factor loadings are decreasing in magnitude with

284 equally-spaced intervals across indicators. Specifically, the factor loadings for $x_{1j} \sim x_{3j}$ are

285 set to (.60, .70, .80), and those for $m_{1j} \sim m_{12j}$ are set to (.30, .35, .40, .45, .50, .55, .60, .65,

286 .70, .75, .80, .85). The reliability measure for categorical indicators can be computed using

287 Green and Yang's (2009b) alternative reliability estimate, ω_{cat}, for unidimensional

288 categorical items. Flora (2020) presented the way of computing ω_{cat} using the function

289 `reliability` from the R package `semTools` (Jorgensen et al., 2022), and the output of

290 estimate `omega` and `omega2` is based on Green and Yang's (2009b) method. Hence, the scale

291 internal consistency can be obtained using `reliability` function.

292   To better approximate the real substantive study, $x_{1j} \sim x_{3j}$ will be simulated with 2

293 categories (i.e., $K = 2$), which means that they are binary indicators, and $m_{1j} \sim m_{12j}$ will be

294 simulated with 5 categories (i.e., $K = 4$), which means that they are indicators with multiple

295 categories. In the case of symmetric distribution, the thresholds for $x_{1j} \sim x_{3j}$ will be set to 0

296 so that the success probability of indicators is .5, calculated by the R function `plogis`. For

297 $m_{1j} \sim m_{12j}$, ks are simulated for symmetric distribution and skewed distribution. the

298 threshold values will be centered around 0. Specifically, $\beta_{\mathbf{m_1}}$ is a threshold vector of (-1.5, -.5,

299 .5, 1.5) across indicators. For the skewed condition, the thresholds for $x_{1j} \sim x_{3j}$ will be set to

300 -2.2 and the success probability of indicators is .9; the threshold vector for $m_{1j} \sim m_{12j}$ is set

301 to (.05, .75, 1.55, 2.55) across indicators and the distributions of simulated observed

302 indicators are apparently right-skewed.

303   The final dataset should be composed of 15 observed indicators of ξ_{x_j} and ξ_{m_j}, and

304 one exogenous observed variable y_j predicted by ξ_x, ξ_m, and the interaction term $\xi_{x_j}\xi_{m_j}$.

305 Note that for UPI and LMS, the subscript j can be removed because they assume the same

306 likelihood function for each observation.

307   The 2S-PA-Int method will be implemented by user-defined function based on the R

308 package `lavaan` (Rosseel, 2012) and `OpenMx` (Boker, 2023), and estimated using maximum

309 likelihood (`ML`). The factor scores using the EAP method can be calculated using the

310 function `fscores` with the argument `full.scores.SE = TRUE` from the `mirt` package

311 (Chalmers, 2012). The UPI method will be implemented using the function `IndProd` in the

312 R package `semTools`, and the `sem` function in `lavaan`. The UPI models will be estimated

313 using maximum likelihood with robust standard error (`MLR`). Since there is not a reliable R

314 package for LMS, the LMS model will be estimated in `Mplus` (Muthén & Muthén,

315 1998-2024) using the Gauss-Hermite quadrature integration algorithm with 16 integration

316   points. The robust standard errors will be used as well.

317   To investigate Type I error rate of latent interaction effect, a condition of $\gamma_{xm} = 0$

318   will be added in the study design to compare with $\gamma_{xm} = 0.3$. The latent variables ξ_x and ξ_m

319   are assumed to follow standard normal distributions (i.e., $N \sim [0, 1]$) and they are allowed to

320   correlate with three levels of correlation (i.e., $Corr[\xi_x, \xi_m] = 0,\ 0.3,\ 0.6$). Three sample size

321   conditions (i.e., 100, 250, and 500) will be included in the study design to examine the

322   impact of sample size on the estimation of latent interaction effect across methods, according

323   to the study design of Hsiao et al. (2021). Besides, past studies have examined the impact of

324   reliability of observed indicators, I plan to manipulate the reliability of underlying

325   continuous indicators (i.e., x_{ij}^*) with three levels: 0.7, 0.8, and 0.9. According to the equation

326   (5) of composite reliability, and given that the factor loadings are fixed values, I can adjust

327   the error variances of each observed continuous indicator before categorizing them. Suppose

328   the items $x_{1j} \sim x_{3j}$ are manipulated to have a reliability of 0.7, the sum of error variances

329   will be 1.89. To achieve congeneric items, the distribution of error variance proportions will

330   be 44%, 33%, and 23%, and the specific error variances of $x_{1j} \sim x_{3j}$ will be 0.83, 0.62, and

331   0.44. For $m_{1j} \sim m_{12j}$, the distribution of error variance proportions can be 5%, 3%, 12%, 6%,

332   10%, 2%, 15%, 9%, 8%, 11%, 7%, 12%.

333   In summary, the tentative study design will be a $2 \times 2 \times 3 \times 3$ study with two effect

334   sizes of the interaction effect, two distributions of indicators, three sample sizes, and three

335   reliability levels between the latent predictor and the latent moderator. The Monte Carlo

336   Simulation will be structured and conducted using the R package `SimDesign` (Chalmers,

337   2020), and the tentative number of replication will be 2,000 for each condition. Across all the

338   models, the standardized point estimate of the interaction effect (i.e., $\hat{\gamma}_{xm}$) with standard

339   error (i.e., $\hat{SE}[\hat{\gamma}_{xm}]$) will be compared.

Evaluation Criteria

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

Convergence Rate

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

Standardized Bias

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

$$B(\gamma_{xm}) = R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm}), \tag{20}$$

$$SB = \frac{B(\gamma_{xm})}{SE_{\gamma_{xm}}}, \tag{21}$$

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

361 considered acceptable for each replication condition (Collins et al., 2001).

Robust Relative Standard Error (SE) Bias

363   The robust relative SE bias will be computed as:

$$Robust\ Relative\ SE\ Bias = \frac{MDN(\widehat{SE_r}) - MAD}{MAD}, \tag{22}$$

364 where MDN will be the median of the estimated SE values and MAD will be the empirical

365 median-absolute-deviation of SE values. An absolute value of robust relative SE bias within

366 10% range will be considered acceptable (Hooglan & Boomsma, 1998).

Root Mean Squared Erorr (RMSE)

368   The RMSE is defined as the squared root of the sum of squared bias:

$$RMSE = \sqrt{R^{-1}\Sigma_{r=1}^{R}(\hat{\gamma}_{xm_r} - \gamma_{xm})^2}. \tag{23}$$

369 RMSE measures the average difference between calculated interaction estimates and their

370 true value, which can account for both bias, the systematic deviation from the true value,

371 and variability, the spread of estimates across replications. In a 2,000 replication simulation,

372 lower RMSE indicates greater accuracy in estimating $\hat{\gamma}_{xm}$. RMSE provides the most

373 informative comparison across methodologies when key factors, including sample size, model

374 complexity, and disturbance level, are held constant in the simulation.

Empirical Type I Error and Statistical Power

376   The empirical Type I error rate will be computed as the proportion of replications in

377 which the Wald test rejects the true null hypothesis $H_0 : \gamma_{xm} = 0$ at the significance level

378 $\alpha = .05$ for the condition $\hat{\gamma}_{xm} = 0$. The empirical power will be computed similarly for the

379 condition $\gamma_{xm} \neq 0$.