2S-PA-Int-Cat

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

Two-stage path analysis with interaction for categorical variables.

Keywords: keywords

Word count: X

2S-PA-Int-Cat

Methods

We adopted a fully crossed design with varying conditions of sample size, composite reliability of scale, interaction effect, and item skewness, based on the study design of Aytürk et al. (2020) and (**hsiaoModelingMeasurementErrors2021?**). We compared the accuracy of UPI with three product–indicator formations (all-pair, matched-pair, and parceled-pair), LMS for categorical items, and 2S-PA-Int in recovering the latent interaction effect.

Population Structural Model

We considered a latent regression model with two exogenous latent variables for person j with $j = 1, \ldots, N$, ξ_{x_j} and $\xi_{_j}$, and one endogenous latent outcome, $\xi_{_j}$, as our population model. The primary estimand was the standardized latent interaction effect of ξ_{x_j} and ξ_{m_j} on ξ_{y_j}, denoted γ_{xm}:

For $j = 1, \ldots, N$,

$$\begin{bmatrix} \xi_{x_j} \\ \xi_{m_j} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \quad \xi_{y_j} = \alpha + \gamma_x \, \xi_{x_j} + \gamma_m \, \xi_{m_j} + \gamma_{xm} \, \xi_{x_j} \xi_{m_j} + \zeta_j, \tag{1}$$

where α was the constant intercept set to 1.2. ξ_{x_j} was the first-order latent predictor with a fixed main effect $\gamma_x = 0.3$, and ξ_{m_j} was the first-order latent moderator also with a fixed main effect $\gamma_m = 0.3$. Both exogenous factors were factors were standardized with zero means and unit variances. Additionally, they were pre-specified with a fixed correlation of $\rho = 0.3$, and they were allowed to freely correlated with the latent interaction term. We examined two population values of the latent interaction effect: $\gamma_{xm} = 0$ to test the null hypothesis (H_0) and $\gamma_{xm} = 0.3$ (medium effect; Cohen, 1992) to test the alternative hypothesis (H_1). The variance of latent outcome variable ξ_{y_j} was set to 1 when $\gamma_{xm} = 0$ under the H_0 condition, so that the variance of disturbance term $\sigma^2_\zeta = 1 - \left(\gamma_x^2 + \gamma_m^2 + \gamma_{xm}^2 + 2\,\gamma_x\gamma_m\,\rho\right)$. Therefore, the values of σ^2_ζ were adjusted according to the latent interaction effect size (e.g., $\sigma^2_\zeta = 1 - (0.3^2 + 0.3^2 + 0 + 2 \times 0.3 \times 0.3 \times 0.3) = 0.766$ when $\gamma_{xm} = 0$, which indicated that the first-order latent predictors and the latent interaction term jointly contributed to explain 23.3% variance in ξ_{y_j}.

Population Measurement Model

After drawing the person-level latent scores from the structural model, we generated observed indicators for each construct. Let i index items with $j = 1, \ldots, K$. The latent outcome variable ξ_{y_j} was measured by three continuous indicators (y_{1j}, y_{2j}, y_{3j}) with differential loadings:

$$y_{ij} \;=\; \lambda_{y_i}\,\xi_{y_j} \;+\; \delta_{y_{ij}}, \qquad i = 1, 2, 3, \tag{2}$$

where λ_ys were sepcified as $\{.50,\ .70,\ .90\}$, and the error variances followed $\delta_{y_{ij}} \sim \mathcal{N}\!\left(0,\ 1 - \lambda^2_{y_i}\right)$.

Following the design in Aytürk et al. (2020), we set ξ_j with three items (i.e., $x_{1j}, \ldots, x_{2j}$) and ξ_m with twelve items (i.e., $m_{1j}, \ldots, m_{12j}$). We first formed continuous precursors with varying item loadings,

$$x^*_{ij} = \lambda_{x_i}\,\xi_{x_j} + \delta_{x_{ij}}, \qquad i = 1, 2, 3, \tag{3}$$

$$m^*_{ij} = \lambda_{m_i}\,\xi_{m_j} + \delta_{m_{ij}}, \qquad i = 1, \ldots, 12, \tag{4}$$

where x^*_{ij} was the score of the underlying latent continuous variable for each observed categorical item i. $\delta_{x_{ij}}$ was the individual-specific error term for each observed indicator i which followed a normal distribution with a zero mean and a variance $\theta_{x_{ij}}$.

Categories of Observed Indicators for ξ_x and. Observed categorical indicators (e.g., x_{ij} for ξ_{x_j}) were generated from a normal–ogive (cumulative probit) graded response model with differential factor loadings (Cho, 2023). For person j and item i measuring latent factor η_x,

$$x^*_{ij} = \lambda_{x_i}\xi_{x_j} + \delta_{x_{ij}}, \qquad \delta_{x_{ij}} \sim \mathcal{N}(0,1), \tag{5}$$

where x^*_{ij} is the score of underlying latent continuous variable for each observed categorical item i. $\delta_{x_{ij}}$ is the individual-specific error term for each observed indicator i, which follows a standard normal distribution. Given x^*_{ij}, the observed categorical item x_{ij} can be created with multiple categories through thresholding:

$$x_{ij} = \begin{cases} 0 & \text{if } x^*_{ij} < \beta_{x_{i1}} \\ k & \text{if } \beta_{x_{ik}} \leq x^*_{ij} < \beta_{x_{i(k+1)}} \\ K-1 & \text{if } \beta_{x_{i(K-1)}} \leq x^*_{ij} \end{cases}, \tag{6}$$

where β_{ik} is the threshold parameter between the kth and $(k+1)$th category for $k = 1, 2, ..., K$. The observed items for ξ_{m_j} (e.g., m_{ij}) were generated analogously. $\{\tau_{m_i,k}\}$.

Simulation Design

Material

Procedure

Data analysis

We used R (Version 4.5.1; R Core Team, 2025) and the R-packages *papaja* (Version 0.1.3; Aust & Barth, 2024) and *tinylabels* (Version 0.2.5; Barth, 2025) for all our analyses.

Results

Discussion

References

Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://doi.org/10.32614/CRAN.package.papaja

Barth, M. (2025). *tinylabels: Lightweight variable labels*. https://doi.org/10.32614/CRAN.package.tinylabels

R Core Team. (2025). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/