Assessing Spurious Interaction Effects in Structural Equation Modeling: A Cautionary Note

Educational and Psychological
Measurement
2015, Vol. 75(5) 721–738
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0013164414565007
epm.sagepub.com

Jeffrey R. Harring¹, Brandi A. Weiss², and Ming Li¹

Abstract

Several studies have stressed the importance of simultaneously estimating interaction and quadratic effects in multiple regression analyses, even if theory only suggests an interaction effect should be present. Specifically, past studies suggested that failing to simultaneously include quadratic effects when testing for interaction effects could result in Type I errors, Type II errors, or misleading interactions. Research investigating this issue has been limited to multiple regression models. Contrarily, structural equation modeling is a more appropriate analysis when hypotheses include latent variables. The current study utilized Monte Carlo simulation to investigate whether quadratic effects should be included in the latent variable interaction model. Consistent with previous research, it was found that including latent variable quadratic effects in the model successfully reduced the frequency of spurious interaction effects but at a cost of low power to detect true interaction effects, inaccurate parameter estimates, inaccurate standard error estimates, and reduced convergence rates. Based on findings from the current study, we recommend that researchers hypothesizing interactions between latent variables should test for these relations using the latent variable interaction model rather than the interaction quadratic model. If researchers are concerned about spurious interactions, then they may want to consider including quadratic effects in the model, provided that they have sample sizes of at least 500 and high

¹University of Maryland, College Park, MD, USA

²George Washington University, Washington, DC, USA

Corresponding Author:

Jeffrey R. Harring, University of Maryland, 1230 Benjamin Building, College Park, MD 20742-1115, USA.
Email: harring@umd.edu

indicator reliability. We caution all researchers to base higher order effects models on theory.

Keywords

latent interaction models, interaction effects, nonlinear structural equation modeling, spurious effects

Structural equation modeling (SEM) refers to an analytic method for modeling a hypothesized pattern of linear relations among a set of observed and latent variables (Bollen, 1989). In recent years, this type of modeling has been extended to integrate nonlinear effects in the structural model. Like their manifest (nonlatent) variable regression counterparts, structural models that presume an additive, linear relation among latent variables may be too inflexible to adequately account for the complexities of the exogenous–endogenous relation often found in practice (Harring, 2013; Wall, 2009; Wall & Amemiya, 2007). For example, if a slope relating a continuous latent predictor and endogenous criterion is thought to systematically change across levels of a second latent predictor, the inclusion of a cross-product term could augment the simple additive structural model in a straightforward manner to account for this relation, which is often referred to as an interaction effect (Kenny & Judd, 1984).

Another example in which the inclusion of a cross-product term could be beneficial is the case of a quadratic relation in which the effect of the exogenous variable on the endogenous variable changes depending on the level of the exogenous variable itself. For example, consider an exogenous variable, ξ , and an endogenous variable, η . One instance of a positive quadratic relation exists when as ξ increases, η also increases up until a maximum point, after which the relation between ξ and η decreases. Interaction effects and quadratic relations are often considered in conjunction because a quadratic effect can be viewed as a special type of interaction effect, and similarly an interaction effect is nonlinear in nature. Nonlinear relations like those described above—especially multiplicative relations like interactions—are just the types of relations that often are of special interest and provide insight into effects of substantive variables.

Several studies have stressed the importance of simultaneously estimating interaction and quadratic effects in multiple regression analyses to control for spurious interactions (Cortina, 1993; Ganzach, 1997; Lubinski & Humphreys, 1990). Because multiple regression for measured variables can be cast as a special case of regression models falling within the SEM framework, it may appear on the surface that the results from these past studies would generalize to the SEM context in a straightforward manner. However, we would argue that because SEM takes into account random measurement error of the measured variables that reliability (or lack thereof) of the measures will be pivotal in whether or not these multiplicative effects in a latent variable framework will be spurious. Simulation studies that have studied interaction,

quadratic, and other nonlinear effects in a latent framework have been limited in that they have only included simulated data from a population with no interaction effect in order to determine if one (spurious interaction) appears. Previous research has not investigated the impact that testing for unhypothesized quadratic effects could have on Type II errors, and thus power. The current study used Monte Carlo simulation to address this model specification issue in order to provide recommendations to applied researchers hypothesizing these types of relations.

The Evolution of the Constrained Approach

The evolution of estimating interactions and quadratic effects in structural equation models began with Kenny and Judd (1984) who hypothesized the interaction between two latent exogenous variables and used cross-products to represent the latent interaction effect. Hayduk (1987) was the first to expand the Kenny–Judd model to include a latent endogenous variable. Many methods have been suggested to estimate nonlinear latent variable models (Arminger & Muthén, 1998; Bollen & Paxton, 1998; Klein & Moosbrugger, 2000; Ping, 1996; Wall & Amemiya, 2003). A convenient summary of work on the latent variable interaction model is the edited volume by Schumacker and Marcoulides (1998) that featured methods or critical assessments that specifically underscored approaches known as product indicator methods.

In SEM, the LISREL specification for the structural model in which two exogenous variables interact with one another can be written as

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta, \quad (1)$$

where η is an endogenous latent variable, ξ_1 and ξ_2 are first-order exogenous latent variables, $\xi_1\xi_2$ is a latent interaction term, α represents the regression intercept, γ_1 , γ_2 , and γ_3 are direct path estimates, and ζ denotes the latent regression residual.

Jöreskog and Yang (1996) extended the Kenny–Judd model to include a mean structure. They noted that in the Kenny–Judd model, even if the observed variables were mean-centered, their products would not necessarily be mean-centered. Consequently, the latent interaction variable, $\xi_1\xi_2$, would also not be mean-centered, and thus a mean structure was always necessary.

Algina and Moulder (2001) extended on the Jöreskog and Yang (1996) model and that advocated by Yang-Wallentin and Jöreskog (2001) by mean-centering the independently observed variables. Algina and Moulder (2001) referred to this model as the “constrained” model and found that it was more likely to converge, was less biased, and had better Type I error control than the Jöreskog and Yang (1996) uncen-tered model. The Algina and Moulder (2001) constrained approach applied to a latent interaction model imposed four types of constraints on the model.

The first constraint required the loadings for each of the interaction indicators to be constrained to equal the product of the two indicators that created the interaction indicator. The second constraint required the mean of the interaction to be set equal to the covariance between ξ_1 and ξ_2 (i.e., $\kappa_3 = \phi_{21}$). The constrained model is based

on the assumption that ξ_1 and ξ_2 are normally distributed and that the correlation between $\xi_1\xi_2$ and each of the first-order terms (i.e., ξ_1 and ξ_2) are zero; $\phi_{31} = \phi_{32} = 0$. Accompanying this assumption is the third constraint that is imposed on the variance of the interaction latent variable: $\phi_{33} = \phi_{11}\phi_{22} + \phi_{21}^2$. The final constraint was imposed on the errors of each of the indicators for the interaction.

The Unconstrained Approach

Marsh, Wen, and Hau (2004) introduced an unconstrained model in which all four of the constraints were released. Like the generalized appended product indicator (Wall & Amemiya, 2001), this model also did not require the stringent normality assumption of ξ_1 and ξ_2 by allowing $\xi_1\xi_2$ to covary with ξ_1 and ξ_2 . Marsh, Wen, and Hau (2006) noted that this unconstrained model was much easier for researchers to implement than the constrained model because it does not necessitate the specification of nonlinear constraints. Like the constrained approach, the unconstrained approach is based on the assumption that ξ_1 and ξ_2 were jointly multivariate normal, and the distribution of $\xi_1\xi_2$ was nonnormal. Previous research has shown that when data are normally distributed, the unconstrained approach is unbiased in most conditions, and almost as unbiased as the constrained approach (Marsh et al., 2004; Weiss, 2010).

For the empirical study described below, we assume that there are three measured indicator variables for each of the two exogenous factors as well as for the endogenous factor. Furthermore, a matched-paired cross-product indicator strategy, proposed by Marsh et al. (2004), is adopted in which the product of measured indicators from each of the linear factors is used to create indicators of the latent interaction term (and subsequent quadratic terms as well). Figure 1 depicts the unconstrained approach with the following specifications. The structural model for the unconstrained approach is the same as in Equation 1. The measurement model connecting the measured variables to the latent variables is specified as

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \tau_{y1} \\ \tau_{y2} \\ \tau_{y3} \end{pmatrix} + \begin{pmatrix} 1 \\ \lambda_{y2} \\ \lambda_{y3} \end{pmatrix} \eta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}, \quad (2)$$

$$\begin{pmatrix} x_1 - \tau_{x1} \\ x_2 - \tau_{x2} \\ x_3 - \tau_{x3} \\ x_4 - \tau_{x4} \\ x_5 - \tau_{x5} \\ x_6 - \tau_{x6} \\ (x_1 - \tau_{x1})(x_4 - \tau_{x4}) \\ (x_2 - \tau_{x2})(x_5 - \tau_{x5}) \\ (x_3 - \tau_{x3})(x_6 - \tau_{x6}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & \lambda_6 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_8 \\ 0 & 0 & \lambda_9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{pmatrix}, \quad (3)$$

$$\Theta_\varepsilon = \text{diag}(\theta_{\varepsilon1}, \theta_{\varepsilon2}, \theta_{\varepsilon3}), \quad (4)$$

Figure 1. A path diagram of a latent interaction model with three observed indicators for each latent variable under the unconstrained model.

$$\Theta_{\delta} = \text{diag}(\theta_{\delta 1}, \theta_{\delta 2}, \theta_{\delta 3}, \theta_{\delta 4}, \theta_{\delta 5}, \theta_{\delta 6}, \theta_{\delta 7}, \theta_{\delta 8}, \theta_{\delta 9}), \quad (5)$$

where τ_y and the τ_{xi} terms are the means of y and x variables, and λ_i , $i = 1, \dots, 9$ are the coefficients relating each latent variable in $\xi' = (\xi_1, \xi_2, \xi_1\xi_2)$ to its measured variable indicators. Under the model specified in Equations 1 to 3, the exogenous latent variables ξ_1 and ξ_2 and the regression error terms ζ and δ_i are assumed to be multivariate normal with mean vector of zero and mutually uncorrelated (except that ξ_1 and ξ_2 are permitted to correlate). The mean vector and covariance matrix for ξ_1 , ξ_2 , and $\xi_1\xi_2$ are, respectively,

$$E[\xi] = \begin{pmatrix} 0 \\ 0 \\ \kappa_3 \end{pmatrix}, \quad \text{Cov}(\xi) = \begin{pmatrix} \varphi_{11} & & \\ \varphi_{21} & \varphi_{22} & \\ \varphi_{31} & \varphi_{32} & \varphi_{33} \end{pmatrix}. \quad (6)$$

The Latent Quadratic Model

Similarly to the latent interaction model, one may express the latent quadratic model as

$$\eta = \alpha + \gamma_1 \xi_1 + \gamma_2 \xi_1^2 + \zeta, \quad (7)$$

where η is the latent criterion, α is the intercept, γ_1 and γ_2 are the structural coefficients of the latent first-order and quadratic exogenous latent predictors, ξ_1 and ξ_1^2 , respectively, and ζ denotes the disturbance term. Similarly to the interaction model,

each indicator for the first-order latent variable can be squared to create product-indicators for the latent quadratic effect.

Model Misspecification Issues With Interaction and Quadratic Effects

When first-order terms are perfectly correlated, then there is no need to differentiate between an interaction effect and a quadratic effect because they are the same. When the correlation between first-order terms is zero, then one can clearly see the differences between interaction effects and quadratic effects. However, as the relation between the first-order terms increases (i.e., as multicollinearity between first-order terms increases), it becomes more difficult to differentiate between interaction effects and quadratic effects. Thus, an interaction effect may be statistically significant because an endogenous variable and an exogenous variable are quadratically related, or vice versa.

Several studies have stressed the importance of simultaneously estimating interaction and quadratic effects in multiple regression analyses, even if theory only suggests an interaction effect (Cortina, 1993; Ganzach, 1997; Lubinski & Humphreys, 1990). Specifically, these studies found that failing to simultaneously test for quadratic effects when testing for an interaction effect could have resulted in one of the following situations:

1. A significant interaction effect may be observed when in reality there is no interaction effect (Type I error, also referred to as a spurious interaction; Cortina, 1993; Lubinski & Humphreys, 1990).
2. An interaction may appear to be nonsignificant when in fact there is an interaction effect (Type II error; Ganzach, 1997).
3. The interaction may appear positive when in fact it was negative (misleading interaction; Ganzach, 1997).

There have been mixed recommendations about whether or not one should test for quadratic effects in order to exert Type I error control when interactions are hypothesized (Aiken & West, 1991; Cortina, 1993; Ganzach, 1997; Lubinski & Humphreys, 1990; Shepperd, 1991). Some researchers suggested a more tolerant approach, requiring theory before including nonlinear effects in a model (Aiken & West, 1991). Other researchers stated that nonlinear effects should only be included when the correlation between the first-order effects is nonzero (i.e., $\phi_{21} \neq 0$; Shepperd, 1991). More stringent recommendations have stated that researchers should always test for both quadratic and interaction effects (Cortina, 1993; Ganzach, 1997; Lubinski & Humphreys, 1990). This is a model specification issue.

The majority of previous studies addressing this issue are limited in that they only investigated this recommendation using multiple regression models (Cortina, 1993; Lubinski & Humphreys, 1990; MacCallum & Mar, 1995; Shepperd, 1991).

Contradictorily, many social and behavioral science researchers hypothesize interactions between latent variables, which contain measurement error. Therefore, when interactions are hypothesized between latent variables, latent variable analyses (such as SEM) are more appropriate than the traditional measured variable analyses.

To date, one study investigated the simultaneous estimation of quadratic and interaction effects using latent variable models (Klein, Schermelleh-Engel, Moosbrugger, & Augustin, 2009). The authors of this study found that testing for quadratic effects in addition to an interaction effect resulted in lower Type I error and thus fewer spurious interactions. However, this study was limited in that it utilized one approach of testing for nonlinear effects (the latent moderated structures approach) when many other methods of testing for nonlinear effects also exist. Additionally, the simulation design of the Klein et al. (2009) study only investigated spurious interaction effects (i.e., situation one above, but neither two nor three).

While controlling for spurious interaction effects is a valid concern, controlling for Type II errors should also be of concern. If a researcher solely focuses on avoiding spurious interactions, true interactions may go undetected. This is a power issue, which has not been addressed in previous studies. Additionally, testing for quadratic effects that are not based in theory could potentially result in spurious quadratic effects. Thus, when attempting to avoid making a Type I error with regard to an interaction effect, one may in turn make a Type I error with regard to a quadratic effect. The goal of the current study was to use Monte Carlo simulation to investigate whether or not quadratic effects should be included in the latent variable interaction model. This study builds on previous research in that Type II errors, power, and spurious quadratic effects are considered in addition to spurious interaction effects and spurious quadratic effects.

Method

To address the research questions, three models were compared, a first-order effects only model (Model A)

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \zeta, \quad (8)$$

an interaction model like that in Equation 1 (Model B), and an interaction-quadratic model (Model C),

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \gamma_4\xi_1^2 + \gamma_5\xi_2^2 + \zeta. \quad (9)$$

Using these three models, two scenarios were created. In the first scenario, Model A was true, and Models B and C were tested. This scenario allowed for investigation of spurious interactions and spurious quadratic effects. In the second scenario, Model B was true, and Models B and C were tested. This scenario allowed for investigating Type II errors associated with the interaction effect, as well as spurious quadratic effects.

All variables were simulated to come from a population in which

$$\eta = 0.4\xi_1 + 0.4\xi_2 + \gamma_3\xi_1\xi_2 + \zeta. \quad (10)$$

Thus, while γ_1 and γ_2 paths were set equal to 0.4 based on values used by Marsh et al. (2004), the γ_3 path varied depending on whether or not an interaction effect was present. Although all data were simulated to come from a population in which Equation 10 was true, and thus $\alpha = 0$ in the population; α is known to not necessarily be zero when the model is estimated (Jöreskog & Yang, 1996; Marsh et al., 2004). Thus, the structural model that was estimated included α .

The effect size associated with the interaction effect represents the additional variance that the interaction explains in η above and beyond that which can be explained by the first-order effects. This value was described by Marsh et al. (2004) and is represented by

$$R^2_{\gamma_3} = \frac{\gamma_3^2(\phi_{11}\phi_{22} + \phi_{21}^2)}{\sigma_\eta^2}. \quad (11)$$

Previous studies have set this value at 5% (Harring, Weiss, & Hsu, 2012; Jaccard & Wan, 1995; Klein & Muthén, 2007; Little, Bovaird, & Widaman, 2006; Marsh et al., 2004; Moulder & Algina, 2002; Weiss, 2010). The current study used similar values allowing the interaction to account for 0% (Model A), and 5% of the unique variance in η (Model B).

The measurement model depicted in Figure 1 was used throughout, except to allow for the different structural models. Reliability of the indicators has been known to affect estimates of structural coefficients (Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007). Commensurate with past simulation studies (see, e.g., Algina & Moulder, 2001; Jaccard & Wan, 1995) loadings for indicators were set equal to either 0.5 or 0.9 in the population-generating model. Less reliable loadings were thought to result in lower power, which corresponds to an increase in the probability of making a Type II error (Lubinski & Humphreys, 1990). Thus, the less reliable loading condition was chosen to investigate the power Models B and C had to identify real interaction effects. The loading of 0.9, which denotes a high level of measured variable reliability, was chosen based on previous research (Marsh et al., 2004; Weiss, 2010). We chose the loadings to be equal across the indicator variables. This decision was made in part by the practical realities of executing the simulation although we do recognize that in applied research settings that the loadings (and thus the reliabilities) for each observed variable indicator could very well be distinct.

Two different sample sizes were used (referred to as small and large here), which corresponded to $n = 100$ and $n = 500$, respectively. The sample sizes were chosen partially based on past simulation studies that used similar sample sizes (Jaccard & Wan, 1995; Klein & Muthén, 2007; Marsh et al., 2004; Moulder & Algina, 2002; Schermelleh-Engel, Klein, & Moosbrugger, 1998). Furthermore, it is known that small sample size has an indirect effect on power. Thus, the sample size of 100 was

chosen to investigate the power of Models B and C. Wall and Amemiya (2003) evaluated methods using sample sizes as large as 1,000. However, in preliminary analyses, little difference was found in the bias and precision of parameter estimates between sample sizes of 500 and 1,000.

The correlation between the two first-order latent variables ξ_1 and ξ_2 was varied. Most previous studies investigating latent variable interaction effects used values ranging from .20 to .50 (Jaccard & Wan, 1995; Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Little et al., 2006; Marsh et al., 2004; Moulder & Algina, 2002; Schermelleh-Engel et al., 1998). One study (Marsh et al., 2004) used a value as high as .70. Because the impact of high multicollinearity on estimating interaction and quadratic effects is central to the current research question, including a condition with a large value of ϕ_{21} was desirable. Therefore, the current study manipulated ϕ_{21} to be .25, .50 (to represent values that have been used in previous studies), or .75 (to represent a high degree of multicollinearity) in the population-generating model (given that the latent variables have unit variance $\phi_{11} = \phi_{22} = 1$).

In summary, the simulation design was 2 (Model A and Model B) $\times$ 2 (loadings) $\times$ 2 (sample sizes) $\times$ 3 (correlations between first-order effects), resulting in 24 conditions. All 24 conditions were analyzed with both Models B and C, resulting in 48 sets of analyses. All data were simulated to come from a normal distribution. Due to the complicated nonlinear constraints associated with the constrained approach, the unconstrained approach was used to test for the interaction and quadratic effects. Previous research has shown that when data are normally distributed, the unconstrained approach is unbiased in most conditions, and almost as unbiased as the constrained approach (Marsh et al., 2004; Weiss, 2010). Data were simulated using SAS 9.2. Maximum likelihood was used to estimate model parameters. The Hessian matrix from the Newton–Raphson algorithm was used in the computation of the standard errors. Data for each of the possible 24 conditions was simulated 1,000 times. Simulated data were analyzed using LISREL 8.8 and SAS 9.2.

Results

Convergence Rates

Table 1 contains the converge rates for each condition. In all conditions, Model B was more likely to converge than Model C. Convergence rates decreased as ϕ_{21} increased, and this pattern was more pronounced for Model C than Model B. As expected, convergence rates were lower for when loadings were 0.5 and sample size was 100.

Spurious Interactions

By definition, spurious interactions occur when a statistically significant interaction effect is observed when in reality there is no interaction effect. Spurious interactions are numerically represented by Type I error rates. Type I error for the interaction

Table 1. Convergence Rates (Expressed in Percentages).

		$\lambda = 0.50$			$\lambda = 0.90$		
		ϕ_{21}	0.25	0.50	0.75	0.25	0.50
True Model A							
$n = 100$	Model B	73	75	68	100	100	100
	Model C	66	63	59	100	99	96
$n = 500$	Model B	93	93	92	100	100	100
	Model C	86	86	83	100	100	100
True Model B							
$n = 100$	Model B	75	71	63	100	100	100
	Model C	67	61	59	100	100	95
$n = 500$	Model B	93	91	90	100	100	100
	Model C	88	86	81	100	100	100

effect was computed from the proportion of converged solutions that had a statistically significant interaction effect in the simulated data when the population interaction effect was zero (i.e., the proportion of times that γ_3 was statistically significant, when Model A was true). An α level of .05 was used to compute Type I error, thus values around .05 were desirable.

Scenario 1, in which Model A was true and Models B and C were tested, allowed for investigation of spurious interactions. Figure 2 contains the Type I error rates for Scenario 1. When loadings were 0.5, and Model B was tested, Type I error was above the desired .05 level, indicating that there was a higher than desired rate of spurious interactions. When loadings were 0.5, Model C (which included quadratic effects) resulted in successful reduction of Type I error rates. When loadings were 0.9, spurious interactions were not problematic for either Model B or Model C. Altering the correlation between first-order effects did not affect the rate of Type I errors for most conditions. In nearly all conditions, increasing the sample size resulted in lower Type I error rates, with the exception of the Model C, $\phi_{21} = .75$ conditions. Of note, Model C resulted in lower than desired (approximately .02 rather than .05) Type I error rates when loadings were 0.5 and the sample size was 500.

Power

Power was calculated as one minus the proportion of converged solutions in which γ_3 was not statistically significant, when the true γ_3 value was nonzero (i.e., one minus the proportion of times that γ_3 was not statistically significant when Model B was true). Most methodologists suggest that power values of .70 or higher are sought. Scenario 2, in which Model B was true and Models B and C were tested, allows one to examine the power these models had to detect the interaction effect.

Figure 3 contains the power estimates when Model B was the true model and Models B and C were tested. As expected, Model B had more power to detect true

Figure 2. Type I error for interaction effect (true Model A, tested Models B and C).

Figure 3. Empirical power for interaction effect (true Model B, tested Models B and C).

Figure 4. Spurious quadratic effects when Model C is tested (true Model A).

interaction effects than Model C in all conditions. However, for Model B the conditions in which loadings were 0.9 were the only conditions in which statistical power reached values that were near desirable. That is, when loadings were 0.5, power was low regardless of sample size and regardless of which model was tested. In most conditions, power for both models also decreased as multicollinearity increased.

Spurious Quadratic Effects

Both Scenarios 1 and 2 allowed for the investigation of spurious quadratic effects. Type I error for the quadratic effects (i.e., the rate of spurious quadratic effects) was computed by calculating the proportion of converged solutions that had a statistically significant quadratic effect in the simulated data when the population quadratic effect was zero (i.e., the proportion of times that γ_4 and γ_5 in Equation 9 were statistically significant, when Models A and B were true). An α level of .05 was used to compute Type I error, and thus values around .05 were wanted.

Similar Type I error rates were found for both quadratic effects in Model C; thus, only the rates for the first quadratic effect associated with ξ_1^2 are reported. Figure 4 contains the Type I error rates for the first quadratic effect, γ_4 , when the first-order

Figure 5. Spurious quadratic effects when Model C is tested (true Model B).

effects model (Model A) was true, but the interaction and quadratic effects model (Model C) was tested whereas the Type I error rates for the quadratic effect, γ_4 , when the interaction only model (Model B) was true, but the interaction and quadratic model (Model C) was tested is shown in Figure 5. Evidently, the Type I error rates were near the desired level for all of the conditions, indicating that spurious quadratic effects were rarely present. Of note, the conditions with both loadings of 0.5 and sample sizes of 500 had lower than desired Type I error. Also, the condition with 0.5 loading, sample size of 100, and high first-order correlation had a slightly higher than desired Type I error rate.

Bias and Relative Bias

Relative bias was also used as an evaluation criterion for investigating spurious interaction effects. Relative bias is defined by the average difference between the parameter estimates and the population-generating value divided by the population-generating parameter value. In the conditions in which the population-generating parameter value γ_3 was equal to zero the bias could not be divided by zero, therefore bias—the difference between the average parameter estimates and the population-generating—were computed.

Table 2. Bias and Relative Bias for the Interaction Effect.

		$\lambda = 0.50$			$\lambda = 0.90$			
		ϕ_{21}	0.25	0.50	0.75	0.25	0.50	0.75
True Model A (bias)								
$n = 100$	Model B		.310	.090	.295	.007	.001	-.006
	Model C		.117	.518	.036	.031	.040	.071
$n = 500$	Model B		.006	.011	-.027	.000	-.001	.001
	Model C		.006	.028	.014	.000	-.002	.005
True Model B (relative bias)								
$n = 100$	Model B		.102	.121	.999	.053	.058	-.059
	Model C		.008	-.598	.094	.339	.415	.808
$n = 500$	Model B		.227	.096	-.202	.072	.052	-.033
	Model C		-.602	-.671	.026	.160	.201	.789

Table 2 contains the bias and relative bias estimates for the interaction effect across the 48 conditions. Values close to zero are desirable because they indicate low bias. Hoogland and Boomsma (1998) suggested that unbiased parameter estimates should be within an absolute value of 0.05. When Model B was tested, bias was low when loadings were 0.9 or when the sample size was 500. The strength of the relation between the first-order factors did not have a meaningful impact on bias across most conditions.

This is an indication that regardless of whether or not an interaction effect is present, testing for the interaction effect results in relatively unbiased estimates of the interaction for the conditions simulated in this study, provided that loadings are 0.9 or the sample size is 500.

When there was no true interaction effect (i.e., Model A was true), Model C also resulted in unbiased estimates of the interaction effect when loadings were 0.9 or sample size was 500. However, when a true interaction effect existed (i.e., Model B was true), Model C yielded biased estimates of the interaction in almost all conditions. In most conditions, this bias was positive, indicating that Model C overestimated the interaction effect.

Precision

Precision was evaluated by calculating a ratio of the average estimated standard error, $SE(\theta)$, to the observed standard deviation of the parameter estimates, $SD(\theta)$, across the replications in which estimation converged. The ratio indicates how the average standard error estimates compare to the average empirical variance estimates. Values close to one are desirable and indicate that the standard error values computed by the model are representative of the population values. Values less than one indicate the model underestimates the standard error, while values greater than one would indicate that the model overestimates the standard error estimates.

Table 3. SE/SD Ratio for the Interaction Effect.

		$\lambda = 0.50$			$\lambda = 0.90$		
ϕ_{21}		0.25	0.50	0.75	0.25	0.50	0.75
<hr/>							
True Model A							
$n = 100$	Model B	1.96	4.16	1.30	0.96	0.94	0.92
	Model C	2.66	3.77	2.56	1.84	0.82	1.06
$n = 500$	Model B	0.92	1.96	1.68	0.96	1.01	1.00
	Model C	0.46	1.22	4.57	0.96	0.97	0.90
True Model B							
$n = 100$	Model B	2.94	2.91	2.54	0.94	0.92	0.86
	Model C	2.95	3.21	3.04	0.68	0.87	1.69
$n = 500$	Model B	1.60	1.24	1.68	0.97	0.98	0.92
	Model C	0.90	1.10	4.02	0.95	0.95	0.62

Table 3 contains the standard error to standard deviation ratios associated with estimating the interaction effect. When loadings were 0.9, ratios were closer to 1.0 for Model B than for Model C in all conditions. With loadings of 0.9 and sample size of 100, Model C resulted in inaccurate standard error estimates. When loadings were 0.5, both Model B and Model C resulted in inaccurate standard error estimates; in the majority of conditions these standard errors were overestimated.

Discussion

Several previous studies have stressed the importance of simultaneously estimating interaction and quadratic effects, even if theory only suggests an interaction effect exists (Cortina, 1993; Ganzach, 1997; Klein et al., 2009; Lubinski & Humphreys, 1990; Shepperd, 1991). The current study investigated the impact that incorporating nontheoretical quadratic effects had on Type I error rates, power, relative bias estimates, standard error estimates, and convergence rates in a latent variable framework. Consistent with previous research, we found that including latent variable quadratic effects in the model successfully reduced the frequency of spurious interaction effects. However, this also resulted in a large reduction of statistical power to detect interaction effects particularly when loadings (and thus reliability) of indicators were 0.5, samples were 100, and first-order terms were strongly, linearly related. Only in the most ideal conditions (loadings of 0.9 and sample size of 500) was statistical power acceptable (.70 or greater) when nontheoretical quadratic effects were included in the model, and in these conditions the interaction-quadratic model resulted in 16% to 79% bias of the interaction effect.

Also consistent with previous research, testing the interaction-only model resulted in spurious interaction effects in some conditions. However, provided that loadings (and thus reliability) of indicators were 0.9, the rate of spurious interaction effects was consistent with alpha, and thus similar to expected values. Also, in most

conditions the convergence rates were higher and the parameter estimates and standard error estimates associated with the interaction effect were more accurate with the interaction-only model in comparison to the interaction-quadratic model.

Based on findings from the current study, we recommend that researchers hypothesizing interactions between latent variables test for these relations using the latent variable interaction model, instead of the interaction-quadratic model. The latent variable interaction model, as tested here, had higher convergence rates, adequate Type I error rates, higher power (and thus lower Type II error rates), resulted in unbiased estimates of the interaction effect, and had more accurate standard error estimates across most conditions when compared with the interaction-quadratic model. Our recommendation is contrary to that from previous studies in which it is recommended that researchers should always test for both quadratic and interaction effects (Cortina, 1993; Ganzach, 1997; Klein et al., 2009; Lubinski & Humphreys, 1990).

Of note, the sample size and strength of the relation between first-order effects did not impact the bias of the interaction effects with the interaction-only model. Rather, the reliability of the indicators had the greatest impact on Type I error rates associated with the interaction effect. This finding is similar to those from previous empirical studies (see, e.g., Dimitruk et al., 2007; Jaccard & Wan, 1995). Also, because the relation between first-order factors is related to the value of the interaction effect, relation between first-order factors, and the conditional variance of η , power is directly related to these effects, and thus we are limited in making direct comparisons of the relation of the first-order factors and statistical power.

While the interaction-quadratic model may not be necessary to estimate interaction effects when loadings (and thus reliability) of indicators is adequate, it could be beneficial in reducing Type I error rates when loadings are low. If researchers are concerned about spurious interactions (e.g., if high-stakes decisions are to be made based on findings of statistically significant interaction effects), then they may want to consider including quadratic effects in the model to reduce spurious findings. However, if non-theoretical quadratic effects are to be included in models for this purpose, sample sizes should be at least 500 and indicator reliability estimates should be high. Of course, caution should be taken in interpreting results when loadings are low with any model, not just the interaction-quadratic model.

Recommendations from this study do not mean that quadratic effects should never be included in latent interaction models. Researchers should test the model that they theorize. That is, if a researcher hypothesizes a quadratic effect, then he or she should test for that effect. This recommendation is consistent with Aiken and West (1991) who state theory should be present before incorporating nonlinear effects into a model. Thus, we caution all researchers to base higher order effects models on theory.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was funded by a grant from the Institute of Education Sciences, U.S. Department of Education, to Boston College (#R305A140114). The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: 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: A Multidisciplinary Journal*, 8, 40-52.
- Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hasting algorithm. *Psychometrika*, 63, 271-300.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley.
- Bollen, K. A., & Paxton, P. (1998). Interactions of latent variables in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 5, 267-293.
- Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management*, 19, 915-922.
- Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology*, 3, 100-114.
- Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, 2, 235-247.
- Harring, J. R. (2013). A piecewise regression model for latent variables. *Educational and Psychological Measurement*, 74, 197-214.
- Harring, J. R., Weiss, B. A., & Hsu, J. C. (2012). A comparison of methods for estimating quadratic effects in nonlinear structural equation models. *Psychological Methods*, 17, 193-214.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL*. Baltimore, MD: Johns Hopkins University Press.
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and meta-analysis. *Sociological Methods & Research*, 26, 329-367.
- Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, 117, 348-357.
- 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.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96, 201-210.
- Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65, 457-474.
- 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, 647-673.

- Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Augustin, K. (2009). Assessing spurious interaction effects. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 13-28). Rotterdam, Netherlands: Sense.
- Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 13, 497-519.
- Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Quantitative Methods in Psychology*, 107, 385-393.
- MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin*, 118, 405-421.
- 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.
- Marsh, H. W., Wen, Z., & Hau, K. T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 225-265). Greenwich, CT: Information Age.
- Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 1-19.
- Ping, R. A. (1996). Latent variable regression: A technique for estimating interaction and quadratic coefficients. *Multivariate Behavioral Research*, 31, 95-120.
- Schermelleh-Engel, K., Klein, A., & Moosbrugger, H. (1998). Estimating nonlinear effects using a latent moderated structural equations approach. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 203-238). Mahwah, NJ: Lawrence Erlbaum.
- Schumacker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Shepperd, J. A. (1991). Cautions in assessing spurious "moderator effects." *Psychological Bulletin*, 110, 315-317.
- Wall, M. M. (2009). Maximum likelihood and Bayesian estimation for nonlinear structural equation models. In R. Millsap & A. Maydeu-Olivares (Eds.), *Handbook of quantitative methods in psychology* (pp. 540-567). Thousand Oaks, CA: Sage.
- Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, 26, 1-29.
- Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, 56, 47-63.
- Wall, M. M., & Amemiya, Y. (2007). A review of nonlinear factor analysis and nonlinear structural equation modeling. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 337-362). Mahwah, NJ: Erlbaum.
- Weiss, B. A. (2010). *A comparison of methods for testing for interaction effects in structural equation modeling* (Unpublished doctoral dissertation). University of Maryland, College Park.
- Yang-Wallentin, F., & Jöreskog, K. G. (2001). Robust standard errors and chi-squares for interaction models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New development and techniques in structural equation modeling* (pp. 159-171). Mahwah, NJ: Lawrence Erlbaum.