Results and Discussion

Jimmy

Results and Discussion

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

Average Raw Bias and Standardized Bias for γ_{xm}

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

31 absolute SB and B for RAPI first increased from .14(.08) to .18(.03) and then decreased to

32 .08(.01) as $Corr[\xi_x, \xi_{xm}] = 0$, and those for matched-pair UPI first increased from .10(.03) to

33 .11(.02) and then decreased to .03(.00) as $Corr[\xi_x, \xi_{xm}] = 0.6$.

34   The B values generally became smaller as sample size increased for the three methods,

35 which aligned with the statistical property of SEM models such that larger sample sizes tend

36 to provide more accurate and reliable parameter estimates and reduce sampling errors.

37 Nevertheless this pattern was not exactly consistent with the absolute SB because the

38 empirical standard deviation of B decreased as the sample size increased, which might

39 amplify the absolute SB. For instance, when $\rho = .70$ and $Corr[\xi_x, \xi_m] = 0$, the magnitude of

40 raw average biases decreased from .08 to .01 for RAPI while the absolute SB first increased

41 from .14 to .21 and then decreased to .19. The above findings revealed that the pattern

42 displayed through B values might be masked by the corresponding empirical standard

43 deviation, and the comparability of raw average biases need to be cautiously considered in

44 standard units.

45   It was found that the absolute SB of RAPI and matched-pair UPI were almost

46 positive while some of the 2S-PA-Int estimates were negative across simulation conditions.

47 The results were consistent with previous findings for RAPI and matched-pair UPI such that

48 they tended to provide overestimated interaction estimates with high correlations between

49 first-order latent predictors and low reliability (Marsh et al., 2004; Hsiao et al., 2018).

50 2S-PA-Int did not show a clear sign of over or underestimation, indicating that the absolute

51 SB values were more randomly distributed. Nevertheless, all the methods yielded comparably

52 low standardized biases across simulation conditions, which was acceptable for practical use.

53 **Relative SE Bias of γ_{xm}**

54   Table 2 showed the robust relative standard error (SE) bias ratio with outlier

55 proportions of SE when $\gamma_{xm} = 0.3$. All the values outside the -10% ~ 10% range were bolded.

56 Generally, the values of robust relative SE bias were all below 10% for RAPI, matched-pair

UPI, and 2S-PA-Int across the conditions of medium to high reliability level. The ranges

were from .56%(1.55%) to 8.33%(1.65%) for RAPI, .09%(1.55%) to -8.96%(5.85%) for

matched-pair UPI, and -.57%(1.40%) to -7.39%(1.30%) for 2S-PA-Int, which implied that

the estimated SE values of $\hat{\gamma}_{xm}$ were not biased and $\hat{\gamma}_{xm}$ estimated by the three methods

under medium to high reliability showed less variability across other conditions. Compared

to 2S-PA-Int, matched-pair UPI produced two relative SE values outside the acceptable

range under the conditions of small sample size ($N = 100$) and low reliability ($\rho = .70$):

-11.52%(8.15%) and -14.14%(8.40%), meaning that the SE values were negatively biased. As

for RAPI, unacceptable relative SE biases were generated across various conditions under

low reliability ($\rho = .7$). The outlier proportions of SEs identified by the IQR method showed

declining trends as sample size increased and reliability levels improved for all the methods,

meaning that the estimation of γ_{xm} became more accurate and stable with less extreme

values. For instance, the proportion of outliers for RAPI decreased from 10.90% to 5.55%

and to 2.60% under $N = 100$ and $\rho = .70$, while within the condition of $N = 100$ the

proportion decreased from 10.90 to 5.40 to 1.90 as ρ increased.

However, the robust relative SE bias did not demonstrate a clear pattern associated

with the population reliability value and sample size for the three methods. It was found

that even though under the condition of large sample size and high reliability, the relative SE

bias could be higher then those under worse conditions. For example, when $N = 500$ and

$\rho = .90$, the relative SE bias produced by 2S-PA-Int was -2.41%, while the value was -1.02

under $N = 100$ and $\rho = .90$. It implied that the estimated SEs were unstably deviated from

the true reference value (i.e., empirical standard deviation of γ_{xm}). Additionally, the overall

values of the relative SE bias were negative for matched-pair UPI and 2S-PA-Int, and almost

positive for RAPI, which indicated that the SEs were systematically underestimated for

matched-pair UPI and 2S-PA-Int but overestimated for RAPI.

Coverage Rate of 95% CI of γ_{xm}

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

RMSE of γ_{xm}

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

Discussion

Applied researchers often focus on complex relationships between variables, such as interactions. However, classical regression models, which assume variables are free of

measurement error, have been shown to produce biased estimates. Consequently, latent variables approaches with the SEM framework are increasingly being considered. In this study, we reviewed and compared the performance of matched-pair UPI and RAPI with 2S-PA-Int in estimating interaction effects on congeneric items with varying factor loadings and errors.

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

Higher coverage rates with 95% CI for RAPI around 95% ~ 97% were observed in our results, implying that RAPI has higher chance and accuracy in capturing true interaction effects within the 95% confidence intervals, compared to matched-pair UPI and 2S-PA-Int. 2S-PA-Int estimated interaction effects with acceptable coverage rates as well, though

slightly lower than those estimated by RAPI, implying that 2S-PA-Int is able to capture the true effects with high likelihood. Matched-pair UPI was affected mostly by small sample size and low reliability level, which implied that it was not as robust as RAPI and 2S-PA-Int and not recommended to use under this condition. Our results showed consistency with past research mentioned above; however, Marsh et al. (2004) did not test matched-pair UPI on fully congeneric items and it may imply that matched-pair UPI has less chance of capturing true effects with varied error variances within first-order indicators.

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

162   for latent interaction estimation.

163   Revisiting Marsh's criteria of a good model of estimating latent interaction effects,
164   2S-PA-Int is practically preferable in terms of simple model specification as a single-indicator
165   model and comprehensive usage of information by using factors scores based on all first-order
166   indicators. Specifically, models overloaded with indicators may have difficulties in reaching
167   convergence due to the intricate covariance structures to be estimated, potentially resulting
168   in non-identifiable models (Bollen, 1989). Furthermore, Byrne (2016) highlights that
169   excessive indicators can introduce redundancy, complicating the model unnecessarily and
170   increasing the likelihood of estimation problems. Thus, 2S-PA-Int should be a safer
171   alternative to matched-pair UPI especially with small sample size and low reliability level.
172   Compared to RAPI, 2S-PA-Int is more advantageous in terms of stability and accuracy of
173   interaction estimates.

Limitations

175   A few limitations in the current study are discussed below. First, Hsiao et al. (2018)
176   mentioned that RAPI may be more approachable when researchers do not have the access of
177   original data and have to analyze secondary data since composite scores are usually reported
178   with reliability index (e.g., usually Cronbach's α). Since reporting factor scores with
179   standard errors is still not a commonly applied practice, some secondary dataset may not
180   contain computed factors scores and thus 2S-PA-Int is not applicable in this case. Second,
181   currently the congeneric items in this study design are all continuous with normal
182   distributions. Given that categorical data is frequently used in psychology research to
183   capture the qualitative aspects of human behavior, attitudes, and characteristics (Brown,
184   2015; Kline, 2016), 2S-PA-Int has not been evaluated and should be studied with categorical
185   items in the future. Third, the study designs in the past methodological paper on latent
186   interaction effects were almost simply structured with two latent predictors and one
187   interaction term, which could be insufficient to accommodate more complicated real-world

188 scenarios, such as multiple interaction terms. Besides, multilevel design rencently is

189 increasingly used in educational, counseling, and organizational research (e.g., students

190 nested in classrooms, patients nested in clinics, employees nested in companies). Thus, it is

191 worth exploring the potential of 2S-PA-Int with complicated data types and structures with

192 varied sample sizes and reliability levels.

Table 1

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

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

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

Table 2

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

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

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

Table 3

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

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

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

Table 4

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

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

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