Testing Moderation in Business and Psychological Studies with Latent Moderated Structural Equations

Gordon W. Cheung[1] · Helena D. Cooper-Thomas[2] · Rebecca S. Lau[3] · Linda C. Wang[4]

Abstract

Most organizational researchers understand the detrimental effects of measurement errors in testing relationships among latent variables and hence adopt structural equation modeling (SEM) to control for measurement errors. Nonetheless, many of them revert to regression-based approaches, such as moderated multiple regression (MMR), when testing for moderating and other nonlinear effects. The predominance of MMR is likely due to the limited evidence showing the superiority of latent interaction approaches over regression-based approaches combined with the previous complicated procedures for testing latent interactions. In this teaching note, we first briefly explain the latent moderated structural equations (LMS) approach, which estimates latent interaction effects while controlling for measurement errors. Then we explain the reliability-corrected single-indicator LMS (RCSLMS) approach to testing latent interactions with summated scales and correcting for measurement errors, yielding results which approximate those from LMS. Next, we report simulation results illustrating that LMS and RCSLMS outperform MMR in terms of accuracy of point estimates and confidence intervals for interaction effects under various conditions. Then, we show how LMS and RCSLMS can be implemented with Mplus, providing an example-based tutorial to demonstrate a 4-step procedure for testing a range of latent interactions, as well as the decisions at each step. Finally, we conclude with answers to some frequently asked questions when testing latent interactions. As supplementary files to support researchers, we provide a narrated PowerPoint presentation, all Mplus syntax and output files, data sets for numerical examples, and Excel files for conducting the loglikelihood values difference test and plotting the latent interaction effects.

The assessment of interactions between variables is critical to theoretical and empirical advances in business and psychological research. An interaction indicates that the relationship between two variables is either contingent upon or varies across levels of another variable, with such interactions commonly proposed in business and psychological theories. While most researchers report results of their measurement models for latent variables based on multiple-item surveys used to collect data, and employ structural equation modeling (SEM) with latent variables to control for measurement errors

✉ Gordon W. Cheung
gordon.cheung@auckland.ac.nz

Helena D. Cooper-Thomas
helena.cooper.thomas@aut.ac.nz

Rebecca S. Lau
rebeccasylau@gmail.com

Linda C. Wang
lindawang@hsu.edu.hk

1 Department of Management and International Business, The University of Auckland, Room 469, Owen G Glenn Building, 4/F., 12 Grafton Road, Auckland, Private Bag 92019, Auckland 1142, New Zealand

2 Management Department, Faculty of Business Economics & Law, Auckland University of Technology, Auckland, New Zealand

3 Hong Kong, China

4 Department of Management, School of Business, The Hang Seng University of Hong Kong, Lee Quo Wei Academic Building, Hang Shin Link, Siu Lek Yuen, Shatin, Hong Kong

when estimating path coefficients in models without latent interaction, the majority of them revert to regression analysis (such as moderated multiple regression (MMR)) or path analysis with observed variables (such as PROCESS; Hayes, 2013) when estimating interaction effects. In their recent review, Cortina, Markell-Goldstein, Green, and Chang (2019) confirmed this, showing that two-thirds of tests for interaction effects adopted non-latent approaches, predominantly MMR.

One major concern with using MMR to test for interaction effects is the detrimental influence of measures with low reliability (Aguinis, 1995). While most researchers pay attention to the reliability of their variables, few have paid attention to the reliability of the interaction term that can be much lower than the reliability of the two variables forming the interaction (Aguinis, Edwards, & Bradley, 2017; Busemeyer & Jones, 1983), and which results in larger bias in the estimated interaction effect. Consequently, researchers have focused on developing methods for estimating interaction effects using SEM (for example, Kenny & Judd, 1984; Marsh, Wen, & Hau, 2004). Prior proposed SEM methods, such as the product-indicator approach (PI; Kenny & Judd, 1984), require creating a latent interaction term from the indicators and imposing sophisticated nonlinear constraints. Such complexity in creating and testing interaction effects with SEM has likely limited its adoption. Contrasting this, the interaction term in MMR is simply computed from the product of observed variables.

Recently, Cortina et al. (2019) presented a flowchart with a list of estimation approaches for testing latent interaction effects, starting with the most accurate estimation method and moving down to less accurate ones. Notably, they place the PI approach and the latent moderated structural equations (LMS) approach (Klein & Moosbrugger, 2000) at the top of the list. While conceptually easier to understand, PI has a number of drawbacks including the complexity of creating numerous nonlinear constraints among latent variables and indicator residuals; violation of the normal distribution assumption for latent variables and their indicators resulting in incorrect estimates; and lack of information on how to use PI with more complex models, such as models with multiple interactions or a three-way interaction.

On the other hand, recent developments for implementing LMS in Mplus allow users to estimate latent interaction effects easily with relatively simple syntax, even when the number of indicators of the latent variables forming the latent interaction is large, and for complicated models such as those with a curvilinear three-way interaction. Several simulation studies have demonstrated the superiority of LMS over PI in terms of higher accuracy of estimates, smaller standard errors, and higher power (Kelava et al., 2011; Klein & Moosbrugger, 2000). Yet a limitation of LMS is that, because it requires a

more complex estimation method, a model with multiple latent interactions and a large sample size may require a long time to run or computing resources that go beyond ordinary personal computers. In such cases, reliability-corrected single-indicator LMS (RCSLMS) provides an approximation approach that adopts LMS with summated scales and correction of measurement errors (Cheung & Lau, 2017).

Despite a number of studies explaining the benefits of LMS when estimating latent interaction effects (for example, Kelava et al., 2011; Klein & Moosbrugger, 2000), and a number of teaching notes elaborating the procedures (for example, Cheung & Lau, 2017; Maslowsky, Jager, & Hemken, 2015; Muthén, 2012; Sardeshmukh & Vandenberg, 2017), only 6% of the analyses in Cortina et al.'s (2019) review applied LMS. Moreover, MMR remains the prevalent approach for assessing interaction effects in nonexperimental studies (Cortina et al., 2019) in spite of evidence that various methodological artifacts negatively affect MMR when testing for interaction effects (for example, Aguinis, 1995; Aguinis et al., 2017).

To address this situation, the aim of this teaching note is to motivate and enable researchers to employ LMS when investigating latent interaction effects, including RCSLMS when appropriate. The remaining paper is divided into four sections. In the first section, we briefly explain the conceptual basis of LMS and RCSLMS approaches to estimating latent interactions since one possible reason for the low utilization of LMS is that it is more difficult to understand than the PI approach (Cortina et al., 2019). In the second section, we establish the superiority of LMS and RCSLMS compared with MMR by providing the results of two simulation studies. In the third section, we demonstrate a 4-step procedure to facilitate the easy implementation of LMS for testing latent interaction effects using Mplus. Notwithstanding the advantages of LMS, a further possible reason for its low adoption rate may be the practical concerns raised in previous LMS-oriented teaching notes, such as challenges in computing time, estimating the model fit and the effect size of the interaction term, and obtaining standardized coefficients (Maslowsky et al., 2015; Sardeshmukh & Vandenberg, 2017). More specifically, while Maslowsky et al. (2015) provided a tutorial on applying LMS, we demonstrate how the three limitations they raised can be resolved, and show how moderating effects can be better probed by the Johnson-Neyman approach (Johnson & Neyman, 1936). Building upon the work of Dawson (2014) on moderation in management research, we also demonstrate LMS for simple slope tests, examining three-way interaction and other curvilinear effects. In the final section, we provide answers to some frequently asked questions regarding the application of LMS, including tips that increase the computing capacity and lower the computing requirements of LMS.

Early SEM Approaches to Estimating Interaction Effects

Estimating interaction effects using MMR assumes that observed variables are measured without errors (Kline, 2016).[1] As such, an interaction term constructed from observed variables includes not only the true score of the latent interaction, but also random measurement errors. Because SEM isolates indicator residuals from latent constructs (Bollen, 1989), estimating interaction effects from the latent constructs controls for measurement errors; this substantially reduces any bias in estimates caused by measurement errors.

Researchers have developed various approaches for estimating the effects of interactions between latent variables in SEM. Early work on PI, referred to as the *constrained approach*, proposed using the products of all possible pairs of indicators of two latent variables as indicators of the latent interaction term (Algina & Moulder, 2001; Bollen, 1989, pp. 405–406; Jöreskog & Yang, 1996; Kenny & Judd, 1984). While this PI approach is straightforward conceptually, Cortina et al. (2019) highlight a major limitation in that it requires researchers to define numerous nonlinear constraints for the loadings, intercepts, and residual variances of the products of indicators. Typically, when one product indicator is added, five additional concomitant constraints are required (Jöreskog & Yang, 1996). This process rapidly becomes unwieldy when the two latent variables have more than two or three indicators. Hence, Cortina et al. (2019) recommend creating two to three parcels for each latent variable before creating the product indicators and provide R-code accordingly. Nonetheless, the results are unstable as different combinations of the indicators in parcels will yield different results (Hall, Snell, & Foust, 1999).

Unlike the traditional *constrained approach*, which imposes complicated nonlinear constraints, the *unconstrained approach* introduced by Marsh et al. (2004) imposes no nonlinear constraints. However, the unconstrained approach also raises concerns, one being that the indicators for the latent interaction are created with matched pair products, implying that different pairings of the indicators of the two latent variables may produce different results. In addition, it is unclear how to form the matched pair product indicators when the two latent variables have different numbers of indicators. While the two-stage least square (2SLS) method proposed by Bollen (1995) and Bollen and Paxton (1998) avoids employing complicated nonlinear constraints, as only one of the indicators of the outcome variable is used, the results may vary when different indicators are used. A second issue with PI, together

with its variations, is that the maximum likelihood (ML) estimator assumes that latent variables and their indicators are normally distributed. However, when nonlinear effects exist, the endogenous latent variable (dependent variable) and its indicators are not normally distributed. As a result, PI may provide not only incorrect chi-square values but also biased standard errors and significance tests for interaction effects (Bollen, 1989, p. 406; Cortina, Chen, & Dunlap, 2001; Kelava et al., 2011; Klein & Moosbrugger, 2000; Sardeshmukh & Vandenberg, 2017). A final issue with PI is that, to date, there is no explanation of how it should be implemented for more complicated models, including those with multiple moderators, three-way interaction, moderated quadratic effects, and polynomial regression.

Cortina et al. (2001) provided a comprehensive review of the problems of various early approaches to modeling latent interactions, while Cortina et al. (2019) and Sardeshmukh and Vandenberg (2017) provided brief reviews on PI and more recent approaches. Interested readers are referred to these reviews outlining the issues inherent to various approaches.

Conceptual Basis of Latent Moderated Structural Equations

Investigations of interaction effects stem from the same conceptual foundation. Equation 1 below shows the general equation that specifies the relationships among two exogenous latent variables (independent variables) X and Z, and one endogenous latent variable (dependent variable) Y with both interaction and quadratic terms (Kelava et al., 2011, p. 467).[2]

$$Y = \alpha + \gamma_1 X + \gamma_2 Z + \omega_{12} XZ + \omega_{11} X^2 + \omega_{22} Z^2 + \zeta \quad (1)$$

where α is the intercept of the regression equation, γ_1 and γ_2 are the main effects of X and Z respectively, ω_{12} is the interaction effect of XZ, ω_{11} and ω_{22} are the quadratic effects of X and Z respectively, and ζ is the residual of the regression equation. An important development in testing for latent interactions was the introduction of LMS by Klein and Moosbrugger (2000). LMS estimates the regression coefficient ω_{ij} of all freely estimated two-way interactions of exogenous latent variables with the following equation (Klein & Moosbrugger, 2000, p. 460):

$$\eta = \alpha + \Gamma\xi + \xi'\Omega\xi + \zeta \quad (2)$$

where η is the latent endogenous variable, α is the intercept, ξ are the latent exogenous variables, Γ are the main effects of ξ

---

[1] As in most of the discussions in SEM, measurement errors in this article refer to random measurement errors only. Systematic measurement errors, such as method effects and researcher-introduced bias, can only be accounted for using specific research designs and other modeling approaches.

[2] Equation 1 shows that either X or Z can be treated as the moderator. In other words, both independent variable and moderator are treated in the same way statistically. Theory and hypotheses should be used to determine which is the independent variable and which is the moderator (Dawson, 2014).

on η, Ω are the nonlinear effects of ξ on η, and ζ is the residual of η.

Hence, Eq. 1 can be expressed as follows (Kelava et al., 2011, p. 467):

$$Y = \alpha + (\gamma_1 \ \gamma_2) \cdot \begin{pmatrix} X \\ Z \end{pmatrix} + (X \ \ Z) \cdot \begin{pmatrix} \omega_{11} & \omega_{12} \\ 0 & \omega_{22} \end{pmatrix} \cdot \begin{pmatrix} X \\ Z \end{pmatrix} + \zeta \quad (3)$$

While PI treats the product term XZ and the quadratic terms X^2 and Z^2 in Eq. 1 as latent variables and therefore requires the creation of product indicators to "measure" these latent variables, Eq. 3 shows that LMS uses matrix multiplication of X and Z to estimate the interaction and quadratic effects on Y without creating latent variables to represent the product term XZ and quadratic terms X^2 and Z^2. Thus, LMS does not require the creation of product indicators to represent the latent interaction and quadratic terms, which avoids the need to impose complicated nonlinear constraints. This differentiates LMS as superior, contrasting with PI which uses product indicators and nonlinear constraints.

Technical explanations of how LMS estimates model parameters are beyond the scope of this teaching note. Interested readers are referred to Kelava et al. (2011), Klein and Moosbrugger (2000), and Sardeshmukh and Vandenberg (2017) for more details. In short, LMS employs distribution analytic methods to resolve the problems encountered by PI that result from a nonnormally distributed endogenous latent variable and its indicators, and uses an iterative estimation procedure—the expectation-maximization (EM; Dempster, Laird, & Rubin, 1977) algorithm—to provide an ML estimation of model parameters (Kelava et al., 2011). In LMS, only error variables and latent independent variables, but not latent dependent variables, are assumed to be normally distributed. Previous simulation investigations have provided evidence that LMS offers unbiased standard errors and efficient parameter estimates even for data that deviate slightly from a normal distribution (Klein & Moosbrugger, 2000).

Reliability-Corrected Single-Indicator Latent Moderated Structural Equations

While it is advantageous to estimate latent interaction effects with LMS, there are various conditions (which will be explained later) under which LMS is not ideal. Under such conditions, and if all indicators of the construct are measured with the same response scale, we recommend adopting reliability-corrected single-indicator latent moderated structural equations (RCSLMS) to test latent interactions (Cheung & Lau, 2017). RCSLMS is a modified version of LMS in which the values of all indicators for a latent variable are averaged to form a scale score such that each latent variable in the model has only one indicator. The factor loading between the latent variable and the single indicator is then fixed to 1. Instead of

assuming no measurement errors as in MMR, or estimating the residual variances as in LMS, RCSLMS fixes the residual variance of the single indicator to $(1 - \text{reliability})$ times the variance of the observed scale score (Bollen, 1989; Hayduk, 1987). One convenient approach is to use Cronbach's alpha as a measure of reliability. For example, if the scale has Cronbach's alpha of 0.8 and the variance of the average score is 1.5, the residual variance for this latent variable is fixed to $(1 - 0.8)(1.5) = 0.3$. Finally, LMS is employed to estimate the interaction effects of latent variables with single indicators. RCSLMS has been shown to provide estimates that are good approximations to those produced by LMS with all indicators and superior to those provided by regression-based approaches (Cheung & Lau, 2017; Su, Zhang, Liu, & Tay, 2019).

Two Simulations Comparing MMR, LMS, and RCSLMS

In this section, we report the results of two simulations that compared the performance of MMR, LMS, and RCSLMS for estimating latent interaction effects. Similar simulations have been conducted in specific contexts including moderated-mediation (Cheung & Lau, 2017) and polynomial regression (Su et al., 2019). The two simulations in this teaching note examined the impact of various factors on the performance of these three approaches to test for interaction effects in a more general context, in terms of the accuracy and precision of estimated parameters, confidence interval (CI) coverage rates, type I error rates, and statistical power. These simulations were conducted based on the interaction model in Eq. 4, which has been employed in previous LMS simulation studies (Klein & Moosbrugger, 2000; Klein & Muthén, 2007).

$$Y = \alpha + \gamma_1 X + \gamma_2 Z + \omega_{12} XZ + \zeta \quad (4)$$

where α is the intercept, γ_1 is the main effect of X, γ_2 is the main effect of Z, ω_{12} is the interaction effect of XZ, and ζ is the residual of Y.

Simulation 1—Effects of Model Parameters on Performance of MMR, LMS, and RCSLMS

This first simulation examined accuracy and precision of estimates, coverage, and completion rates comparing MMR, LMS, and RCSLMS. Four indicators for each of the three latent variables, X, Z, and Y were created. The levels of model parameters were chosen to approximate models observed in empirical research. In the first simulation, three different sample sizes ($N = 100$, 200, and 500) and three levels of correlation between X and Z ($rXZ = 0.1$, 0.3, and 0.5) were manipulated to represent small, medium, and large correlations

between X and Z (Cohen, 1988, pp. 79–80). Four levels of interaction effects ($\omega_{12} = 0$, 0.12, 0.3, and 0.42) were used. Note that the condition of $\omega_{12} = 0$ was used to examine the type I error rate for the interaction effect. Cohen (1988, pp. 413–414) defined small, medium, and large effect sizes for regression as $f^2 = 0.02$, 0.15, and 0.35 respectively, which can be translated to $\omega_{12} = 0.12$, 0.3, and 0.42, respectively when $rXZ = 0.3$. Cohen's f^2 for the same interaction effects are larger when $rXZ = 0.5$ and are smaller when $rXZ = 0.1$. Since the objective of the first simulation was to examine the accuracy of parameter estimates, we manipulated ω_{12} instead of f^2. The average effect sizes across all rXZ conditions were $f^2 = 0$, 0.02, 0.16, and 0.36. Three levels of residual variance δ of each indicator were set based on Cronbach's alpha of the aggregated score of each latent variable ($\alpha = 0.7$, 0.8, and 0.9). Finally, three patterns of factor loadings for indicators of Z were adopted ($z1 = 1$, $z2 = 1$, $z3 = 1$, $z4 = 1$; $z1 = 1$, $z2 = 1$, $z3 = 0.8$, $z4 = 1.2$; $z1 = 1$, $z2 = 1$, $z3 = 0.8$, $z4 = 0.8$) to examine the effect of unequal factor loadings on the performance of the three approaches. As a result, the standardized factor loadings, which vary as functions of unstandardized factor loadings and size of residual variance, covered a wide range of values from 0.49 to 0.91 in the population parameters. Factor loadings for all indicators for X and Y were set at 1. In total, there were 324 population conditions. Similar to Klein and Moosbrugger (2000) and Klein and Muthén (2007), the values of the main effects were set to $\gamma_1 = 0.4$ and $\gamma_2 = 0.2$. The residual variances ζ for Y were fixed based on the interaction effects. All intercepts in the model, including the indicator intercepts and intercept for Y, were set to 0. All latent variables and indicators were set to be normally distributed with variance equal to 1. Five hundred data sets were simulated with Mplus 8.4 (Muthén & Muthén, 1998-2014) for each of the 324 conditions, giving a total of 162,000 simulated data sets.

Each simulated data set was analyzed with the three approaches for testing interaction effects. For MMR, the values of the four indicators for each latent variable were averaged to form the aggregate score for the analysis. For LMS, the raw data (indicator values) were analyzed with the latent variable interaction function in Mplus 8.4. For RCSLMS, the aggregated scores (instead of indicator values) were analyzed with the latent variable interaction function in Mplus 8.4. Also for RCSLMS, the R-package was used to calculate the variance and Cronbach's alpha of each variable and to create the Mplus input file in which the residual variance of each variable was fixed at $(1 - \text{reliability})$ times the variance of the aggregated score.

When comparing the performance of each approach, we first examined the average estimated interaction effects across 500 simulated data sets for each population condition, which indicates the accuracy of the estimated interaction effect. Then absolute bias was calculated as the average of the absolute difference between the point estimate and the population value (Ledgerwood & Shrout, 2011). Because the standard error

does not reflect the deviation of the point estimate from the population value, the absolute bias is a better representation of the precision of the estimate produced by each approach. The coverage of the 95% CI was then examined. The coverage was calculated as the proportion of the 500 simulated samples for which the population value was included in the 95% CI, thus serving as a measure of CI accuracy. It is expected that 95% of the simulated samples have estimated 95% CI that contain the true value. Finally, one frequently raised concern inherent in LMS is the higher chance of nonconvergence when the sample size is small because a large number of parameters are estimated (Sardeshmukh & Vandenberg, 2017). Therefore, we examined the completion rates of the simulated samples.

Results—Simulation 1 on the Effects of Model Parameters[3]

Accuracy of Estimates The average estimated interaction effects of each approach are shown in Table 1. On average, LMS produced very accurate estimates irrespective of the population conditions. RCSLMS also produced very accurate estimates, with average estimated interaction effects deviating from population values by less than 1% even under the unequal factor loading conditions. While MMR produced very accurate estimates when there was no interaction effect in the population, on average, the interaction effects were underestimated by -21.62%. The attenuation for MMR was more serious when reliability was lower, with the interaction effects underestimated by -33.33% when $\alpha = 0.7$. For all three approaches, increased sample size did not result in more accurate estimates.

Precision of Estimates The average absolute biases are presented in Table 2. Precision is the combined effect of accuracy and variation of estimates. In general, the estimated interaction effect was more precise when both reliability and sample size increased because the standard error of estimates was smaller. Because LMS and RCSLMS include measurement errors in the calculation of standard error of estimates, LMS had the largest standard error of estimates, closely followed by RCSLMS, whereas MMR resulted in the smallest standard error of estimates (as presented in Table S3 in the supplementary files). When there was no interaction effect or a small interaction effect, MMR produced the most precise estimates because of the smaller standard errors and only small attenuation of estimated parameters. With medium and large interaction effects, both LMS and RCSLMS produced more precise estimates, whereas the attenuation of estimated parameters in MMR became more severe.

---

[3] Results presented in Tables 1 to 3 are averages across all three patterns of factor loadings. Tables showing results for each pattern of factor loadings are available in the supplementary files.

Coverage of 95% Confidence Interval LMS resulted in the most accurate 95% CI, with an average coverage of 94.30%. This was closely followed by RCSLMS with an average coverage of 92.8%. On the other hand, the attenuation of estimated parameters and the smaller standard error of estimates produced by MMR together resulted in inaccurate CIs, with an average coverage of only 69.7%. Since increased sample size led to smaller standard errors but not more accurate estimates by MMR, the coverage of CI was lowered when sample size was larger. This was most obvious when $N = 200$ or 500, reliability was at 0.7, and the interaction effect was medium or large. In such conditions, the average coverage of the 95% CI produced by MMR was only 14.0%. This is of particular concern because an increasing number of studies, in particular moderated-mediation studies, report CIs generated from MMR or related regression approaches (such as PROCESS; Hayes, 2013) which are likely to be

inaccurate. Coverages of CI are presented in Table S5 in the supplementary files.

Completion Rates In the first simulation, MMR completed estimation for all 162,000 simulated data sets. Both LMS and RCSLMS also had very high completion rates, probably because there are only twelve indicators and one latent interaction in the model. More specifically, out of the 162,000 simulated data sets, LMS completed all but 14 cases, and RCSLMS completed all but six cases. All incomplete estimations were under conditions with $N = 100$.

Simulation 2—Comparing Type I Error Rates and Power of MMR, LMS, and RCSLMS

We conducted a second simulation to examine type I error rates and power. In this simulation, we only used the condition

Table 1 Simulation 1: average estimated parameters

rXZ	Alpha	N	$\omega_{12} = 0$			$\omega_{12} = .12$			$\omega_{12} = .30$			$\omega_{12} = .42$		
			MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS
.1	.7	100	.004	.005	.004	.077	.119	.120	.184	.286	.291	.258	.406	.413
		200	.004	.007	.006	.077	.123	.124	.185	.295	.299	.258	.412	.417
		500	.004	.006	.006	.077	.124	.125	.186	.299	.301	.259	.419	.421
	.8	100	.005	.006	.005	.089	.122	.122	.217	.297	.299	.300	.410	.413
		200	.004	.006	.006	.090	.124	.124	.216	.298	.300	.300	.415	.417
		500	.004	.005	.005	.089	.123	.124	.218	.301	.302	.303	.419	.420
	.9	100	.005	.006	.006	.106	.123	.124	.255	.298	.298	.356	.415	.416
		200	.004	.005	.005	.105	.123	.124	.257	.300	.300	.357	.416	.417
		500	.003	.004	.004	.105	.123	.123	.258	.301	.302	.361	.420	.420
.3	.7	100	.003	.005	.004	.080	.118	.119	.192	.286	.289	.270	.407	.411
		200	.004	.007	.006	.081	.123	.124	.197	.298	.300	.273	.413	.417
		500	.004	.006	.006	.081	.124	.125	.198	.302	.303	.274	.419	.420
	.8	100	.004	.005	.004	.092	.121	.121	.224	.293	.294	.314	.410	.413
		200	.005	.006	.006	.094	.124	.124	.228	.300	.300	.318	.417	.419
		500	.004	.005	.006	.094	.124	.124	.229	.302	.302	.318	.419	.420
	.9	100	.005	.006	.005	.109	.123	.123	.264	.295	.296	.370	.414	.415
		200	.005	.005	.006	.110	.124	.124	.269	.301	.301	.373	.417	.417
		500	.004	.004	.004	.110	.123	.123	.269	.300	.300	.376	.420	.420
.5	.7	100	.002	.004	.002	.086	.119	.118	.211	.291	.291	.297	.410	.413
		200	.004	.006	.006	.089	.123	.123	.215	.295	.297	.301	.416	.418
		500	.004	.005	.006	.090	.124	.124	.217	.300	.301	.302	.418	.419
	.8	100	.003	.004	.003	.099	.121	.120	.244	.293	.294	.338	.408	.409
		200	.005	.006	.006	.103	.124	.124	.248	.298	.299	.346	.417	.418
		500	.004	.005	.005	.103	.123	.124	.251	.301	.302	.349	.420	.420
	.9	100	.004	.005	.004	.117	.122	.122	.286	.298	.298	.398	.414	.415
		200	.005	.005	.006	.119	.124	.124	.289	.299	.299	.405	.418	.419
		500	.004	.004	.004	.119	.122	.123	.291	.300	.300	.407	.419	.420

MMR moderated multiple regression approach, *LMS* latent moderated structural equations approach, *RCSLMS* reliability-corrected single-indicator latent moderated structural equations approach, ω_{12} interaction effect

of equal factor loadings, and all population parameters were the same as those in the first simulation except for the sample size and the magnitude of interaction effects. We manipulated six levels of sample size ($N = 100, 150, 200, 300, 400,$ and 500) and four levels of f^2 ($0, 0.02, 0.15,$ and 0.35). We manipulated f^2 instead of ω_{12} such that readers can find the power and required sample size for small, medium, and large effect sizes from the results. The second simulation included 216 population conditions, giving a total of 108,000 simulated data sets. We calculated the proportion of the 500 simulated samples in which the 95% CI did not include zero. This proportion represented a type I error when the population interaction effect was zero and represented power when the population interaction effects were at $f^2 = 0.02, 0.15,$ and 0.35.

Results—Simulation 2 Comparing Type I Error Rates and Power

Type I Error and Power Table 3 shows the results of simulation 2. Columns on the left show the type I error rates of the three approaches when the population interaction effect $f^2 = 0$. The average type I error rates for MMR, RCSLMS, and LMS across all samples were 0.070, 0.057, and 0.070 respectively. The type I error rates for MMR were higher than 0.080 when $N = 100$ and 150 (0.085 and 0.083 respectively), and for RCSLMS when $N = 100$ (0.091). It should be noted that the type I error rates for LMS were above 0.08 when there was high reliability ($\alpha = 0.9$), high rXZ (0.5), and $N = 200$ or smaller. None of the approaches had adequate power (0.8) for testing a small interaction effect except under the most favorable conditions in which $N = 400$ or higher, $rXZ = 0.5$, and $\alpha = 0.9$. Power increased with larger sample size and higher reliability. For the most part, all three approaches produced adequate power for testing medium and large interaction effects with $N = 150$ or higher when $\alpha = 0.8$ or higher.

When to Adopt MMR, LMS, or RCSLMS

Given the simulation results reported above, and from previous studies (Cheung & Lau, 2017; Su et al., 2019), we provide some general rules for choosing between MMR, LMS, and RCSLMS when testing interaction effects, which are summarized in Fig. 1. In general, similar to the advice by Su et al. (2019), we recommend using LMS whenever possible, which is also in line with Cortina et al.'s (2019) recommendation to test interactions with a fully latent approach. The two key advantages of LMS are that it provides the best estimates of both the size of measurement errors and the relationships between the indicators and latent variables by allowing factor loadings to vary across indicators. In turn, LMS produces the most accurate estimated parameters and CIs. LMS also allows higher model complexities. One shortcoming of LMS indicated by the simulations is that, when $N = 100$, it

yields large standard errors, large absolute differences, and low power. Nevertheless, this shortcoming can be resolved by increasing the sample size. If LMS is not feasible, then RCSLMS is usually the next best option because it provides more accurate estimates and better CI coverage than MMR.

However, MMR is the only choice when both X and Z are measured directly without errors, because both LMS and RCSLMS require at least one of the variables to be latent. There may be instances where X and Z are measured directly but the researcher suspects there are measurement errors, yet Cronbach's alpha cannot be estimated from the sample. The best solution is to use RCSLMS with reliability estimated from other methods, such as split-half and test-retest reliabilities. If reliability cannot be estimated from the sample, RCSLMS can be used with reliability taken from previous studies. In such cases, measurement errors should be fixed with slightly overestimated reliability (by $+ 0.05$) to avoid over-adjustment of measurement errors (Savalei, 2019).[4] If either X or Z is a categorical variable and the other is a continuous latent variable, multiple group analysis should be adopted and interaction effects tested by comparing the effects on Y using a direct comparison method (Lau & Cheung, 2012).

If either or both X and Z are latent, then LMS should be used when feasible because it provides greater modeling flexibility. When one of the variables is observed, that variable should be modeled as a single indicator latent variable by fixing the factor loading to 1 and residual variance to 0 (or 0.01 when Mplus gives an error message since residual variance at 0 may sometimes cause problems in estimating the standard errors). This converts the exogenous observed variable into a latent variable with the benefit that, by default, Mplus estimates all the covariances among exogenous latent variables, whereas this is not the case for observed variables, for which covariances would have to be coded. A further benefit of converting the observed variable into a single indicator latent variable is that, by default, Mplus centers the variable.

LMS may not be feasible when:

(a) The sample size is smaller than the number of estimated parameters. Models in empirical studies frequently involve many indicators for multiple variables, but sample size may be limited, such as in experiments or studies at group or organization levels.

---

[4] The reliability level should be set a priori; a sensitivity analysis using multiple levels of reliability is not recommended because that would be exploratory (Savalei, 2019). Based on the results of a simulation study, Savalei (2019) recommended fixing the measurement error a priori by a slightly over-estimated reliability ($+ 0.05$) when the researcher is comfortable in guessing the reliability. However, if the reliability is underestimated ($- 0.05$) or overestimated by a larger extent ($+ 0.15$), the recommended approach will result in less accurate and less precise parameter estimates, as well as lower coverage of CI and inflated type I error when compared with either SEM or the single indicator with Cronbach's alpha correction approaches. These negative consequences were more obvious when the sample size approached 200.

Table 2 Simulation 1: absolute difference from population values

rXZ	Alpha	N	$\omega_{12}=0$			$\omega_{12}=.12$			$\omega_{12}=.30$			$\omega_{12}=.42$		
			MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS
.1	.7	100	.079	.129	.128	.086	.129	.128	.129	.126	.123	.166	.119	.116
		200	.056	.091	.090	.067	.090	.089	.118	.084	.083	.163	.082	.082
		500	.033	.053	.054	.050	.053	.053	.114	.050	.051	.161	.047	.048
	.8	100	.078	.108	.109	.082	.109	.109	.104	.104	.105	.129	.099	.099
		200	.056	.076	.077	.061	.075	.075	.091	.071	.071	.122	.068	.069
		500	.033	.045	.046	.042	.045	.045	.083	.042	.043	.117	.040	.041
	.9	100	.075	.089	.089	.078	.090	.090	.084	.086	.086	.091	.079	.080
		200	.054	.064	.064	.056	.063	.063	.063	.059	.059	.075	.056	.056
		500	.032	.037	.037	.034	.037	.037	.048	.035	.035	.062	.032	.032
.3	.7	100	.078	.122	.120	.085	.123	.121	.121	.118	.116	.155	.112	.109
		200	.055	.084	.083	.064	.083	.082	.108	.077	.077	.148	.075	.075
		500	.034	.051	.051	.047	.050	.051	.102	.048	.049	.146	.046	.047
	.8	100	.077	.103	.102	.081	.104	.103	.100	.101	.100	.118	.092	.092
		200	.055	.071	.071	.058	.070	.070	.082	.066	.065	.107	.062	.062
		500	.034	.043	.044	.040	.043	.044	.072	.041	.041	.102	.038	.038
	.9	100	.074	.085	.085	.076	.086	.085	.081	.083	.082	.086	.077	.077
		200	.053	.059	.059	.054	.059	.059	.058	.054	.054	.066	.051	.051
		500	.032	.036	.036	.033	.036	.036	.041	.033	.033	.049	.031	.032
.5	.7	100	.078	.110	.107	.083	.111	.108	.109	.105	.104	.134	.099	.097
		200	.055	.074	.072	.060	.073	.072	.094	.069	.068	.123	.066	.065
		500	.034	.045	.046	.043	.045	.045	.084	.044	.044	.118	.040	.040
	.8	100	.077	.095	.093	.079	.095	.094	.090	.089	.089	.105	.085	.085
		200	.054	.064	.063	.055	.063	.063	.070	.060	.059	.085	.054	.054
		500	.034	.039	.040	.037	.039	.039	.055	.037	.037	.073	.034	.035
	.9	100	.074	.079	.078	.075	.079	.079	.076	.074	.073	.078	.070	.070
		200	.052	.053	.053	.052	.053	.053	.054	.050	.050	.054	.045	.045
		500	.032	.033	.033	.032	.032	.032	.032	.030	.030	.034	.027	.028

MMR moderated multiple regression approach, *LMS* latent moderated structural equations approach, *RCSLMS* reliability-corrected single-indicator latent moderated structural equations approach, ω_{12} interaction effect

(b) The model does not converge, usually due to model complexities and small sample size.

(c) Complex models that require greater computing capacities than found in most personal computers. The larger the number of latent variables involved in one or more latent interactions, the greater the number of numerical integration points required. Mplus will provide a warning message when computing capacities are exceeded, which is likely to occur when five or more latent variables are involved in multiple interactions.

(d) Models require excessive amounts of time for estimation. This is normally the case when bootstrapping is required, such as in moderated-mediation models or in testing simple slope effects. For example, when the number of indicators is large or the sample size is large, it may take about 1 minute to estimate a model. Multiplying 1 minute by 2000 bootstrap samples gives over 30 hours required for estimation, indicating the excessive amount of time for testing a model.

(e) When Mplus is not available. Currently, among the most commonly used SEM software programs, LMS is only available with Mplus. The R-based lavaan package does not provide LMS estimation. Although the R-based nlsem package (Umbach, Naumann, Brandt, & Kelava, 2017) supports fitting nonlinear structural equation models with LMS, results from nlsem are very sensitive to the starting values used, and hence, even the factor loadings may not be comparable with results from lavaan or Mplus. When using nlsem, the recommendation is to run the model 100 times with different starting values and then average the results (private communication with N. Umbach). Hence, until this problem with nlsem has been resolved, it is not recommended for implementing LMS. When the full version of Mplus is not available, one may consider estimating the model with RCSLMS using the Mplus demo version, which can be downloaded at no cost from the Mplus homepage

(https://www.statmodel.com/demo.shtml). However, one limitation of estimating latent interactions with RCSLMS using the Mplus demo version is that only six variables are allowed in the model, although that should be enough for many empirical studies. If there are more than six variables in the model, then the full version of Mplus should be acquired to test the model.

Compared to LMS, RCSLMS involves fewer parameter estimates. This simpler model yields several benefits, namely that more complex models can be estimated when the sample size is small, and that the elapsed time for estimating the model is much shorter compared with the LMS approach. Several simulation studies have shown RCSLMS performs equally well or even slightly better than LMS (Cheung & Lau, 2017; Su et al., 2019) when there are equal factor loadings across indicators, and our simulation results show that RCSLMS performs equally well in conditions with slightly unequal factor loadings. However, we also find that RCSLMS results in inflated type I error when sample size is small and rXZ is high. In summary, although RCSLMS mostly performs well, it is not appropriate under the following conditions:

(a) When sample size is 100 or smaller.
(b) Response scales for the indicators of a variable are not the same.
(c) The latent variable is not unidimensional, such as in higher-order constructs, in which case LMS should be used because it can model latent interactions with higher-order constructs.
(d) When the same or similar indicators are used for measuring both X and Z, in which case measurement errors for the same indicators are likely to be correlated across X and Z; this can only be modeled in LMS. An example of this is congruence research, which usually uses similar measures for person values (X) and organization values (Z).

These four limitations for RCSLMS also apply to MMR and PI.

The simulation results presented in Table 3 also suggest the need to adopt specific criteria for reliability. Despite accounting for measurement errors, LMS and RCSLMS have lower power when reliability is low. Although the manipulations in our simulations were based on Cronbach's alpha, the values of construct reliability are usually very similar to the values of Cronbach's alpha (Hair, Black, Babin, & Anderson, 2009). Hence, we recommend a minimum level of construct reliability at 0.7 for LMS and a minimum level of Cronbach's alpha at 0.7 for RCSLMS. At this level of reliability, there is still adequate power for testing a moderate interaction effect with a sample size between 200 and 300. This recommended level is also suggested by Hair et al. (2009, p. 619) as denoting good reliability.

Numerical Examples Demonstrating LMS and RCSLMS Implementation

In this section, we demonstrate a 4-step procedure for estimating latent interaction effects with LMS using Mplus 8.4 on simulated data developed from Study 1 in Du, Derks, Bakker, and Lu (2018). A summary of the 4-step procedure, including the decisions and actions in each step, is provided in Fig. 2. Most interaction effects can be tested by estimating one SEM model in each step, giving rise to Model 1, Model 2, Model 3, and Model 4. Hence, for Example A, the models are Model A1 to Model A4. For more complex interactions, a further model is added in Step 3 (Model 3b) to estimate the effect size. Mplus output files for all models, as well as data sets used in the numerical examples, are available in the supplementary files. The Mplus syntax can be found under "Input Instructions" at the beginning of each output file. Key results data are highlighted in each output file. Note that the supplementary files include also a narrated PowerPoint presentation working through the examples presented below.

The original study by Du et al. (2018) examined the moderating effects of homesickness on the relationships between job resources and both task performance and safety behavior using MMR. They also hypothesized the moderating effect of homesickness was moderated by emotional stability and openness to experience. For simplicity of exposition, we demonstrate the model with only task performance as the dependent variable, as well as the interactions among job resources, homesickness, and emotional stability. In the first example, we demonstrate a model testing the two-way interaction between job resources and homesickness (Du et al., 2018, p. 102 – Step 3). Then we illustrate the models for testing three-way interaction (Du et al., 2018, p. 102 – Step 6), followed by testing for quadratic effects and moderated quadratic effects.

The original data included responses from a two-wave survey of 442 employees in a Chinese manufacturing company. Job resources were measured using Karasek's (1985) three-item feedback scale and Van Veldhoven, de Jonge, Broersen, Kompier, and Meijman's (2002) three-item social support scale. Homesickness was measured with the 20-item Utrecht Homesickness Scale (Stroebe, van Vliet, Hewstone, & Wills, 2002). The six-item scale for measuring emotional stability and the six-item scale for measuring openness to experience were both from Judge, Rodell, Klinger, Simon, and Crawford (2013). Task performance was measured at time 2 with the three-item scale from Goodman and Svyantek (1999). We included only one control variable, job demands, measured using the three-item scale from Peterson et al. (1995). Except for emotional stability and openness to experience, which were measured on 7-point response scales, all other variables were measured on 5-point response scales. Interested readers are referred to Du et al. (2018, pp. 99–100) for the conceptual model, sample, and measures. The

Table 3 Simulation 2: type I error/power for testing interaction effects

rXZ	Alpha	N	$f^2=0$			$f^2=.02$			$f^2=.15$			$f^2=.35$		
			MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS
.1	.7	100	.090	.038	.096	.158	.092	.156	.540	.348	.550	.774	.586	.780
		150	.084	.044	.070	.202	.124	.212	.696	.596	.710	.926	.858	.928
		200	.060	.044	.068	.220	.182	.238	.794	.712	.824	.974	.940	.976
		300	.058	.040	.060	.272	.254	.306	.928	.914	.956	.996	.992	.998
		400	.060	.050	.058	.370	.306	.372	.982	.970	.982	1.000	1.000	1.000
		500	.052	.034	.042	.386	.374	.436	.992	.992	.996	1.000	1.000	1.000
	.8	100	.090	.056	.092	.212	.156	.202	.708	.624	.708	.916	.860	.918
		150	.080	.064	.076	.248	.216	.246	.868	.818	.874	.982	.970	.986
		200	.064	.064	.066	.312	.282	.324	.936	.914	.936	.998	.998	.998
		300	.054	.046	.058	.394	.370	.412	.986	.984	.986	1.000	1.000	1.000
		400	.056	.046	.058	.480	.460	.490	.998	.996	.998	1.000	1.000	1.000
		500	.048	.040	.042	.566	.568	.578	1.000	1.000	1.000	1.000	1.000	1.000
	.9	100	.078	.066	.076	.288	.252	.288	.846	.826	.838	.974	.970	.980
		150	.082	.066	.082	.352	.314	.348	.952	.944	.950	.998	.998	.998
		200	.078	.060	.068	.410	.398	.410	.982	.982	.984	1.000	1.000	1.000
		300	.058	.054	.056	.530	.520	.530	.998	.998	.998	1.000	1.000	1.000
		400	.056	.046	.048	.636	.634	.642	1.000	1.000	1.000	1.000	1.000	1.000
		500	.048	.048	.046	.730	.724	.730	1.000	1.000	1.000	1.000	1.000	1.000
.3	.7	100	.078	.040	.090	.174	.108	.174	.546	.368	.548	.768	.598	.770
		150	.082	.048	.066	.218	.150	.220	.712	.594	.724	.928	.844	.932
		200	.072	.052	.078	.238	.218	.258	.800	.728	.826	.970	.942	.970
		300	.064	.048	.070	.314	.300	.344	.936	.922	.946	.994	.992	.996
		400	.060	.058	.070	.420	.380	.434	.980	.974	.986	1.000	1.000	1.000
		500	.056	.042	.052	.444	.444	.468	.994	.994	.996	1.000	1.000	1.000
	.8	100	.088	.066	.098	.226	.188	.234	.698	.620	.710	.904	.862	.910
		150	.086	.062	.078	.266	.240	.286	.850	.824	.874	.984	.970	.994
		200	.076	.068	.076	.332	.308	.328	.928	.920	.936	.998	.996	.998
		300	.064	.054	.062	.440	.410	.446	.982	.984	.990	1.000	1.000	1.000
		400	.058	.054	.060	.540	.524	.554	.994	.996	.996	1.000	1.000	1.000
		500	.046	.048	.054	.606	.608	.612	1.000	1.000	1.000	1.000	1.000	1.000
	.9	100	.094	.080	.096	.290	.266	.280	.826	.816	.826	.976	.964	.970
		150	.078	.074	.080	.366	.352	.358	.950	.942	.952	.998	.996	.998
		200	.086	.080	.088	.456	.438	.452	.984	.982	.986	1.000	1.000	1.000
		300	.054	.046	.050	.574	.568	.576	.998	.998	.998	1.000	1.000	1.000
		400	.064	.060	.068	.698	.690	.696	1.000	1.000	1.000	1.000	1.000	1.000
		500	.054	.050	.050	.782	.778	.782	1.000	1.000	1.000	1.000	1.000	1.000
.5	.7	100	.070	.038	.086	.208	.130	.202	.566	.410	.588	.782	.632	.802
		150	.076	.056	.066	.248	.180	.262	.726	.634	.762	.926	.866	.946
		200	.092	.060	.088	.278	.254	.318	.814	.768	.860	.972	.950	.980
		300	.066	.052	.060	.366	.362	.406	.938	.938	.966	.996	.994	.996
		400	.058	.064	.076	.468	.454	.512	.984	.982	.986	1.000	1.000	1.000
		500	.058	.046	.056	.534	.532	.566	.994	.998	.998	1.000	1.000	1.000
	.8	100	.086	.062	.096	.262	.222	.270	.710	.646	.716	.908	.858	.920
		150	.086	.064	.082	.316	.278	.334	.876	.838	.886	.988	.978	.994
		200	.088	.080	.082	.388	.370	.412	.936	.928	.944	.998	.996	.998
		300	.066	.058	.070	.508	.504	.538	.990	.990	.992	1.000	1.000	1.000
		400	.070	.072	.076	.626	.622	.652	.994	.996	.996	1.000	1.000	1.000

Table 3 (continued)

rXZ	Alpha	N	$f^2 = 0$			$f^2 = .02$			$f^2 = .15$			$f^2 = .35$		
			MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS	MMR	LMS	RCSLMS
		500	.058	.050	.054	.680	.698	.714	1.000	1.000	1.000	1.000	1.000	1.000
	.9	100	.092	.084	.090	.340	.310	.330	.840	.808	.842	.982	.978	.980
		150	.092	.084	.088	.426	.408	.418	.956	.950	.960	.998	.996	.998
		200	.094	.084	.084	.520	.514	.516	.982	.984	.984	1.000	1.000	1.000
		300	.058	.050	.054	.682	.666	.684	.998	.998	.998	1.000	1.000	1.000
		400	.072	.072	.078	.788	.792	.800	1.000	1.000	1.000	1.000	1.000	1.000
		500	.052	.042	.046	.850	.858	.858	1.000	1.000	1.000	1.000	1.000	1.000

MMR moderated multiple regression approach, *LMS* latent moderated structural equations approach, *RCSLMS* reliability-corrected single-indicator latent moderated structural equations approach, f^2 effect size of interaction

data set (Example1.dat in the supplementary files) was simulated based on the reported descriptive statistics, correlations, and results of estimated effects with $N = 422$ (Du et al., 2018, pp. 101–102). LMS works only on raw data but not summary statistics (means and covariance matrix).

Example A: Two-Way Interactions (Models A1, A2, A3, and A4)

Step 1: Measurement Model Estimation (Model 1)

As in all other SEM analyses, the first step is conducting a confirmatory factor analysis (CFA) by estimating the measurement model; this assesses the quality of measures by examining their convergent and discriminant validity. We assessed the fit of the six-factor measurement model (Model A1), in which the 44 observed indicators loaded on six defined factors. Indicators for the dependent variable are likely nonnormally distributed because of the interaction effects; hence, we recommend and used the maximum likelihood with robust standard error (MLR) estimator due to its higher robustness to nonnormality. In the supplementary Mplus files, the indicators for job resources (JOBRES) are denoted as $JobRes_1 - JobRes_6$; those for homesickness (HOMES) are denoted as $Homes_1 - Homes_{20}$; those for emotional stability (EMOSTAB) are denoted as $EmoStab_1 - EmoStab_6$; those for openness (OPEN) are denoted as $Open_1 - Open_6$; and those for task performance (TASKP) are denoted as $TaskP_1 - TaskP_3$. The indicators for the control variable, job demands (JOBD), are denoted as $JobD_1 - JobD_3$. To set the scale for latent variables, the factor loading of the first indicator for each latent variable is set to 1, and the means of all latent variables are set to 0, which centers all latent variables (these are the defaults in Mplus). This centering facilitates the interpretation of the main effects when there is a statistically significant interaction effect (Aguinis & Gottfredson, 2010).[5]

For this empirical example, the measurement model generated $\chi^2 = 946.54$ ($df = 887$). Other fit indices—RMSEA = 0.013, CFI = 0.982, and SRMR = 0.040—indicated that the measurement model fit the data well. Unstandardized factor loadings of all indicators were statistically significant ($p < 0.01$), and standardized factor loadings were all greater than the threshold value of 0.40 suggested by Ford, MacCallum, and Tait (1986). Moreover, the correlation coefficients among all latent variables ranged from -0.382 to 0.463, indicating that they were six distinct variables because the absolute value of all correlation coefficients were significantly lower than 1. Construct reliability of the six variables was estimated using the following formula (Fornell & Larcker, 1981, p. 45):

$$\rho = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum Var(\varepsilon_i)} \tag{5}$$

---

[5] Little, Slegers, and Card (2006) suggested an effects-coding method as an alternative way of identifying and scaling latent variables by fixing the sum of factor loadings across indicators to the number of indicators (for example, sum of factor loadings is fixed at 4 when there are four indicators for the construct), and the sum of indicator intercepts is fixed at 0. While the effects-coding method is equivalent to the marker variable method that fixes the factor loading of a marker variable to one by just changing the scale and intercept of the latent variable, the effects-coding method is particularly useful for interpreting latent means. However, the effects-coding method may not be appropriate for estimating latent interactions since latent variables that form the latent interaction should be centered (such that the mean is zero) to improve interpretability of the estimated regression coefficients. Besides, interpreting unstandardized regression coefficients will be more challenging when the effects-coding method is adopted. Moreover, the effects-coding method requires all indicators to have the same response scale while using the marker item method in LMS does not require this. We also caution against achieving identification by standardizing the latent variables (fixing the variance of latent variables at 1) because the standard errors of estimated parameters may be biased. Given these various issues, we recommend adopting the marker variable method to provide identification when LMS is used.
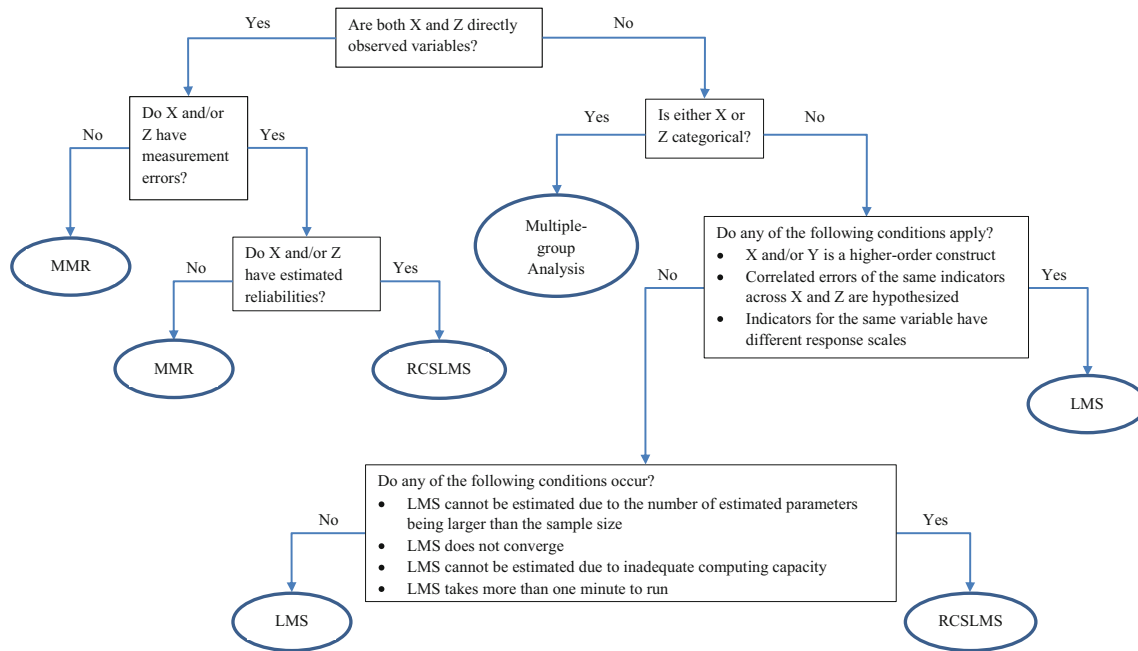
Fig. 1 Recommendations on which approach to use for testing interactions

where ρ is the construct reliability, λ_i is the standardized factor loading, and $Var(\varepsilon_i)$ is the residual variance of indicator i obtained from the Mplus output. As noted above, construct reliability should be at least 0.70; this criterion was met in our numerical example, with values ranging from 0.70 to 0.86. These results provided evidence for adequate reliability, as well as both convergent validity and discriminant validity. If the measurement model does not provide a good fit to the data, one should review the measurement model for any coding errors, for example, higher-order constructs or omitted correlated errors for the same items across time or across sources. If there is insufficient evidence for the appropriateness of the measures, testing for latent interactions should not proceed further with this data set.

Step 2: SEM Without Latent Interaction (Model 2)

The second step is to estimate the structural equation model with only hypothesized linear effects. The objective of this step is to provide baseline model fit indices to allow a subsequent comparative evaluation of overall model fit for the model with latent interactions. More specifically, because the nonlinear effects of a latent interaction model result in biases to the chi-square value and other fit indices derived from the chi-square value (Kelava et al., 2011), Mplus does not provide conventional fit indices for evaluating the overall model fit when estimating an LMS model. Thus, we followed the procedure described in Muthén (2012) to evaluate model fit of an LMS model. First, we estimated a structural model without the latent interaction term (Model A2), that is, we estimated only the main effects of job resources, homesickness, emotional

stability, openness, and job demands on task performance, whereas we excluded from the model the interaction effect between job resources and homesickness on task performance. If Model 2 provides an adequate fit to the data, one may proceed to estimate the structural equation model with the latent interaction (Model A3 in the next step) and examine if inclusion of the latent interaction effect improves model fit significantly.[6] If Model 2 does not provide adequate fit to the data, one may want to revise the model based on alternative theories, noting that modification of model parameters based on modification indices is exploratory and sample specific, and hence requires a separate sample for confirmation. Since we wanted to compare the model fit of Model 2 against the model with the latent interaction using the MLR estimator, that is, Model 3, we estimated both Model 2 and Model 3 with the MLR estimator. In our example, this structural equation model (Model A2) resulted in χ^2 (887) = 946.54, RMSEA = 0.013, CFI = 0.982, and SRMR = 0.040. Hence, we concluded that this structural equation model without the latent interaction fitted the data well. The results showed that only the main effect of job resources ($b = 0.287$, $p < 0.01$, $\beta = 0.323$) on task performance was statistically significant. Significance tests for regression coefficients should be based on unstandardized coefficients because the estimated standard errors for standardized coefficients may be biased (Cudeck, 1989).

---

[6] Although rare, there may be instances where Model 2 does not fit the data adequately and yet Model 3 fits the data significantly better than Model 2. Unfortunately, we will not be able to tell if Model 3 fits the data adequately in such situations.

All five independent variables together accounted for 17.4% of the variance in task performance.

Step 3: LMS Model Estimation (Model 3)

Because the structural equation model without the latent interaction in Step 2 fitted the data well, we proceeded to the next step which was to estimate the 2-way interaction effect between job resources and homesickness on task performance (Model A3). In Mplus, under the ANALYSIS command, the statement "TYPE = RANDOM;" was added to request estimates of the random slopes, and the statement "ALGORITHM = INTEGRATION;" was added to request the numerical integration for the estimation of interaction effects. The latent interaction term was defined under the MODEL command with the statement "JRxHS | JOBRES xwith HOMES;" where *JRxHS* denotes the latent interaction between job resources and homesickness.

Model Fit of the LMS Model We followed the procedure in Muthén ([2012](#)) to evaluate the model fit of the LMS model by comparing the model fit of the models with and without the latent interaction term. Although Sardeshmukh and Vandenberg ([2017](#)) suggested comparing the Akaike information criterion (AIC) between the two models, Klein and Moosbrugger ([2000](#)) and Muthén ([2012](#)) recommended comparing the loglikelihood values because the two models are nested, while AIC is more appropriate for comparing nonnested models. While Maslowsky et al. ([2015](#)) suggested comparing the scaled loglikelihood values directly, Bryant and Satorra ([2012](#)) pointed out that the difference between two scaled loglikelihood values is not chi-square distributed. The proper comparison is a chi-square difference test based on the loglikelihood values and scaling correction factors estimated from the two models both estimated with the MLR estimator (Asparouhov & Muthén, [2013](#), p. 2; Satorra & Bentler, [2010](#)):

$$TRd = 2(L_1 - L_0)(p_1 - p_0)/(c_1 p_1 - c_0 p_0) \qquad (6)$$

where *TRd* is the chi-square distributed difference in scaled loglikelihood ratios with degrees of freedom equals to ($p_1 - p_0$) and, for the model with latent interaction (Model A3), L_1 is the loglikelihood H0 value, c_1 is the H0 scaling correction factor for MLR, and p_1 is the number of free parameters. Similarly, for the model without latent interaction (Model A2) estimated in Step 2, L_0 is the loglikelihood H0 value, c_0 is the H0 scaling correction factor for MLR, and p_0 is the number of free parameters. In our example, $L_1 = -23,453.387$, $c_1 = 0.9767$, $p_1 = 148$ (from output of Model A3), $L_0 = -23,457.537$, $c_0 = 0.9780$, and $p_0 = 147$ (from output of Model A2); therefore, $TRd(df = 1) = 10.565$ ($p =$

0.0012) which indicated that the improvement in model fit through inclusion of the latent interaction effect was statistically significant. An Excel spreadsheet (Loglikelihood Values Difference.xlsx) is provided in the supplementary files to facilitate calculation of the difference in scaled loglikelihood values.

Testing an Interaction Effect The results of Model A3 showed that the regression coefficient of *JRxHS* on *TASKP* (task performance), that is, ω_{12} in Eq. [4](#), was -0.426 ($p < 0.01$, $\beta = -0.177$). Thus, we conclude that the interaction effect was statistically significant and the effect of job resources on task performance depended on the level of homesickness.

Standardized Coefficients and R^2 Standardized coefficients and R^2 for all endogenous variables can be obtained by including "OUTPUT: STDYX;" under the OUTPUT command in the Mplus syntax. Results from Model A3 show that the R^2 for task performance was 0.217, that is, all the independent variables in the model together explained 21.7% of variance in task performance.

Effect Size For a significant interaction, it is valuable to know the effect size of the interaction effect, with Cohen's ([1988](#)) f^2 recommended (Dawson, [2014](#)). The effect size f^2 represents the proportion of residual variance in the dependent variable accounted for by the latent interaction, over and above what is accounted for by the main effects and other control variables. The effect size f^2 of the interaction term can be obtained with the following formula (Cohen, [1988](#), p. 410; Dawson, [2014](#), p. 14):

$$f^2 = \frac{R_2^2 - R_1^2}{1 - R_2^2} \qquad (7)$$

where R_2^2 and R_1^2 are the variance of the dependent variable explained in the models with and without the latent interaction (Model A3 and Model A2), respectively. In our example, $R_2^2 = 0.217$ and $R_1^2 = 0.174$. Hence, the change in R^2 is 0.043, and thus, inclusion of the latent interaction accounted for an additional 4.3% of the variance in task performance with an effect size $f^2 = 0.055$.

Step 4: Interpreting the Interaction Effect (Model 4)

The objective of Step 4 is to interpret the latent interaction effect by conducting simple slope tests and probing the latent interaction effects via two different figures. While this step involves post-hoc analyses conducted only if an interaction effect is statistically significant, researchers are encouraged to specify the form of all interactions a priori in their
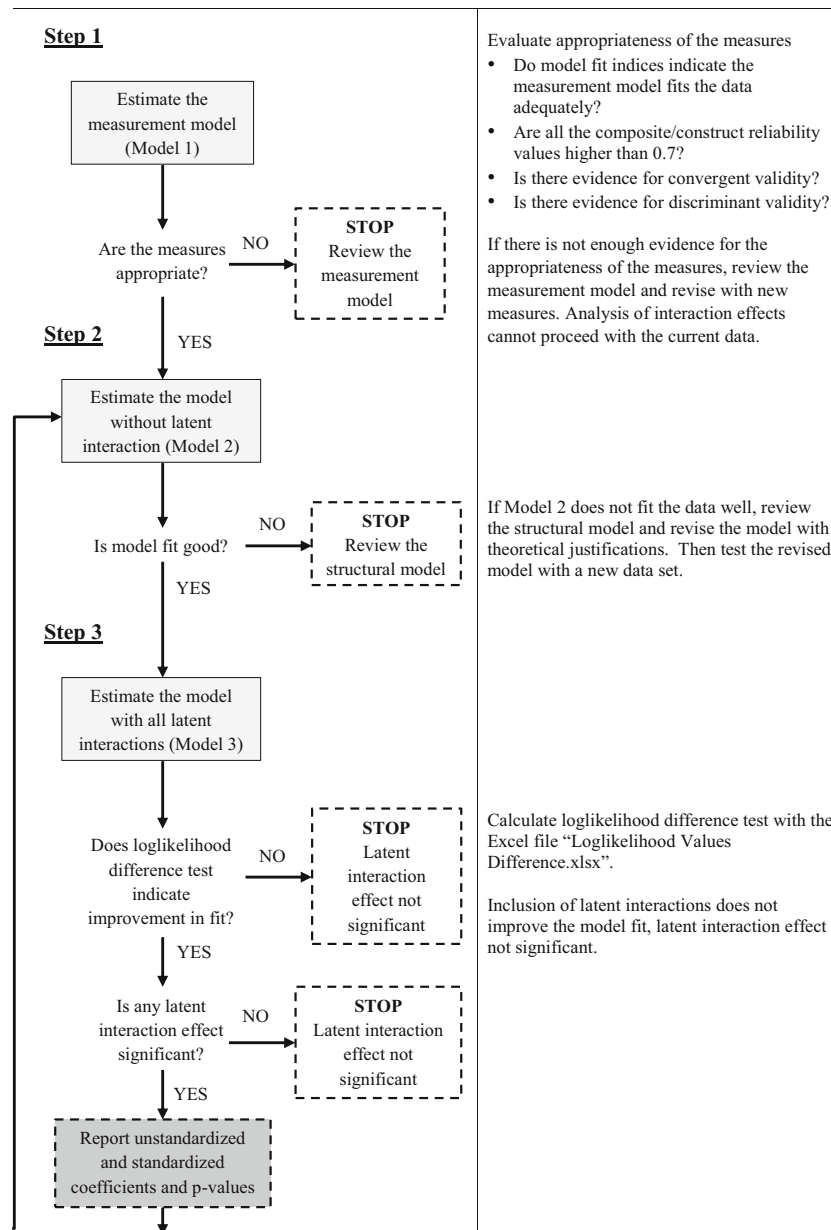
Step 1

Estimate the measurement model (Model 1)

Are the measures appropriate? →NO→ **STOP** Review the measurement model

Evaluate appropriateness of the measures
- Do model fit indices indicate the measurement model fits the data adequately?
- Are all the composite/construct reliability values higher than 0.7?
- Is there evidence for convergent validity?
- Is there evidence for discriminant validity?

If there is not enough evidence for the appropriateness of the measures, review the measurement model and revise with new measures. Analysis of interaction effects cannot proceed with the current data.

Step 2

↓ YES

Estimate the model without latent interaction (Model 2)

Is model fit good? →NO→ **STOP** Review the structural model

If Model 2 does not fit the data well, review the structural model and revise the model with theoretical justifications. Then test the revised model with a new data set.

↓ YES

Step 3

Estimate the model with all latent interactions (Model 3)

Does loglikelihood difference test indicate improvement in fit? →NO→ **STOP** Latent interaction effect not significant

Calculate loglikelihood difference test with the Excel file "Loglikelihood Values Difference.xlsx".

Inclusion of latent interactions does not improve the model fit, latent interaction effect not significant.

↓ YES

Is any latent interaction effect significant? →NO→ **STOP** Latent interaction effect not significant

↓ YES

Report unstandardized and standardized coefficients and p-values

Fig. 2 Procedures for testing latent interactions with LMS and RCSLMS approaches *(continued on next page)*

hypotheses, and these tools then "test" the hypothesized forms of interactions (Dawson, 2014). For example, when hypothesizing that the positive relationship between X and Y is stronger when Z is higher, one can further specify a boundary condition that the positive relationship only exists when Z is higher than 3 (on a 7-point scale). An alternative example is to hypothesize a positive relationship between X and Y when Z is higher than 5 (on a 7-point scale) and a negative relationship when Z is lower than 2. Step 4 requires running Model A4, which is a modified version of Model A3 in Step 3, in which one removes the request for standardized outputs, and adds estimation of simple main effects and a request for bootstrapping the results.

Simple Slope Tests If the latent interaction effect estimated in Step 3 using Model A3 is statistically significant, one can conduct simple slope tests in a new model (Model A4) to examine the statistical significance of the simple main effect of X on Y at various levels of the moderator Z (Aiken & West, 1991). In our example, we examined homesickness moderating the relationship between job resources and task performance. Based on Eq. 4, the relationship between job resources and task performance can be expressed as follows:

$$TASKP = (\gamma_1 - \omega_{12}HOMES)*JOBRES \tag{8}$$

By default in Mplus, the means of latent variables are set to zero and the variances of exogenous variables are estimated.
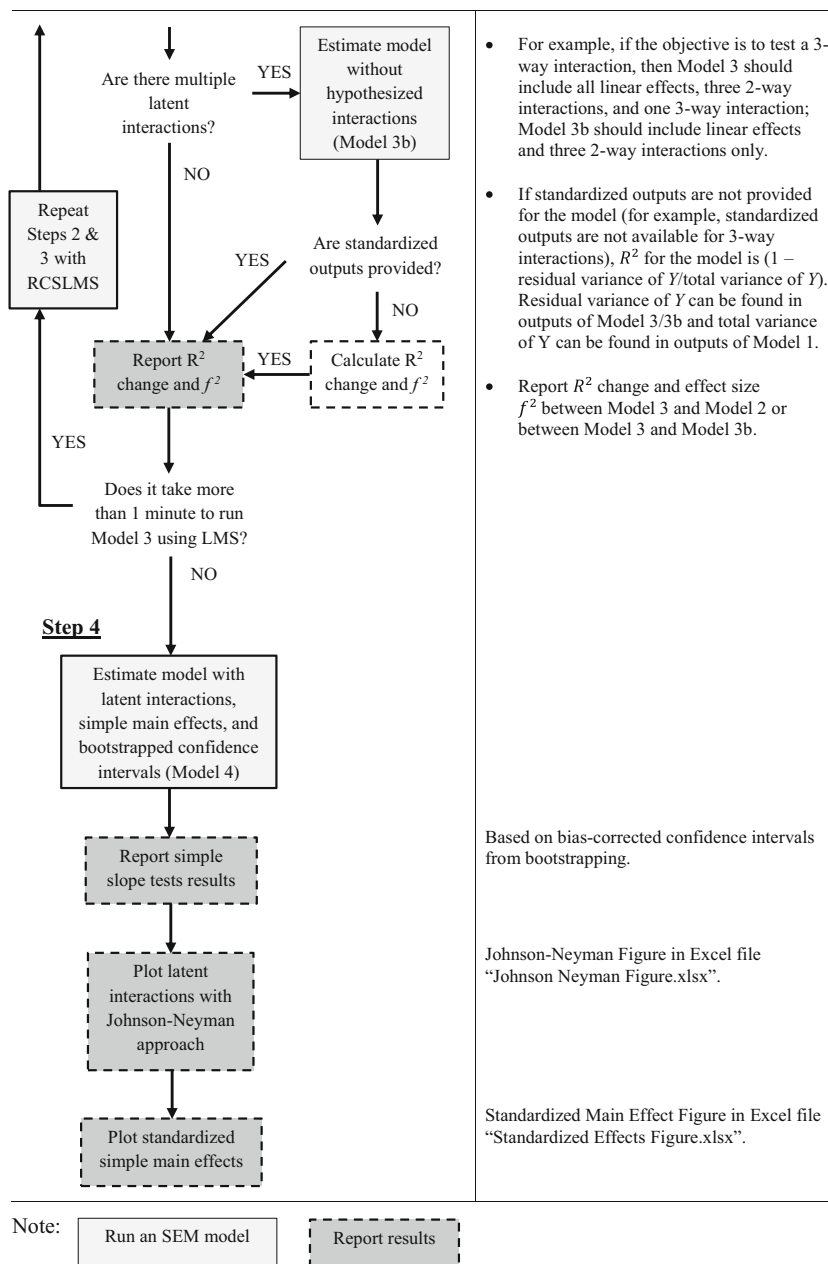
Fig. 2 (continued)

The value of one standard deviation of the moderator, in this case homesickness, can be obtained by first labeling the estimated variance of homesickness (varHS) and then creating a new variable (stdHS) for the standard deviation of homesickness and defining that as the square root of the variance of homesickness using the MODEL CONSTRAINT option in Mplus. For example (and as per Model A4),

HOMES (varHS);

MODEL CONSTRAINT :

NEW (stdHS);

stdHS = SQRT(varHS);

The common method for examining the simple main effects, as suggested by Aiken and West (1991), is to assess these at two levels of the moderator by simply substituting the values of mean plus 1 standard deviation and mean minus 1 standard deviation of homesickness into Eq. 8. Instead, to provide greater detail, we estimate the simple main effects of job resources on task performance at five levels of homesickness as follows: (a) 2 standard deviations below the mean, (b) 1 standard deviation below the mean, (c) at the mean, (d) 1 standard deviation above the mean, and (e) 2 standard

deviations above the mean.[7] Estimating the conditional main effects at five or more levels enables probing of the interaction effects in a figure using the Johnson-Neyman approach as described below. In order to substitute different levels of homesickness into Eq. 8, in the Mplus syntax (Model A4), we first labeled γ_1 and ω_{12} in Eq. 8 as b1 and b3, and then created new variables by using the MODEL CONSTRAINT option. For example:

TASKP ON JOBRES (b1);

TASKP ON JRxHS (b3);

MODEL CONSTRAINT :

NEW (Slope_2L Slope_L Slope_M Slope_H Slope_2H);

Slope_2L = b1 + b3*(−2*stdHS);

Slope_L = b1 + b3*(−1*stdHS);

Slope_M = b1 + b3*(0*stdHS);

Slope_H = b1 + b3*(1*stdHS);

Slope_2H = b1 + b3*(2*stdHS);

Then a statistical test was conducted for the simple main effect of job resources on task performance at each specified level of homesickness by estimating the CI of each simple main effect using bootstrapping. Since Mplus does not allow for bootstrapping with the MLR estimator, the ML estimator is used instead. Note that both the MLR and ML estimators give the same estimated parameters, and because bootstrapping is used to estimate the standard errors of the estimated parameters, there is no need to adjust the estimated standard errors with the MLR estimator. The ML estimator is specified by including "ESTIMATOR = ML;" under the ANALYSIS command. We suggest generating 2000 bootstrap samples by including "BOOTSTRAP = 2000;" under the ANALYSIS command. Bias-corrected confidence intervals (BCCI) are obtained by including "OUTPUT:

CINTERVAL(BCBOOTSTRAP);" under the OUTPUT command. The results from the output file of Model A4 show that when homesickness was one standard deviation below the mean, the effect of job resources on task performance was statistically significant ($b = 0.463$; $p < 0.01$; 95% CI [0.257, 0.771]). When homesickness was one standard deviation above the mean, the effect of job resources on task performance was not statistically significant ($b = 0.144$; $p > 0.10$; 95% CI [-0.044, 0.348]).

Plotting the Interaction Effect with the Johnson-Neyman Technique For a moderator which is a continuous variable, such as homesickness, the moderating effect can be displayed graphically using the Johnson-Neyman technique (Johnson & Neyman, 1936) which plots the unstandardized simple main effects (Y-axis) at various levels of the moderator (X-axis). We provide an Excel file in the supplementary files (Johnson Neyman Figure.xlsx), which plots the unstandardized effects of job resources on task performance at various levels of homesickness as per Fig. 3. The entries in the highlighted cells were the estimated standard deviation of homesickness (moderator), the estimated simple main effects, and the corresponding 95% BCCI that can be found under "New/Additional Parameters" of the Mplus output file of Model A4. This Excel file can be adapted by researchers using their own Mplus output values. Draw a horizontal line at the value of zero on the Y-axis; note where this line intercepts with the CI (either the lower limit or the upper limit) and draw a vertical line down to the X-axis. This indicates the boundary between significant and nonsignificant regions. The region where the horizontal line is outside the CI is significant, and the region where the horizontal line is within the CI is nonsignificant. Figure 3 reveals that the higher the level of homesickness, the lower was the positive effect of job resources on task performance. Figure 3 also shows that we have identified a boundary condition for the effect of job resources on task performance. That is, job resources only had a statistically significant positive effect on task performance when the level of homesickness was lower than 0.3 standard deviations above the mean. However, one should interpret the level of the moderator that defines the boundary conditions cautiously because this is sample specific and sample size will affect the width of the CIs and hence the boundary conditions.

Plotting the Standardized Effects at Various Levels of Moderator While the graph using the Johnson-Neyman technique is superior (Fig. 3) because it provides substantially more information, reviewers may still request the standard graph plotting the simple main effects of the moderator at one standard deviation above and below the mean. Hence, we also plotted the standardized effects of job resources on task performance at homesickness values of one standard deviation above and below the mean to illustrate the nature of the

---

[7] However, these five levels are arbitrarily selected and if other levels (at least five levels for a continuous variable because the confidence intervals are not linear) of the moderator can be meaningfully identified, these more specific levels should be used instead. An example is when the moderator is a directly observed variable measured on a 5-point scale, then using 1, 2, 3, 4, and 5 as the five levels may be more meaningful. Note that since the moderator is mean-centered when converted into a latent variable before estimating the latent interaction effects, each level of the moderator in a simple slope test should be mean-centered as well. For example, suppose the mean of the moderator Z is 3.25 on a 5-point scale, this mean should be subtracted from each of the five levels of the moderator and the simple slope of the relationship between X and Y at Z equals 1 can be defined as the following under MODEL CONSTRAINT:

Slope_2L = b1 + b3*(−2.25); !1−3.25 = −2.25.

interaction effect (Fig. 4). The standardized coefficients can be obtained from the output of Model A3. The results show that the standardized form of Eq. 8 can be expressed as follows:

$$TASKP^* = (0.336 - 0.177*HOMES^*)*JOBRES^* \qquad (9)$$

where $TASKP^*$, $HOMES^*$, and $JOBRES^*$ are the standardized values of $TASKP$, $HOMES$, and $JOBRES$, respectively. We plotted two lines in Fig. 4. The first line was plotted by substituting $HOMES^*$ with 1 and $JOBRES^*$ with values from −3 to 3, and the second line was plotted by substituting $HOMES^*$ with −1 and $JOBRES^*$ with values from −3 to 3. To facilitate this, we provide an Excel file (Standardized Effects Figure.xlsx) in the supplementary files that plots the standardized effects in Fig. 4. Researchers can adapt this Excel file by inputting the values from their Mplus output into the relevant highlighted boxes. Note that this graph is more meaningful if the moderator is a dichotomous variable. If the moderator is a continuous variable, one should interpret this graph cautiously because the two levels of moderator chosen may not meaningfully represent the sample (Dawson, 2014).

We next present Examples B to G covering a range of other interactions researchers may wish to investigate.

Example B: Three-Way Interactions

(Models B3, B3b, and B4) Du et al. (2018) also hypothesized that the moderating effect of homesickness was moderated by emotional stability. The test of this hypothesis involves three two-way interactions and one three-way interaction. For this example, the measurement model in Step 1 (Model A1), and the model with only linear effects in Step 2 (Model A2) are the same as those in Example A described above; therefore, we do not repeat these first two steps.

In Step 3, we added to Model A2 the three-way interaction among job resources, homesickness, and emotional stability, as well as the three two-way interactions, to produce Model B3. Since Mplus 8.4 does not provide standardized outputs for three-way interactions yet, the request for standardized outputs on the OUTPUT command was removed. Comparing the loglikelihood values between Model A2 ($L_0 = -23,457.537$, $c_0 = 0.9780$, and $p_0 = 147$) and Model B3 ($L_1 = -23,450.593$, $c_1 = 0.9739$, and $p_1 = 151$) resulted in $TRd(df = 4) = 16.870$ ($p < 0.01$), which indicated that adding the three-way interaction and three two-way interactions improved the model fit significantly. Results from the output of Model B3 show that the effect of the three-way interaction among job resources, homesickness, and emotional stability on task performance was statistically significant ($b = -0.430$, $p = 0.02$). Following Muthén (2012, p. 7), the standardized coefficient can be estimated by dividing the unstandardized coefficient by the standard deviation of the dependent variable and then

multiplying it by the standard deviations of all the components of the interaction term. The standard deviations of the latent variables can be obtained from the measurement model in Model A1 and the unstandardized coefficient from Model B3. Thus, the standardized effect of the three-way interaction on task performance equaled to $\frac{-0.430*\sqrt{0.160}*\sqrt{0.144}*\sqrt{0.765}}{\sqrt{0.126}} = -0.161$.

Since the R^2 for the three-way interaction model is not provided by Mplus, it is estimated by ($1 -$ residual variance of Y/total variance of Y) where residual variance of Y is obtained from Model B3 and total variance of Y from Model A1. Hence, the R^2 for task performance in Model B3 is estimated as $(1 - \frac{0.094}{0.126}) = 0.254$. In order to estimate the R^2 change and effect size Cohen's f^2 attributable to the three-way interaction, an additional model needs estimating without the three-way interaction effect. This model (Model B3b) has all the linear effects and three two-way interactions; the R^2 for task performance in this model is 0.220. The R^2 change in task performance attributable to the three-way interaction is $(0.254 - 0.220) = 0.034$ and Cohen's $f^2 = \left(\frac{0.034}{1-0.254}\right) = 0.046$.

To conduct Step 4, and following Dawson (2014, p. 5), four different moderating effects with different combinations of high and low levels of homesickness and emotional stability on the job resources–task performance relationship were estimated for the simple slope tests for the three-way interaction. The three-way interaction also requires six pairwise comparisons for differences between slopes. Although Dawson (2014, pp. 5–7) provided the standard error for the simple slope tests and the comparisons between slopes, formulae for those standard errors are complicated and the test statistics likely have nonnormal distributions. Hence, we suggest using bootstrapping to conduct the simple slope test and the comparisons between slopes. Model B4 in the supplementary files shows the Mplus syntax for conducting simple slope tests with bootstrapping using LMS. As it took about 5 minutes to run Model B3, we ran Model B4 with just five bootstrap samples (instead of the usual 2000 samples) to estimate the elapsed time. It took 17.25 minutes to run Model B4 and hence it is estimated that running Model B4 with 2000 bootstrap samples may take up to 4 days. Where researchers find this too long to run a model, we recommend RCSLMS as an alternative approach to testing interaction effects and illustrate this in Example C below.

Example C: Three-Way Interaction with RCSLMS (Models C2, C3, C3b, and C4)

Since RCSLMS has much reduced computational demands compared to LMS, RCSLMS is recommended for testing more complex interactions when bootstrapping is required to conduct the simple slope test and the comparisons between slopes. We next demonstrate how to use RCSLMS to test for a three-way interaction. Identical to Step 1 in LMS, Step 1 in RCSLMS is to conduct a CFA by

estimating the measurement model with all items to assess measurement quality. That was completed in Model A1 in our numerical example. We estimated all subsequent models in Example C with the Mplus demo version.

Step 2 of the RCSLMS approach was to create a new data set to represent each variable using the simple average of item scores (Example2.dat in the supplementary files). Then the variance and Cronbach's alpha for each variable were estimated with R. An initial model with only linear effects (Model C2) was estimated to provide baseline fit indices against which to evaluate the fit of models with latent interactions. Thus, Model C2 was created by replacing each latent variable in Model A2 with the simple average score as a single indicator, fixing the factor loading to 1, and the residual variance to $(1 -$ reliability$)$ times the variance of the variable. For example, the simple average of the indicators for job resources was labeled as SJobRes, which had a variance at 0.289 and a reliability at 0.735. Hence, JOBRES was defined as:

JOBRES BY SJobRes@1; SJobRes@0.076585;

Since the variables are modeled as latent variables with single indicators in RCSLMS, the variables will be centered by default in Mplus. All the fit indices for Model C2 indicated the model fitted the data perfectly. However, since the fit indices are not sensitive to missing nonlinear effects (Mooijaart & Satorra, 2009), we examine if the inclusion of the latent interactions can improve model fit using the loglikelihood values difference test.

Step 3 comprises estimating a model (Model C3) in which the three two-way latent interactions and the three-way interaction are added to Model C2, which parallels the LMS process via Model B3. Comparison of the loglikelihood values between Model C3 and Model C2 resulted in $TRd(df = 4) = 18.514$ $(p < 0.001)$, which indicated that adding the three two-way interactions and the three-way interaction improved the model fit significantly. Results of Model C3 were comparable with those of Model B3. The effects of the interaction between job resources and homesickness, and the three-way interaction among job resources, homesickness, and emotional stability on task performance, were -0.453 $(p = 0.001)$ and -0.436 $(p = 0.011)$, respectively, from Model C3 and were -0.446 $(p = 0.008)$ and -0.430 $(p = 0.020)$, respectively from Model B3. Latent variance of a variable in RCSLMS equals to observed variance minus residual variance, which were all obtained in Step 2. Thus, the standardized three-way interaction effect on task performance equaled to $\frac{-0.436*\sqrt{0.212}*\sqrt{0.148}*\sqrt{0.710}}{\sqrt{0.161}} = -0.162$.

The residual variance for task performance in Model C3 was 0.121. Hence, the R^2 for task performance in Model C3 is

estimated as $(1 - \frac{0.121}{0.161}) = 0.251$. As before, estimation of the additional model (Model C3b) with only the linear effects and three two-way interactions, but not the three-way interaction effect, was required to find the R^2 change and effect size Cohen's f^2 attributable to the three-way interaction. R^2 for task performance in Model C3b was 0.223. The R^2 change in task performance brought by the three-way interaction is $(0.251 - 0.223) = 0.028$ and Cohen's $f^2 = \left(\frac{0.028}{1-0.251}\right) = 0.037$.

Step 4 required conducting simple slope tests of the three-way interaction and comparisons between slopes with Model C4, which added the bootstrapping function to Model C3 with 2000 bootstrap samples to estimate the CIs. Whereas Model C3 required 47 seconds to run, Model C4 took about 26 hours to run. Results show that the relationships between job resources and task performance were not statistically significant when both homesickness and emotional stability were high $(b = -0.116, p > 0.10, 95\%$ BCCI $= [-0.435, 0.150])$ and when both homesickness and emotional stability were low $(b = 0.283, p > 0.05, 95\%$ BCCI $= [-0.050, 0.613])$. However, the relationships between job resources and task performance were statistically significant when homesickness was high and emotional stability was low $(b = 0.218, p < 0.05, 95\%$ BCCI $= [0.008, 0.422])$, and when homesickness was low and emotional stability was high $(b = 0.515, p < 0.01, 95\%$ BCCI $= [0.262, 0.780])$. Based on Dawson (2014), Jeremy Dawson has provided Excel files on his personal web page for probing two-way and three-way interactions (www. jeremydawson.com/slopes.htm), which are also applicable for results obtained from RCSLMS.

Other Models with More Complex Interactions

In this section, analytical procedures for models with more complex interactions are presented based on LMS. All analyses should go through the 4-step procedure described above. Note that Step1 (measurement model), Step 2 (model with linear effects only), and Step 4 (bootstrapping and probing with figures) remain the same when testing various nonlinear effects; only the models for Step 3 vary and therefore will be briefly discussed, with Examples D to G below representing a range of interactions that researchers may want to investigate. The tests for interactions should be based on theoretically-derived hypotheses and here we posit possible interactions for the sake of demonstration only.

Example D: Quadratic Effect (Model D3)

Besides moderating effects, U-shaped or inverted U-shaped relationships representing quadratic effects are also commonly examined in business and psychology studies (Dawson, 2014,

Equation 7). LMS can be used to examine the quadratic effect of a latent variable by simply creating a latent interaction term with the latent variable itself. For example, a latent quadratic term for job resources was defined in Model D3 under the MODEL command with the statement "JRxJR | JOBRES xwith JOBRES;". Comparing the loglikelihood values between Model D3 and Model A2 gave $TRd(df = 1) = 6.985$ ($p = 0.0082$), indicating that adding the latent quadratic effect made a statistically significant improvement to model fit. The results showed that the regression coefficient of $JRxJR$ on $TASKP$ (task performance), that is ω_{11}, was -0.286 ($p < 0.05$, $\beta = -0.124$), and $R_2^2 = 0.212$. Comparing these with results of Model A2 showed the latent quadratic effect of job resources accounted for an additional 3.8% of the variance in task performance with an effect size of $f^2 = 0.048$.

Example E: Moderated Quadratic Effect (Model E3)

The quadratic effect of job resources on task performance might be moderated by homesickness. Thus, both the latent interaction between job resources and homesickness, and that between the quadratic term of job resources and homesickness, were added to Model D3 to produce Model E3 (Dawson, 2014, Equation 8). Comparing the loglikelihood value of Model E3 with that of Model A2 gave $TRd(df = 3) = 11.584$ ($p < 0.01$), suggesting that this model fits the data significantly better than the model without latent interaction. However, the moderating effect of homesickness on the curvilinear relationship between job resources and task performance was -0.034 ($p = 0.903$), which is not statistically significant. The R^2 for task performance in Model E3 is estimated as $(1 - \frac{0.099}{0.126}) = 0.214$. Calculation of the R^2 change in task performance brought by the moderated quadratic effect required estimation of Model E3b that dropped the moderated quadratic effect from Model E3. The R^2 for task performance in Model E3b is also 0.214, the same as Model E3, indicating

that including the moderated quadratic effect explained no additional variance in task performance.

Example F: Polynomial Regression (Model F3)

A simple extension of the quadratic effect is the polynomial regression model shown in Eq. 1. Su et al. (2019) recently used simulation to show that using LMS and RCSLMS to model polynomial regression when examining congruence is superior to the regression approach that ignores measurement errors. In this polynomial regression model (Model F3), the latent quadratic term of job resources, latent quadratic term of homesickness, and the latent interaction between job resources and homesickness were added to Model A2. Comparing the loglikelihood value of Model F3 with that of Model A2 gave $TRd(df = 3) = 13.476$ ($p = 0.0037$), indicating that the inclusion of the latent quadratic terms and the latent interaction improved the model fit significantly. However, the results of Model F3 showed that none of these additional terms had statistically significant effects on task performance. We note a debate between researchers about including quadratic terms. Some researchers (e.g., Cortina, 1993; Edwards, 2009) suggest that whenever an interaction effect is estimated, one should also include the quadratic terms in the model such that the estimated interaction effect will not include spurious effects from omitted quadratic terms. Yet other researchers (e.g., Dawson, 2014) suggest that inclusion of the quadratic terms is only essential when the correlation between X and Z is above 0.5 to provide a more conservative test for the interaction effect. However, Su et al. (2019) found in their simulation that when quadratic effects did not exist in the population, inclusion of the quadratic terms in the model resulted in higher type I error for the quadratic effects and lower power for testing the interaction effects; this was due to attenuation in the estimates of the interaction effects. Hence, joining with other researchers (e.g., Aiken & West, 1991; Cheung, 2009; Harring, Weiss, & Li, 2015; Su et al., 2019), we recommend that the decision as to whether quadratic effects should be
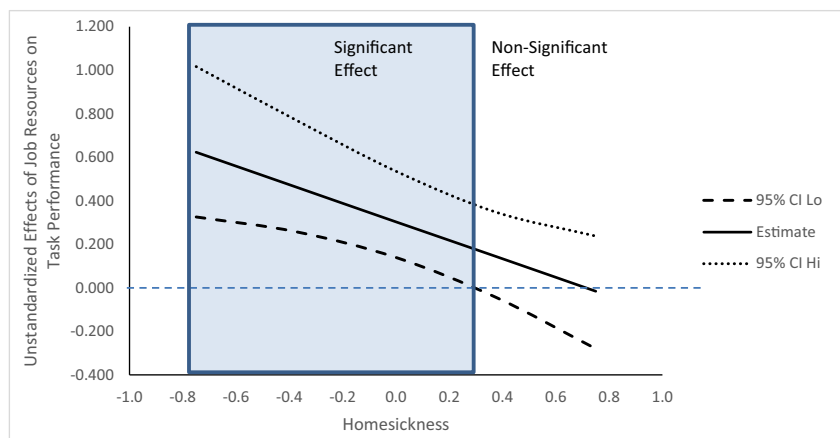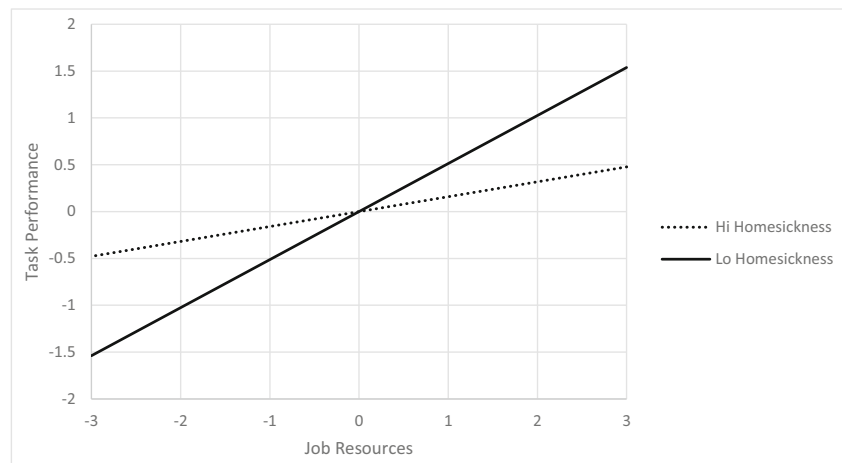

Fig. 3 Unstandardized effects of job resources on task performance conditional on homesickness

included or excluded in a model testing interaction effects should be based on theory and related hypotheses.

Example G: Curvilinear Three-Way Interaction (Model G3)

Finally, the moderating effect of homesickness on the curvilinear (quadratic effect) relationship between job resources and task performance might also be moderated by emotional stability, leading to a curvilinear three-way interaction (Dawson, 2014, Equation 13). This final model, Model G3, combined the three-way interaction and the quadratic effect of job resources, and can be created by inclusion of a second moderator W to the moderated quadratic effect model (Model E3) discussed above. Specifically, the additional linear effect of W, two-way interactions of XW and ZW, three-way interaction of XZW, and the moderated quadratic effects of X^2W and X^2ZW were added to Model E3. Comparing the loglikelihood value of Model G3 with that of Model A2 gave $TRd(df = 8) = 19.758$ ($p < 0.05$), indicating that the model fits the data significantly better than the model with only linear effects. Results of Model G3 showed that the linear three-way interaction ($b = -0.402$, $p < 0.05$) had statistically significant effects on task performance. The R^2 for task performance in Model G3 is estimated as $(1 - \frac{0.093}{0.126}) = 0.262$; comparison of the R^2 obtained in Model E3 (which serves as Model G3b) gives an R^2 change at 0.048 and Cohen's $f^2 = \left(\frac{0.048}{1 - 0.262}\right) = 0.065$.

Some Frequently Asked Questions About Modeling Interactions with LMS and RCSLMS

In this final section, we provide answers to questions that we suspect might otherwise puzzle researchers and restrict their ability to implement LMS and RCSLMS appropriately.

What Sample Size Is Required for LMS and RCSLMS to Test Latent Interactions?

The appropriate sample size to test interactions using LMS or RCSLMS depends on the complexity of the model and effect size being tested. At a minimum, the sample size must be larger than the number of estimated parameters; otherwise, the model will be unidentified and no estimated parameters will be provided. Mplus will provide an error message under such circumstances. In this regard, RCSLMS requires a smaller sample size than LMS because fewer parameters are estimated. Simulation results in Table 3 show that, for both LMS and RCSLMS, $N = 400$ is required to achieve adequate power for testing a small interaction effect and $N = 150$ is required for testing a medium interaction effect. Although in general $N = 100$ provides adequate power for testing a large interaction effect for both LMS and RCSLMS, in particular when α is 0.8 or higher, our simulation results also show that using RCSLMS with a small sample size yields inflated type I error rates. Hence, our overall recommendation is to use a minimum sample size of 150 when testing latent interactions.

How Can the Required Computing Resources Be Reduced?

A common criticism of LMS is that it takes a long time to estimate the model parameters (e.g., Kelava et al., 2011; Klein & Moosbrugger, 2000; Sardeshmukh & Vandenberg, 2017). With ongoing improvements in computing speed, this concern should gradually diminish. Among the models we demonstrated, Model G3 is the most complex, containing eight quadratic and interaction terms; the elapsed time for that model was about 16 minutes on a desktop computer with an Intel Core i7-6700T CPU @ 2.80GHz and 16GB RAM. While elapsed time may be acceptable for a single model, combining bootstrapping and complex models with LMS may take a

much longer time to run. There are a few tricks one can use to shorten the elapsed time. Our first suggestion is to increase the number of processors for Mplus by specifying "PROCESSORS = 8;" under the ANALYSIS command if the computer has eight core processors. While most modern computers have multiple processors (normally eight), Mplus uses only one processor by default. The above syntax instructs Mplus to use multiple processors, shortening the elapsed time by more than 50%.

The second trick is relevant to highly complex models; for these, use RCSLMS to estimate the parameters instead of a full model with all indicators. Multiple simulation studies (Cheung & Lau, 2017; Su et al., 2019), including those in this teaching note, find that RCSLMS provides similar results to those from LMS. Further, unlike PI, RCSLMS does not have the problem of item parceling that gives different results depending on how the items are combined into parcels.

Finally, LMS adopts numerical integration to approximate the mixture of multivariate normal distributions. The most commonly encountered warning and error messages when running LMS with Mplus are "this model requires a large amount of memory and disk space. It may need a substantial amount of time to complete" and "there is not enough memory space to run Mplus on the current input file." The higher the number of integration points, the longer elapsed time it will take and the more accurate the results are. While the Mplus default number of integration points is 15, one may try to reduce the elapsed time by using a smaller number of integration points by specifying, for example, "INTEGRATION = 8;" under the ANALYSIS command. Researchers can then gradually increase the number of integration points, although usually the results show minimal change at 8 or more integration points. However, we do not recommend researchers use small numbers of integration points, such as 4 as suggested by Preacher, Zhang, and Zyphur (2016), as doing so will usually lead to nonconvergent solutions or inaccurate estimates.

What Should I Do with Missing Data?

It is not uncommon to have missing values in the data. An important first step is to identify the nature of the missingness. There are three broad categories of missingness, missing completely at random (MCAR), missing at random (MAR, where missingness is a function of observed variables), and missing not at random (MNAR; Rubin, 1976). One can examine if the missing values are due to MCAR by using Little's MCAR test (Little, 1988) with SPSS or R.[8] Analysis can proceed if data are MCAR or MAR, and the total amount of missing values is less than 10%. When missing values are

defined, Mplus by default uses full information maximum likelihood (FIML) estimation, which is based on the loglikelihood value of individual observations instead of the sample covariance matrix. Interested readers are referred to Enders (2001) and Newman (2014). Missing values can be defined using the MISSING option under the VARIABLE command. For example, if the number 99 is used to represent missing values for the four items x1 to x4 for measuring X, then the following option can be included to define the missing value:

MISSING ARE x1−x4 (99);

When adopting the RCSLMS approach, average scale scores should be calculated based on all available values for a variable. If there are values for three out of four indicators for a variable, then the three values should be averaged to represent the value of the variable. If a variable has all indicator values missing, then the value of that variable will be missing. When data are MNAR or the total amount of missing values is more than 10%, one or more auxiliary variables can be used to convert MNAR to MAR (Graham, 2003; Newman, 2014). An auxiliary variable is a variable that correlates with the substantive variable that has missing data (Collins, Schafer, & Kam, 2001). This can be done by using the AUXILIARY option under the VARIABLE command. For example, if aux1 and aux2 are used as auxiliary variables, then add the following line in the Mplus syntax:

VARIABLE : AUXILIARY = (m) aux1 aux2;

How Should I Test for Interaction Effects When There Are Multiple Independent Variables or Multiple Moderators?

A research design may have multiple independent variables, such as X_1 and X_2, in a study all being moderated by a moderator Z. In this case, all interaction effects should be tested simultaneously in one model. That is,

$$Y = \alpha + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 Z + \omega_{13} X_1 Z + \omega_{23} X_2 Z \quad (10)$$

If one instead specifies multiple models, each including only one moderating effect, the estimated interaction effects may be biased due to the omitted terms in the equations. Similarly, if there is one independent variable X and multiple moderators Z_1 and Z_2, all the interactions should be tested simultaneously in one model such as the following:

$$Y = \alpha + \gamma_1 X + \gamma_2 Z_1 + \gamma_3 Z_2 + \omega_{12} X Z_1 + \omega_{13} X Z_2 \quad (11)$$

An additional consideration, suggested by Dawson (2014, p. 15), is that when multiple moderators are involved, it is probable that the moderators will interact with each other as

---

[8] Little's MCAR test is also available in the R package – BaylorEdPsych by A Alexander Beaujean at https://www.rdocumentation.org/packages/BaylorEdPsych/versions/0.5.

well. If one suspects that the two moderators also interact with each other, then a three-way interaction model is required for the analysis, which can be tested using Model B3 (using LMS) or Model C3 (using RCSLMS) as described above.

What Should I Do with Control Variables?

Many business and psychological studies include one or more control variables. These control variables are treated as exogenous variables in SEM and should be included in all models in the 4-step procedure. In order to ascertain the contributions of the linear effects and interaction effect of X and Z in explaining the variance of Y beyond the control variables, an additional model that includes only the control variables as exogenous variables (Model 2a) can be estimated before estimating Model 2. The R^2 change between Model 2a with only control variables and Model 2 will show the additional proportion of variance in Y that can be explained by X and Z beyond the control variables, and the R^2 change between Model 2 and Model 3 will show the additional proportion of variance in Y that can be explained by XZ. Readers are referred to Becker (2005) and Bernerth and Aguinis (2016) for recommendations on the appropriate use and selection of control variables, and then reporting and interpreting results of analyses with control variables.

Do LMS and RCSLMS Work with Nested Data and Multilevel Models?

All examples in this teaching note assume that all data are at one level; yet quite frequently, data in business and psychological studies collected at a lower level are nested within higher level units. The examples provided here can easily be tailored for nested data by specifying the grouping variable by including "CLUSTER = grouping variable name;" under the VARIABLE command and "TYPE = COMPLEX RANDOM;" under the ANALYSIS command so that the standard errors of the estimated parameters are adjusted for nonindependence. When latent interactions are hypothesized at multiple levels, the same procedure described in this teaching note applies. Cross-level moderations are not discussed in this teaching note because normally no latent interaction term is required. Recently, Preacher et al. (2016) demonstrated an Mplus syntax using latent interactions to examine cross-level moderations. Nevertheless, modeling latent interaction effects at multiple levels requires extensive computing resources and hence the models do not always converge. Readers should note that bootstrapping does not normally work for nested data or multilevel data because the number of level 1 units within each level 2 unit cannot be controlled. Currently, Mplus 8.3 and 8.4 allow for bootstrapping with nested data, with the bootstrap procedure implemented at level 2 units (such as teams). Since sampling is conducted at level 1 units

(such as individuals) in many business and psychological studies, this approach is not recommended. Instead, for multilevel models, we recommend using the Monte Carlo simulation approach within Mplus to find the CI of estimated parameters. Interested readers are referred to Preacher and Selig (2012).

Do LMS and RCSLMS Work with Moderated-Mediation Models?

Although the discussions and examples in this teaching note are about moderation and curvilinear relationships, the Mplus syntax provided in the supplementary files can be easily extended to estimating moderated-mediation effects by multiplying the moderating effects with the additional path coefficients using the MODEL CONSTRAINT option. Cheung and Lau (2017) have demonstrated the procedure for testing moderated mediation with LMS using Mplus, while Sardeshmukh and Vandenberg (2017) have provided Mplus syntax for estimating six frequently examined moderated-mediation models with LMS. Although they commented that LMS may not be applicable to complex moderation and mediation models (Sardeshmukh & Vandenberg 2017, p. 742), our examples reveal that LMS with Mplus can handle complex models. The Mplus syntax and output files for models in the four-step procedure using both LMS and RCSLMS, based on the example in Cheung and Lau (2017), and the data file Example3.dat are also provided in the supplementary files (Models H and I respectively). The Excel file for plotting the Johnson-Neyman graph can be used directly by relabeling the Y-axis to "indirect effect of X on Y through M," instead of the values of Y. The Excel file for plotting the standardized effects can also be adopted to plot the conditional standardized indirect effects by substituting the main effect in the file with the standardized indirect effect (a1*b in the sample output for Model 3) and substituting the interaction effect with the product of the interaction effect and the mediating path (a3*b in the sample output for Model 3).

Conclusion

This teaching note contributes to the literature testing latent interactions in several ways. First, we started our teaching note with a brief explanation highlighting how LMS avoids the complexities of PI in constructing a latent interaction term, with LMS estimating latent interaction effects from their underlying latent variables by matrix multiplications. This removes the requirement to create products of indicators to "measure" the latent interaction and therefore simultaneously avoids the need to derive complicated nonlinear constraints. Second, we presented simulation results to demonstrate how LMS and RCSLMS outperform MMR based on a general

latent interaction model, instead of more specific models such as the moderated-mediation model (Cheung & Lau, 2017) and the polynomial regression model (Su et al., 2019). Third, we went beyond previous simulation studies to demonstrate that RCSLMS is robust for testing latent interactions of constructs with unequal factor loadings. Fourth, we demonstrated comprehensive processes with numerical examples for utilizing LMS to test for a range of latent interaction effects and RCSLMS to test for three-way latent interactions. The recommended 4-step procedure resolves many previously-raised problems in implementing LMS, such as evaluating model fit for models with latent interactions, estimating standardized coefficients, and conducting simple slope tests. The 4-step procedure starts by evaluating measurement quality using CFA in Step 1; establishing baseline fit indices with the model without latent interactions in Step 2; testing the latent interaction effects, and estimating standardized coefficients and the effect size in Step 3; and finally, conducting simple slope tests and probing the interaction effects using figures in Step 4. We have provided separate Excel files that facilitate the calculation of loglikelihood values difference for estimation of model fit and produce two types of figures for probing interaction effects. We also demonstrated how simple slope tests can be conducted using bootstrapping, instead of relying on a complicated formula to estimate the standard errors of the slopes. Finally, we provided sample Mplus syntax and brief explanations for estimating more complex models, including those with three-way interactions, quadratic effects, moderated quadratic effects, moderated three-way interactions, and moderated mediation. A narrated PowerPoint file is provided with the supplementary materials to further support implementation of LMS and RCSLMS.

Limitations and Future Directions The simulations reported in this teaching note used simulated data based on normal distributions for both the latent variables and indicators. Klein and Moosbrugger (2000) found, in their simulation study, that LMS is robust in testing for interaction effects when the variables moderately deviate from the normal distribution assumption. Similarly, Cham, West, Ma, and Aiken (2012) found that LMS performed well when the distributions of variables were normal or symmetric with excess kurtosis of less than 1. Future studies are required to examine the performance of LMS when data are not normally distributed, especially under highly nonnormal conditions.

Additionally, the simulations in this teaching note were based on the simple interaction model shown in Eq. 4. While this is the most commonly studied model in previous simulations that examined the effectiveness of LMS, other simulations in moderated-mediation (Cheung & Lau, 2017) and polynomial regression (Su et al., 2019) have also been conducted to demonstrate the accuracy of LMS and RCSLMS in testing latent interactions. Future simulation research is required to

investigate if the performance of LMS and RCSLMS can be generalized to models with other nonlinear effects, such as models with quadratic terms. Similarly, there is an urgent need to examine whether the problems associated with MMR in testing for two-way interaction effects identified in this study also occur when MMR is used to analyze equations with quadratic terms and three-way interactions.

To summarize, efforts to identify appropriate methods to examine latent interactions can be traced back almost four decades to when Kenny and Judd (1984) first suggested the early PI approach. Meanwhile, LMS has been developed for almost two decades. Despite most empirical researchers recognizing the problems of measurement errors in estimating relationships among variables, and thus using SEM with latent variables in estimating linear effects, typically when estimating moderating effects, they turn to regression-based approaches such as MMR that ignore measurement errors. Multiple simulation studies (for example, Cheung & Lau, 2017), together with the simulation results reported in this teaching note, have shown that ignoring measurement errors when estimating moderating effects will lead to seriously biased estimates and confidence intervals. Hence, results of significance tests may be flawed and any theoretical developments and practical recommendations are similarly likely to be erroneous. Therefore, we recommend future researchers follow the 4-step procedure outlined in this teaching note to estimate moderating and quadratic effects of latent variables with LMS.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10869-020-09717-0.

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management, 21*, 1141–1158. https://doi.org/10.1016/0149-2063(95)90026-8.

Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods, 20*, 665–685. https://doi.org/10.1177/1094428115627498.

Aguinis, H., & Gottfredson, R. K. (2010). Best-practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior, 31*, 776–786. https://doi.org/10.1002/job.686

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, London: Sage.

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*, 40–52. https://doi.org/10.1207/S15328007SEM0801_3.

Asparouhov, T, & Muthén, B. (2013). *Computing the strictly positive Satorra-Bentler chi-square test in Mplus*. Mplus Web Notes: No. 12. Retrieved from http://www.statmodel.com/examples/webnotes/SB5.pdf.

Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods, 8*, 274–289. https://doi.org/10.1177/1094428105278021.

Bernerth, J. B., & Aguinis, H. (2016). A critical review and best practice recommendations for control variable usage. *Personnel Psychology, 69*, 229–283. https://doi.org/10.1111/peps.12103.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley.

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables. In P. V. Marsden (Ed.), *Sociological methodology 1995 (Volume 25)*. American Sociological Association: Washington, D. C.

Bollen, K. A., & Paxton, P. (1998). Two-stage least squares estimation of interaction effects. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 125–151). Mahwah, NJ: Erlbaum.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 372–398. https://doi.org/10.1080/10705511.2012.687671.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin, 93*, 549–562. https://doi.org/10.1037/0033-2909.93.3.549.

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research, 47*, 840–876. https://doi.org/10.1080/00273171.2012.732901.

Cheung, G. W. (2009). A multiple-perspective approach in analyzing data in congruence research. *Organizational Research Methods, 12*, 63–68. https://doi.org/10.1177/1094428107310091.

Cheung, G. W., & Lau, R. S. (2017). Accuracy of parameter estimates and confidence intervals in moderated mediation models: A comparison of regression and latent moderated structural equations. *Organizational Research Methods, 20*, 746–769. https://doi.org/10.1177/1094428115595869.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330–351. https://doi.org/10.1037//1082-989X.6.4.330.

Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management, 19*, 915–922. https://doi.org/10.1177/014920639301900411.

Cortina, J. M., Chen, G., & Dunlap, W. P. (2001). Testing interaction effects in LISREL: Examination and illustration of available procedures. *Organizational Research Methods, 4*(4), 324–360. https://doi.org/10.1177/109442810144002.

Cortina, J. M., Markell-Goldstein, H. M., Green, J. P., & Chang, Y. (2019). How are we testing interactions in latent variable models? Surging forward or fighting shy? *Organizational Research Methods.* https://doi.org/10.1177/1094428119872531.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*(2), 317–327. https://doi.org/10.1037/0033-2909.105.2.317.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology, 29*, 1–19. https://doi.org/10.1007/s10869-013-9308-7.

Du, D., Derks, D., Bakker, A. B., & Lu, C. (2018). Does homesickness undermine the potential of job resources? A perspective from the work-home resources model. *Journal of Organizational Behavior, 39*, 96–112. https://doi.org/10.1002/job.2212.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

Edwards, J. R. (2009). Latent variable modeling in congruence research: Current problems and future directions. *Organizational Research Methods, 12*, 34–62. https://doi.org/10.1177/1094428107308920.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 128–141. https://doi.org/10.1207/S15328007SEM0801_7.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39–50. https://doi.org/10.1177/002224378101800104.

Goodman, S. A., & Svyantek, D. J. (1999). Person-organization fit and contextual performance: Do shared values matter. *Journal of Vocational Behavior, 55*, 254–275. https://doi.org/10.1006/jvbe.1998.1682.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*, 80–100. https://doi.org/10.1207/S15328007SEM1001_4.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2*, 233–256. https://doi.org/10.1177/109442819923002.

Harring, J. R., Weiss, B. A., & Li, M. (2015). Assessing spurious interaction effects in structural equation modeling: A cautionary note. *Educational and Psychological Measurement, 75*, 721–738. https://doi.org/10.1177/0013164414565007.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore, MD: John Hopkins University Press.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs, 1*, 57–93.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–88). Mahwah, NJ: Erlbaum.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98*, 875–925. https://doi.org/10.1037/a0033901.

Karasek, R. (1985). *Job content instrument: Questionnaire and user's guide (rev. 1.1)*. Los Angeles: University of Southern California.

Kelava, A., Werner, C. S., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., Cham, H., Aiken, L. S., & West, S. G. (2011). Advanced nonlinear latent variable modeling: Distribution analytic LMS and QML estimators of interaction and quadratic effects. *Structural Equation Modeling: A Multidisciplinary Journal, 18*, 465–491. https://doi.org/10.1080/10705511.2011.582408.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96*, 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*, 457–474. https://doi.org/10.1007/BF02296338.

Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42*, 647–674. https://doi.org/10.1080/00273170701710205.

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th edition), pp.33-34. New York: The Guilford Press.

Lau, R. S., & Cheung, G. W. (2012). Estimating and comparing specific mediation effects in complex latent variable models. *Organizational Research Methods, 15*, 3–16. https://doi.org/10.1177/1094428110391673.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101*, 1174–1188. https://doi.org/10.1037/a0024776.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198–1202.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal, 13*, 59–72. https://doi.org/10.1207/s15328007sem1301_3.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9*, 275–300. https://doi.org/10.1037/1082-989X.9.3.275.

Maslowsky, J., Jager, J., & Hemken, J. (2015). Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations method. *International Journal of Behavioral Development, 39*, 87–96. https://doi.org/10.1177/0165025414552301.

Mooijaart, A. & Satorra, A. (2009). On insensitivity of the chi-square model test to nonlinear misspecification in structural equation models. *Psychometrika, 74*, 443–455.

Muthén B. (2012). Latent variable interactions. Technical note, 1-9, retrieved from http://www.statmodel.com/download/LV%20interaction.pdf.

Muthén, L. K., & Muthén, B. O. (1998-2014). *Mplus user's guide* (Eighth ed.). Los Angeles, CA: Muthén & Muthén.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods, 17*, 372–411. https://doi.org/10.1177/1094428114548590.

Peterson, M. F., Smith, P. B., Akande, A., Ayestaran, S., Bochner, S., Callan, V., et al. (1995). Role conflict, ambiguity, and overload: A 21-nation study. *Academy of Management Journal, 38*, 429–452. https://doi.org/10.2307/256687.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*, 77–98. https://doi.org/10.1080/19312458.2012.679848.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods, 21*, 189–205. https://doi.org/10.1037/met0000052.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Sardeshmukh, S. R., & Vandenberg, R. J. (2017). Integrating moderation and mediation: A structural equation modeling approach. *Organizational Research Methods, 20*, 721–745. https://doi.org/10.1177/1094428119872531.

Satorra, A. & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*, 243–248. doi: https://doi.org/10.1007/s11336-009-9135-y, 2010.

Savalei, V. (2019). A comparison of several approaches for controlling measurement error in small samples. *Psychological Methods, 24*(3), 352–370. https://doi.org/10.1037/met0000181.

Stroebe, M., van Vliet, T., Hewstone, M., & Willis, H. (2002). Homesickness among students in two cultures: Antecedents and consequences. *British Journal of Psychology, 93*, 147–168. https://doi.org/10.1348/000712602162508.

Su, R., Zhang, Q., Liu, Y., & Tay, L. (2019). Modeling congruence in organizational research with latent moderated structural equations. *Journal of Applied Psychology, 104*(11), 1404–1433. https://doi.org/10.1037/ap10000411.

Umbach, N., Naumann, K., Brandt, H., & Kelava, A. (2017). Fitting nonlinear structural equation models in R with package nlsem. *Journal of Statistical Software, 77*, 1–20. https://doi.org/10.18637/jss.v077.i07.

Van Veldhoven, M., de Jonge, J., Broersen, S., Kompier, M., & Meijman, T. (2002). Specific relationships between psychosocial job conditions and job-related stress: A three-level analytic approach. *Work & Stress, 16*, 207–228. https://doi.org/10.1080/02678370210166399.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.