STRUCTURAL EQUATION MODELING, *13*(4), 497–519

On the Merits of Orthogonalizing Powered and Product Terms: Implications for Modeling Interactions Among Latent Variables

Todd D. Little[a]
University of Kansas

James A. Bovaird[a]
University of Nebraska-Lincoln

Keith F. Widaman
University of California at Davis

The goals of this article are twofold: (a) briefly highlight the merits of residual centering for representing interaction and powered terms in standard regression contexts (e.g., Lance, 1988), and (b) extend the residual centering procedure to represent latent variable interactions. The proposed method for representing latent variable interactions has potential advantages over extant procedures. First, the latent variable interaction is derived from the observed covariation pattern among all possible indicators of the interaction. Second, no constraints on particular estimated parameters need to be placed. Third, no recalculations of parameters are required. Fourth, model estimates are stable and interpretable. In our view, the orthogonalizing approach is technically and conceptually straightforward, can be estimated using any structural equation modeling software package, and has direct practical interpretation of parameter estimates. Its behavior in terms of model fit and estimated standard errors is very reasonable, and it can be readily generalized to other types of latent variables where nonlinearity or collinearity are involved (e.g., powered variables).

Research in the social sciences frequently includes hypotheses about interactive or nonlinear effects on a given outcome variable. When it comes to estimating such effects, however, there is lack of consensus on how to do so properly, particularly when performing structural equation modeling (SEM). A plethora of methods have been proposed and discussed, including those described in Algina and Moulder (2001),

[a]Author contribution by the first two authors was equivalent.

Correspondence should be addressed to Todd D. Little, Schiefelbusch Institute for Life Span Studies, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045–7555. E-mail: yhat@ku.edu; and James A. Bovaird, Department of Educational Psychology, University of Nebraska-Lincoln, 114 Teachers College Hall, Lincoln, NE 68588–0345. E-mail: jbovaird2@unl.edu

Jaccard and Wan (1995), Klein and Moosbrugger (2000), Klein and Muthén (2002), Marsh, Wen, and Hau (2004), Ping (1996a, 1996b), Schumacker and Marcoulides (1998), and Wall and Amemiya (2001, 2003). Most approaches to latent variable interactions are based on a product indicator methodology originated by Kenny and Judd (1984) that requires a level of technical and computational sophistication that renders them quite inaccessible to the average practitioner.

The focus of this article is on the development of an intuitive and technically straightforward approach to estimating interactive and nonlinear effects within SEM. To clarify the position taken, however, methods for estimating nonlinear effects within multiple regression are first reviewed and some of the difficulties inherent within the regression framework are discussed. The focus is on the use of *residual centering* (Lance, 1988) versus *mean centering* (Cohen, 1978; Cronbach, 1987). Next an extension of residual centering (referred to as *orthogonalizing*) to represent nonlinear interaction effects in latent variable models is discussed.

NONLINEAR EFFECTS IN MULTIPLE REGRESSION

Although the ultimate goal here is to explicate the usefulness of residual centering in the context of latent variable analyses, first, some of the merits of residual centering interaction and powered terms in standard multiple regression contexts must be highlighted. In ordinary least squares (OLS) regression, the product of two variables can be used to represent the interactive effect, as seen in Equation 1:

$$Y = b_1X + b_2Z + b_3XZ + e \tag{1}$$

where Y is the outcome variable of interest, e is the assumed error term, X and Z are the first-order predictor variables, and XZ is the multiplicative or product term that represents the interaction effect. Essentially, this regression equation specifies that the slope of the line relating X to Y changes at different levels of Z, or equivalently, that the slope of the line relating Z to Y changes at different levels of X. Saunders (1956) first demonstrated that a product term accurately reflects a continuous variable interaction. Similarly, natural polynomial or powered variables (X^2, X^3, etc.) can be used to represent higher order nonlinear effects of a variable such as a quadratic or cubic trend of age or time.

Researchers often struggle with the fundamental problem that the product term may be highly correlated with the first-order predictor variables from which it is derived, and that a powered term is similarly highly correlated with the original predictor variable from which it is derived. When predictor variables are correlated, the collinearity can lead to problems when estimating regression coefficients. Collinearity means that within the predictor set, one or more of the independent variables (IVs) are highly predicted by one or more of the other IVs. The collinear variables have regression coefficients that are poorly estimated and minor fluctuations in the sample, such as those caused by measurement and sampling error, have major impacts on the weights. The inherent collinearity of powered and product terms with their first-order predictor variables is problematic because it can create instability in the values for the estimated regression weights, leading to bouncing beta weights (Pedhazur, 1982).

Ideally, an interaction term is uncorrelated with (orthogonal to) its first-order effect terms. When modeling the relation between a continuous outcome variable and a categorical predictor variable or a continuous variable that has been measured discretely with a finite range, contrast coding may be used to estimate interactive effects that are either uncorrelated (when *n*s are equal) or minimally correlated (when *n*s are unequal) with first-order effects. Under orthogonal conditions, when the interaction term is entered into a model, the partial regression coefficients representing the magnitudes, directions, and significances of the main effect variables remain precisely the same as they were before the interaction was included. For powered terms, orthogonal polynomial contrast codes are available, but there is a practical limitation that the predictor variable must be measured discretely with a very limited range of values so that tables of contrast coefficients are readily available (see, e.g., Cohen, Cohen, West, & Aiken, 2003, p. 215). With continuous variable interaction terms, the orthogonality property is harder to achieve. Several sources (i.e., Aiken & West, 1991; Cohen, 1978; Cronbach, 1987) have demonstrated that if the first-order effect variables are transformed from a raw-score scaling to a deviation-score scaling by subtracting the variable mean from all observations (i.e., mean centering) the resulting product term will be minimally correlated or uncorrelated with the first-order variables depending on the proximity to bivariate normality.

Despite the equivalent partial regression coefficients representing the relative contributions of uncentered versus mean-centered first-order variables to the regression equation (see Kromrey & Foster-Johnson, 1998, for a convincing demonstration), mean centering predictor variables prior to creating interaction or product terms has two distinct advantages. First, mean centering alleviates the ill conditioning of the correlation matrix among the predictor variables that results from *nonessential multicollinearity* (Marquardt, 1980) among the first-order predictors and their interaction term (or between first-order predictors and even powered terms such as between X and X^2, X^4, or X^6). Thus, the resultant instability of regression estimates and standard errors are stable and robust (i.e., the bouncing beta weight problem is remedied).

The second advantage of mean centering concerns the interpretability of the estimates. The regression coefficient for a mean-centered predictor may be more practically meaningful than the same coefficient for the same predictor with an arbitrary zero point (i.e., interpreting the relative size of change in Y for a one-unit change in X at a given level of Z may be easier if the zero point of Z is the average value of Z rather than an arbitrary and nonmeaningful scale value). Interpretability may also be improved by plotting the predicted relation between X and Y over a range of plausible z scores (e.g., Aiken & West, 1991; Cohen et al., 2003; Mossholder, Kemery, & Bedeian, 1990).

Under most circumstances, mean centering is an adequate solution to the collinearity problem. At times, however, the mean-centered product or powered term will still have some degree of correlation with its first-order variables that can influence the partial regression coefficients. To remedy this lack of complete orthogonality with the mean-centering approach, a simple two-step regression technique called residual centering can be used that ensures full orthogonality between a product term and its first-order effects (Lance, 1988). This technique is also generalizable to powered terms.

Residual centering (i.e., orthogonalizing) is a comparable alternative approach to mean centering that also serves to eliminate nonessential multicollinearity in regression analyses. Residual centering, as originally suggested by Lance (1988), is essentially a two-stage OLS procedure in which a product term or powered term is regressed onto its respective first-order effect(s). The residuals of this regression are then used to represent the interaction or powered effect. The variance of this new orthogonalized interaction term contains the unique variance that fully represents the interaction effect, independent of the first-order effect variance (as well as general error or unreliability). Similarly, the variance of an orthogonalized powered term contains the unique variance accounted for by the curvature component of a nonlinear relation, independent of the linear components.

Residual centering, like mean centering, has a number of inherent advantages for regression analyses. First, the regression coefficients for orthogonalized product or powered terms are stable. That is, the regression coefficients and standard errors of the first-order effect terms remain unchanged when the higher order term is entered. Second, the significance of the product or powered term is unbiased by the orthogonalizing process. Third, unlike mean centering, orthogonalizing via residual centering ensures full independence between the product or powered term and its constituent main effects.[1]

A Simple Regression Example

For this example, two concepts from the area of perceived control and their interaction are used to predict well-being. The first concept, agency for ability, is the degree of belief that one is smart and intellectually able. The second concept is one's degree of belief that successful intellectual performance comes about because of unknown causes. In this example, ample research supports the hypothesis that agency for ability is positively related to well-being or positive affect (for a review, see Skinner, 1995). Depending on the degree to which one believes that unknown

[1] In principle, orthogonalizing is not limited to use on variables that are mathematical manipulations of another variable. The technique can be readily applied on substantively derived variables. For example, Little, Brauner, Jones, Noch, and Hawley (2003) used orthogonalizing to separate two sources of variance from a multidimensional variable.

factors are also responsible for successful performance, this positive relation would be moderated. The reason for this expected interaction is simply that the more one believes unknown factors are responsible for performance, the more that the well-being-enhancing effects of agency would be undermined. Specifically, it was expected that high beliefs in unknown causes would reduce the otherwise positive relation of agency for ability with positive affect.

To test this hypothesis and to compare methods, two product terms for the interaction between agency for ability (agency) and one's belief in unknown causes (causes) were created:

$$INT_U = \text{agency} * \text{causes} \tag{2}$$

$$INT_{MC} = (\text{agency} - \text{mean of agency}) * (\text{causes} - \text{mean of causes}) \tag{3}$$

The interaction term in Equation 2, INT_U, is the product of the uncentered first-order effect variables. The interaction term in Equation 3, INT_{MC}, is the product of the mean-centered first-order effect variables. To create a third orthogonalized interaction term, INT_O, INT_U was regressed on both of the untransformed first-order effect variables (agency and causes), and the residuals were saved as a new variable (INT_O).

Table 1 shows the correlations among the outcome variable of positive affect, the main effect variables of agency and causes, and the three alternative specifications of the interaction term. Table 2 shows the estimated parameters from the three alternative interaction model specifications. Although the correlation between the uncentered interaction term (INT_U) and the first-order effects of agency and causes is at a modest to high level (.37 and .80, respectively), the effect of this level of multicollinearity is pronounced, as shown in Table 2. The regression coefficients for the first-order effects became highly inflated in magnitude when INT_U was entered into the model. For the mean-centered term (INT_{MC}) and the orthogonalized term (INT_O), the regression coefficients for the first-order effects remained quite stable, with one notable exception. Namely, with the inclusion of INT_{MC}, the regression coefficient for causes showed a very slight drop from .05 to .04. This drop is sufficient to change the significance of the effect. Using a one-tailed test (because we know from ample research the direction of the causes effect), causes is a significant predictor of positive affect at the $p < .05$ level in the first-order effects only model as well as in the model when the orthogonalized term (INT_O) is entered. When the mean-centered term is entered (INT_{MC}), causes fails to reach significance at the $p < .05$ criterion. Although the correlation of INT_{MC} with causes was only .06, this degree of collinearity is sufficient to influence a decision about the significance of the first-order effect term. Thus, even this simple example demonstrates that there remains some influence on the regression coefficients between mean-centered predictors and their interaction term when the mean-centered term is not completely orthogonal.

TABLE 1
Simple Statistics and Pearson Correlations Among the Variables Used in the Regressions

	Positive Affect	*Agency*	*Causes*	*INT_U*	*INT_{MC}*	*INT_O*
Positive affect	1.00					
Agency	0.22	1.00				
Causes	–0.01	–0.24	1.00			
INT_U	0.10	0.37	0.80	1.00		
INT_{MC}	–0.10	0.05	–0.06	0.16	1.00	
INT_O	–0.11	0.00	0.00	0.18	0.99	1.00
M	3.03	3.12	2.02	5.35	0.06	0.00
SD	0.68	0.54	0.52	1.68	0.30	0.30

Note. INT_U is the straight product of the two first-order effects (agency and causes); INT_{MC} is the interaction term created after mean centering the two first-order effects; INT_O is the residual-centered interaction term

TABLE 2
Comparison of Results From Three Standard Regression Approaches

		Uncentered		*Mean Centered*		*Residual Centered*	
Predictor	*R_{xx}*	*Beta*	*t*	*Beta*	*t*	*Beta*	*t*
First-order effects only regressions[a]							
Agency	.66	.23	8.79	.23	8.79	.23	8.79
Causes	.74	.05	1.77	.05	1.77	.05	1.77
Adding the respective interaction terms							
Agency	.66	.59	6.81	.23	9.01	.23	8.84
Causes	.74	.62	4.63	.04	1.56	.05	1.77
Interaction	.76	–.62	–4.37	–.11	–4.37	–.11	–4.37

Note. Reliability of positive affect is .92.

[a]The first-order effects only regressions are represented three times only for pedagogic reasons. The regression was only done one time as it is the basis for each subsequent interaction test.

A SIMPLE EXTENSION TO REPRESENT A LATENT VARIABLE INTERACTION

Although mean centering and residual centering are beneficial to the interpretation of interactions in regression models, the estimation of interaction effects within regression models is still fraught with difficulty. Perhaps the largest concern is the effect of measurement error on the power to detect such effects. OLS regression assumes that all variables are measured without error, or are perfectly reliable, an assumption that is often not tenable. Violations of this assumption may result in unknown bias in the parameter estimates (Busemeyer & Jones, 1983). Although the presence of measurement error is problematic for all variables in regression, it is particularly troublesome for an interactive or nonlinear term, for which the reliability is a function of the reliability of its constituent variables. The resultant reliability for the product term is often lower than the minimum reliability of either of the first-order effects. A second, related concern is the differentiation of multiplicative and nonlinear effects under such conditions of low power. A complete discussion of this problem is beyond the scope of this article; for a more thorough treatments see Cortina (1993), Ganzach (1997), Kromrey and Foster-Johnson (1998), Lubinski and Humphreys (1990), and MacCallum and Mar (1995).

SEM represents an important advance in the study of multiplicative or nonlinear effects because of its ability to address properly the presence of measurement error within a statistical model. In SEM, the proportion of variance common to multiple indicators of a given construct is estimated and the structural relations among latent constructs may then be estimated such that the relations are estimated without the attenuating effects of measurement error. Numerous authors have described techniques to represent latent variable interactions within the context of SEM. Most approaches are based on the Kenny and Judd (1984) product indicator model and require complex nonlinear constraints (see, e.g., Algina & Moulder, 2001; Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Schumacker & Marcoulides, 1998; Wall & Amemiya, 2001). Bollen (1995, 1996) and Bollen and Paxton (1998) presented a two-stage least squares (2SLS) approach that does not require the nonlinear constraints but has been found to be less effective than other methods (Moulder & Algina, 2002; Schermelleh-Engel, Klein, & Moosbrugger, 1998). Klein and Moosbrugger (2000) proposed a latent moderated structural (LMS) model approach utilizing finite mixtures of normal distributions that was further refined by Klein and Muthén (2002) as a quasi-maximum likelihood (QML) approach. The QML approach was found to perform well under conditions where first-order indicators are normally distributed (Marsh et al., 2004). Finally, Marsh et al. (2004) proposed an unconstrained product indicator approach, which they reported performed well, especially when underlying distributional assumptions are not met.

All major SEM software programs can implement the nonlinear constraints that are necessary to implement the procedures based on the Kenny and Judd (1984) product indicator model. The 2SLS approach can be directly implemented through PRELIS, the preprocessor for LISREL (Jöreskog & Sörbom, 1996). The LMS approach is implemented directly in M*plus* version 3 (Muthén & Muthén, 1998–2004). Despite the direct software implementation of the 2SLS and LMS approaches and the widespread availability of software to implement nonlinear constraints, modeling nonlinear interactions and powered terms in the SEM context remains an extremely difficult endeavor for the average practitioner. The direct implementations in PRELIS and M*plus* make latent variable interactions more accessible, but the researcher is limited to only those two software programs. In our view, the orthogonalizing technique described here is less technically demanding than other alternative methods of including interactive and powered terms in latent variable models based on nonlinear constraints, it can be implemented in any SEM software platform, and it provides reasonable estimates that are comparable to other existing procedures.

The following sections detail the implementation of the orthogonalizing procedure and provide limited simulation evidence as to the comparability of orthogonalizing to other recently proposed procedures such as Marsh et al.'s (2004) unconstrained approach and the LMS (Klein & Moosbrugger, 2000) approach as implemented in M*plus*. Because the orthogonalizing approach and the LMS approach directly implemented in M*plus* were compared, all statistical analyses were also done in M*plus*. However, any differences in statistical outcomes that may occur when implementing the orthogonalizing approach in other software should be negligible and attributed entirely to the particular set of programming algorithms written into the particular software. M*plus* version 3 was also used to generate and analyze all simulation data.

Extending the Empirical Example to a Latent Variable Interaction

To create othogonalized indicators for a latent interaction construct, each possible product term from two sets of indicators for two latent constructs is formed (much like the Kenny & Judd [1984] approach). Building on the regression example presented in Table 2, we can disaggregate the agency and causes variables to create three unique indicators of each dimension. Specifically, we have three indicators of agency (ag_1–ag_3) and three indicators of unknown causes (uc_1–uc_3). From these indicators, nine product terms are possible:

$$aguc_{11} = ag_1 * uc_1 \tag{4}$$

$$aguc_{12} = ag_1 * uc_2 \tag{5}$$

$$aguc_{13} = ag_1 * uc_3 \tag{6}$$

$$aguc_{21} = ag_2 * uc_1 \tag{7}$$

$$aguc_{22} = ag_2 * uc_2 \tag{8}$$

$$aguc_{23} = ag_2 * uc_3 \tag{9}$$

$$aguc_{31} = ag_3 * uc_1 \tag{10}$$

$$aguc_{32} = ag_3 * uc_2 \tag{11}$$

$$aguc_{33} = ag_3 * uc_3 \tag{12}$$

Each of the resulting nine uncentered product terms is then individually regressed onto the first-order effect indicators of the constructs. For instance,

$$agu c_{11} = b_0 + b_1 ag_1 + b_2 ag_2 + b_3 ag_3 + b_4 uc_1 + b_5 uc_2 + b_6 uc_3 \quad (13)$$

where $ag_{1–3}$ and $uc_{1–3}$ represent the first-order indicators for the constructs agency and unknown causes (causes), respectively. The residual for this regression would then be saved and used as an indicator of the interaction construct. The procedure would be repeated for each of the nine uncentered product terms. The complete SAS code for this data manipulation to create the initial uncentered product terms and residual centered terms can be found in Appendix A.

The next step to representing the latent variable interaction is to include the nine orthogonalized product terms as indicators of a single latent interaction construct. Figure 1 displays the interaction plot based on OLS regression parameter estimates reported in Table 2. Figure 2 displays the SEM path diagram used to represent the latent variable interaction. The SEM parameterization shown in Figure 2 includes two distinct features. First, there is unique variance common to the nine indicators, depending on which first-order effect indicators were used to create them. Accordingly, correlations between the residual variances of the interaction indicators should be specified, such that the indicators $aguc_{11}$, $aguc_{12}$, and $aguc_{13}$ would be allowed to have correlated residuals (as derived from Equations 4–6, each contains the uniqueness of ag_1). Similarly, the indicators labeled $aguc_{21}$, $aguc_{22}$, and $aguc_{23}$ should have correlated residuals (each contains the uniqueness of ag_2; see Equations 7–9). The same pattern of expected residual correlation would be found for $aguc_{31}$, $aguc_{32}$, and $aguc_{33}$, which share ag_3; $aguc_{11}$, $ageuc_{21}$, and $aguc_{31}$, which share uc_1; $aguc_{12}$, $aguc_{22}$, and $aguc_{32}$, which share uc_2; and $aguc_{13}$, $aguc_{23}$, and $aguc_{33}$, which share uc_3.

The second feature of this approach is that the latent interaction term is not allowed to correlate with the main effect latent variables. Because the indicators of the interaction term have been orthogonalized with respect to the main effect latent

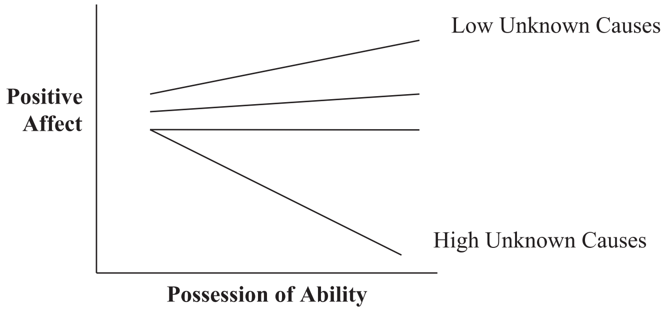

FIGURE 1 The nature of the interaction. *Note.* The slope varies from .23 to –.45 as a function of unknown causes.

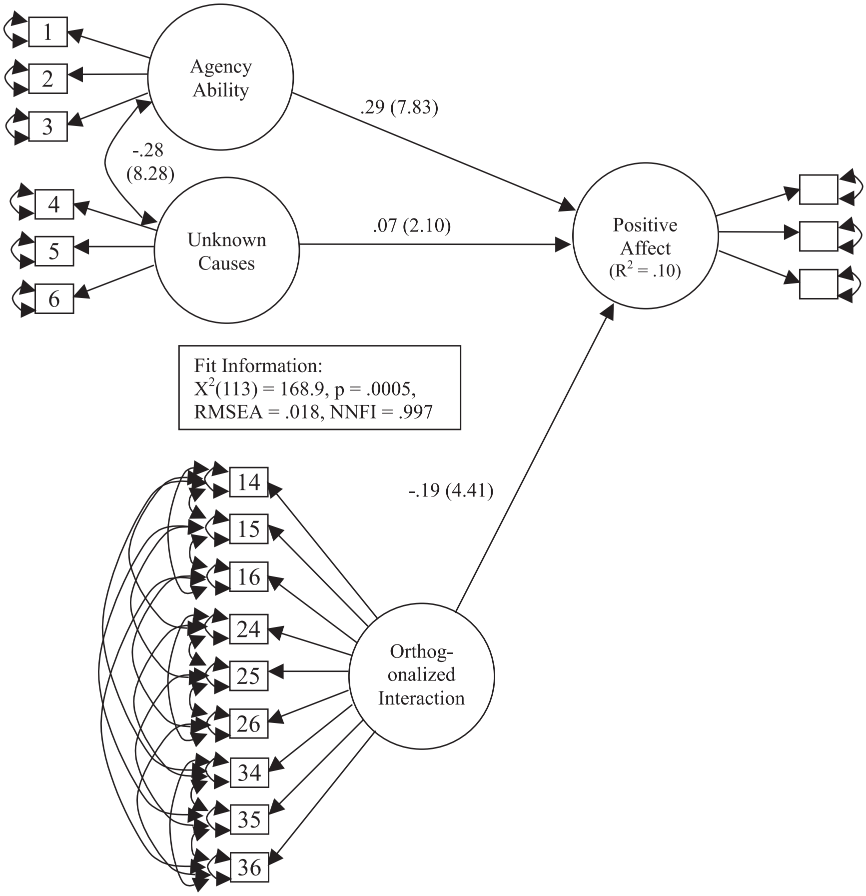

FIGURE 2 A latent variable interaction using orthogonalized indicators. *Note*. The *t* value is listed in the parentheses next to the standardized estimate.

variables, the covariance matrix to be analyzed would contain covariances of precisely zero for the 54 possible relations between the six main effect indicators and the nine interaction indicators. Appendix B contains the LISREL syntax for applying the orthogonalizing approach to the model in Figure 2. Appendix C contains the M*plus* syntax for estimating each of the alternative models.

As shown in the first column of Table 3, the results of the latent variable interaction model showed a pattern of findings that are fully consistent with those from the OLS regressions. The only differences are that the magnitudes of the regression coefficients relating positive affect to agency, causes, and their interaction are larger, and their accompanying significance values smaller than the manifest variable representation shown in Table 2. Such an effect would be expected from using multiple indicators of latent constructs to estimate structural relations disattenuated for measurement error. Also noteworthy is that the fit of the model containing the interaction term is well within acceptable ranges, a condition that is not often found with other techniques for representing latent variable interactions.

The model in Figure 2 was then reparameterized based on the Marsh et al. (2004) unconstrained approach, which is referred to as the mean-centered approach, and the Klein and Moosbrugger (2000) LMS approach. Model results are also reported in Table 3. Note that, according to all reported fit indexes, the orthogonalized approach outperforms the mean-centered approach. Also note that, in terms of the primary model parameters (i.e., latent regressions and correlations), the orthogonalized approach results in comparable parameter estimates and *t* ratios. Most traditional SEM fit indexes are not available in M*plus* when implementing the LMS approach to latent interactions, with the exception of the Akaike's Information Criterion (AIC; Akaike, 1973) and Bayesian Information Criterion (BIC; Schwartz, 1978). Both AIC and BIC are information indexes that can be used when comparing nonnested models—smaller values indicate better models both in terms of model fit and model parsimony. According to both the AIC and

TABLE 3
Comparison of Results From Three Latent Variable Interaction Approaches

	Orthogonalized		*Mean Centered*		*LMS*	
Model fit						
$\chi^2_{(113)}$	174.038		445.61			
p value	0.000		0.000			
CFI	0.994		0.966			
TLI	0.991		0.954			
AIC	38163.29		38434.89		23109.33	
BIC	38471.45		38743.05		23274.04	
RMSEA	0.019		0.044			
	b	*t*	*b*	*t*	*b*	*t*
Latent regressions						
Agency	0.291	7.877	0.296	7.981	0.304	6.822
Causes	0.075	2.158	0.071	2.046	0.070	1.908
Agency × Causes	–0.182	–4.327	–0.173	–4.08	–0.151	–4.275
	r	*t*	*r*	*t*	*r*	*t*
Latent correlations						
Agency with causes	–0.289	–8.486	–0.289	–8.494	–0.294	–7.432

Note. LMS = latent moderated structural model; CFI = comparative fit index; TLI = Tucker–Lewis Index; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root mean squared error of approximation.

BIC, the LMS approach is preferable to both the mean-centered approach and the orthogonalized approach, but all three approaches result in comparable latent parameter estimates and identical inferences.

Thus, the orthogonalized approach results in nearly identical parameter estimates as does the unconstrained mean-centered approach of Marsh et al. (2004) and the LMS approach (Klein & Moosbrugger, 2000). Although the LMS approach does result in lower AIC and BIC values, indicating a more parsimonious model, it achieves this deflated fit because the interaction construct is not estimated directly like the other two approaches and therefore far fewer parameters are estimated. Although both the mean-centered and orthogonalized approaches resulted in excellent model fit, the orthogonalized approach resulted in somewhat better fit, due to the complete orthogonality derived from residual centering that mean centering only approximates.

Simulation Evidence of the Comparability of the Three Approaches

Although the three approaches for estimating latent variable interactions reported in Table 3 produced comparable results, there was not a consistent pattern as to the variability between approaches. For instance, the LMS approach produced a larger estimate for the agency effect with a smaller *t* ratio than both the orthogonalized or mean-centered approaches and a smaller estimate for the interaction effect but a larger *t* ratio than the mean-centered approach. As a simple example of the efficacy of the orthogonalizing approach, three approaches to estimating latent variable interactions were compared: the orthogonalizing approach described in this article, the mean-centered approach described in Marsh et al. (2004), and the LMS method implemented in M*plus* version 3. The orthogonalizing and mean-centered approaches both utilized all possible cross-product indicators to infer the latent constructs.

Population model. M*plus* version 3.12 was used to generate 1,000 replications, each with a sample size of 1,500 participants. The sample size was chosen to reflect the actual sample size of the example data set used to generate the results reported in Table 3. The population (simulated) data were created to mimic the parameters reported in Marsh et al. (2004) but using the random slope parameterization implemented in M*plus* (see Muthén & Asparouhov, 2003) with

$$\begin{aligned} \eta_i &= 0.4\xi_{1i} + 0.4\xi_{2i} + s_i\xi_{1i} + \varepsilon_i \\ s_i &= 0 + 0.2\xi_{2i} + 0 \end{aligned} \tag{14}$$

where s_i is a latent variable that only contributes one additional parameter, $\beta_3 = .2$. The latent exogenous constructs, ξ_1 and ξ_2, were standard normal variables, and the correlation coefficient between ξ_1 and ξ_2 was set at .3. Three indicators were used for each of the latent variables such that y_1, y_2, y_3; x_1, x_2, x_3; and z_1, z_2, z_3 were indicators of η, ξ_1, and ξ_2, respectively. Construct identification was achieved by setting the latent variances to 1.0, and the loadings relating each indicator to its latent variable were all 0.70 in the population-generating model.

Table 4 outlines the results of the small simulation study. All replications resulted in proper solutions for all three methods. Across all 1,000 replications, all factor loadings for η, ξ_1, and ξ_2 were estimated to average 0.70 with a standard error of 0.02 for all three approaches. Factor loadings for the interaction construct ($\xi_1\xi_2$) under the mean-centered and orthogonalizing approaches were estimated to average 0.51 with a standard error of 0.02. The LMS approach does not require indicators, thus factor loadings for the interaction construct ($\xi_1\xi_2$) were not estimated. For both the mean-centered and orthogonalized approaches, 76 free parameters were estimated: 18 factor loadings, 18 residual variances, 18 residual correlations, 3 latent regression coefficients, 1 latent correlation, and 18 intercepts. The LMS approach resulted in only 31 free parameters: 9 factor loadings, 9 residual variances, 3 latent regression coefficients, 1 latent correlation, and 9 intercepts.

Model fit. Chi-square, comparative fit index (CFI), and root mean squared error of approximation (RMSEA) model fit information was only available for the mean-centered and orthogonalized approaches. With the same number of free parameters between both approaches and across 1,000 replications, orthogonalizing resulted in a smaller average chi-square value (by a ratio of 2.25) and a smaller standard deviation. Average CFI was equivalent to two decimal places between both approaches, and the average RMSEA differed by .008 when measured to three decimal places. As would be expected due to the significantly smaller number of free parameters estimated in the LMS approach, AIC and BIC values were substantially smaller for the LMS approach versus both the mean-centered and orthogonalized approaches, but orthogonalizing resulted in lower AIC and BIC values than did mean centering.

Latent coefficients. All three approaches resulted in negligibly biased parameter estimates for both the effect of ξ_1 and ξ_2 with percentage bias for ξ_1 ranging from 0.300% to 0.400% and percentage bias for ξ_2 ranging from 0.025% to 0.125%. The latent correlation between ξ_1 and ξ_2 was estimated with negligible bias across all three approaches as well, ranging from 0.2% to 0.3%.

The primary difference among the three approaches involved the interaction effect. The LMS approach, which was also the generating model, produced only 0.05% bias when estimating the interaction effect, whereas the orthogonalized and mean-centered approaches resulted in 3.9% and 4.2% bias, respectively. Although this bias seems to be a large discrepancy relative to other percentage bias estimates, the parameter estimates differed only in the second decimal place between the mean-centered or orthogonalized approaches and the LMS approach, as shown in Table 4.

TABLE 4
Comparison of Simulation Results From Three Latent Variable Interaction Approaches

	Orthogonalized			*Mean Centered*			*LMS*		
	M	*SD*	*% Bias*	*M*	*SD*	*% Bias*	*M*	*SD*	*% Bias*
Model fit									
$\chi^2_{(113)}$	56.26	11.00	—	126.81	19.54	—	N/A	N/A	—
CFI	1.00	0.00	—	1.00	0.00	—	N/A	N/A	—
AIC	51841.8	594.5	—	51912.6	594.6	—	30070.5	167.7	—
BIC	52245.6	594.5	—	52316.2	594.6	—	30235.3	167.7	—
RMSEA	0.000	0.000	—	0.008	0.006	—	N/A	N/A	—
Latent regressions									
ξ_1	0.4012	0.0382	0.300	0.4013	0.0377	0.325	0.4016	0.0377	0.400
ξ_2	0.4001	0.0368	0.025	0.4001	0.0366	0.025	0.4005	0.0365	0.125
$\xi_1 \times \xi 2$	0.2078	0.0374	3.900	0.2084	0.0374	4.200	0.2001	0.0344	0.050
Latent correlations									
ξ_1 with ξ_2	0.3006	0.0281	0.200	0.3006	0.0281	0.200	0.3009	0.0281	0.300
Standard errors									
ξ_1	0.0365	0.0382	–4.450	0.0365	0.0377	–3.183	0.0370	0.0377	–1.857
ξ_2	0.0365	0.0368	–0.815	0.0365	0.0366	–0.273	0.273	0.0365	1.370
$\xi_1 \times \xi_2$	0.0354	0.0374	–5.348	0.0354	0.0374	–5.348	0.0340	0.0344	–1.163
ξ_1 with ξ_2	0.0288	0.0281	2.491	0.0288	0.0281	2.491	0.0288	0.0281	2.491

Note. LMS = latent moderated structural model; CFI = comparative fit index; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; RMSEA = root mean squared error of approximation.

Standard errors. Also reported in Table 4, average standard errors for the latent parameters, both the effects of the latent constructs ξ_1, ξ_2, and $\xi_1\xi_2$ on η and the correlation between ξ_1 and ξ_2, were very consistent across the three procedures, deviating only in the third decimal place. However, these values, when compared to the empirically obtained standard deviation of the sampling distribution of mean estimates (i.e., the population value for the standard error), tended to be biased. All three procedures resulted in negatively biased standard errors for the effect of ξ_1 and $\xi_1\xi_2$, and both the mean-centered and orthogonalized approaches resulted in negatively biased estimates for the effect of ξ_2. All three procedures tended to overestimate the standard error for the latent correlation between ξ_1 and ξ_2. In general, the LMS approach resulted in a smaller percentage bias than either of the other two procedures, especially for the interaction effect.

DISCUSSION

As mentioned, the goals of this article were twofold. First, some of the merits of residual centering interaction and powered terms in the standard regression context were highlighted. Second, the implications of the orthogonalizing procedure to represent latent variable interactions were outlined. The proposed method for representing latent variable interactions has potential merits. First, the latent variable interaction is derived from the observed covariation pattern among all possible indicators of the interaction. Second, no constraints on particular estimated parameters need to be placed. Third, no recalculations of parameters are required. Finally, this procedure can be implemented in any standard structural model software.

The simulation portion of this article generated data based on the random slopes approach to interactions implemented in M*plus* and, as might be expected, all analyses involving the LMS approach performed better than the mean-centered or orthogonalized approach; however, the practical differences were generally trivial. However, there is a significant limitation in implementing the LMS approach—it requires the use of (a) the specific M*plus* software package, (b) specialized software developed by the originator of the approach, or (c) a specialized software developed by the researcher. Prior to Marsh et al. (2004), the only generally accessible procedures involved complex nonlinear constraints such as those described in Algina and Moulder (2001), Jaccard and Wan (1995), Jöreskog and Yang (1996), Schumacker and Marcoulides (1998), and Wall and Amemiya (2001). Like our orthogonalizing method, Marsh et al. (2004) proposed a relatively easy approach involving a simple mean centering of the indicators prior to forming the cross-products and not implementing the nonlinear constraints. The Marsh et al. approach, referred to here as a mean-centered approach, performed well relative to the comparison procedures, which included an evolution of the LMS approach termed the QML approach (Klein & Muthén, 2002). This simpler mean-centered unconstrained approach was compared with the LMS approach and the proposed orthogonalizing approach. Not only did the orthogonalizing approach perform in a comparable manner to the comparison approaches, but it performed somewhat better than the mean-centered approach.

The approach reported here, like most others, does suffer from the limitation that the standard errors, and thus, significance levels, of the parameters may be biased. Some of the other approaches, in fact, fail to give standard errors (Kenny & Judd, 1984) or, because of the significant number of constraints of estimates of the various parameters, give standard errors that are quite inflated (Jöreskog & Yang, 1996). However, the regression residuals that are used to estimate latent variable interactions from this orthogonalizing procedure are generally fairly normally distributed, thus the standard maximum likelihood estimator likely provides a reasonably robust estimate of standard errors and significance. Although this assertion should be tested more thoroughly in future work, the finding that the observed standard errors were consistent with expectations in this example is quite encouraging. Bootstrap and robust estimation procedures could also be employed to estimate more precisely the standard errors of the parameter estimates associated with the latent variable interaction construct and their accompanying significance levels.

Future work is clearly needed to examine the robustness of the standard errors under more diverse conditions. Moreover, a direct Monte Carlo comparison of all the extant methods needs to be conducted to contrast the strengths and weaknesses of all the various approaches (including alternative parameterizations; Lubinski & Humphreys, 1990) across different conditions, such as sample size, strength of effects, nonnormality, reliability of the main effect indicators, and so on. Although there have been several recent efforts on this front (e.g., Lee, Song, & Poon, 2004; Marsh et al., 2004; Moulder & Algina, 2002), none have been sufficiently comprehensive to resolve this debate.

In summary, a new technique with which to represent latent variable interactions in SEM has been proposed that is a relatively straightforward extension of orthogonalizing in regression frameworks (e.g., Lance, 1988). Moreover, this orthogonalizing technique has the potential to be extended and applied to represent other nonlinear constructs such as quadratic curvatures. The technique proposed here is a potential alternative method that a researcher can choose to implement. It is not highly demanding to specify from a technical perspective, has direct practical interpretation of parameter estimates, and its behavior in terms of model fit and estimated standard errors in this example appears to be very reasonable.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Institutes of Health to the University of Kansas through the Mental Retardation and Developmental Disabilities Research Center (5 P30 HD002528), the Center for Biobehavioral

Neurosciences in Communication Disorders (5 P30 DC005803), a new faculty grant (NFGRF 2301779) from the University of Kansas to Todd D. Little, and an Institutional National Research Service Award (5 T32 HD07525–04) to James A. Bovaird.

REFERENCES

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

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademia Kiado.

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.

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least-squares estimator. *Sociological Methodology, 25,* 223–251.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika, 61,* 109–121.

Bollen, K. A., & Paxton, P. (1998). Interactions of latent variables in structural equation models. *Structural Equation Modeling, 5,* 267–293.

Busemeyer, J. R., & Jones, L. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin, 93,* 549–562.

Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin, 85,* 858–866.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102,* 414–417.

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods, 2,* 235–247.

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., & Sörbom, D. (1996). *LISREL 8: Structural equation modeling.* Chicago: Scientific Software International.

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. 57–89). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201–210.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65,* 457–474.

Klein, A., & Muthén, B. O. (2002). *Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects.* Unpublished manuscript, Graduate School of Education, University of California, Los Angeles.

Kromrey, J. D., & Foster-Johnson, L. (1998). Mean centering in moderated multiple regression: Much ado about nothing. *Educational and Psychological Measurement, 58,* 42–67.
Lance, C. E. (1988). Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. *Applied Psychological Measurement, 12,* 163–175.
Lee, S. Y., Song, X., Y., & Poon, W. Y. (2004). Comparison of approaches in estimating interaction and quadratic effects of latent variables. *Multivariate Behavioral Research, 39,* 37–67.
Little, T. D., Brauner, J., Jones, S., Noch, M., & Hawley, P. H. (2003). A typological examination of the functions of aggression. *Merrill Palmer Quarterly, 49,* 343–369.
Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious “moderator effects”: Illustrated substantively with the hypothesized (“synergistic”) relation between spatial and mathematical ability. *Psychological Bulletin, 107,* 385–393.
MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118,* 405–421.
Marquardt, D. W. (1980). You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association, 75,* 87–91.
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.
Mossholder, K. W., Kemery, E. R., & Bedeian, A. G. (1990). On using regression coefficients to interpret moderator effects. *Educational and Psychological Measurement, 50,* 255–263.
Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling, 9,* 1–19.
Muthén, B. O., & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using Maximum-Likelihood estimation in M*plus* (M*plus* Web Notes No. 6). Retrieved August 30, 2005, from http://www.statmodel.com/download/webnotes/webnote5.pdf
Muthén, L. K., & Muthén, B. O. (1998–2004). *Mplus user’s guide* (3rd ed.). Los Angeles: Muthén & Muthén.
Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart, & Winston.
Ping, R. A., Jr. (1996a). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin, 119,* 166–175.
Ping, R. A., Jr. (1996b). Latent variable regression: A technique for estimating interaction and quadratic coefficients. *Multivariate Behavioral Research, 31,* 95–120.
Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16,* 209–222.
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 Associates, Inc.
Schumacker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6,* 461–464.
Skinner, E. A. (1995). *Perceived control, motivation, and coping.* 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,* 1–30.
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.

APPENDIX A
SAS CODE USED TO RESIDUAL CENTER THE INTERACTION TERMS AND GENERATE THE DATA FOR THE ORTHOGONALIZED LATENT VARIABLE INTERACTION FACTOR

```
DATA RC;
Set raw;
x1z1 = x1*z1; x1z2 = x1*z2; x1z3 = x1*z3;
x2z1 = x2*z1; x2z2 = x2*z2; x2z3 = x2*z3;
x3z1 = x3*z1; x3z2 = x3*z2; x3z3 = x3*z3;
RUN;
PROC REG noprint data=RC;
  Model x1z1 = x1 x2 x3 z1 z2 z3;
  Output out=RC r=o_x1z1; RUN;
PROC REG noprint data=RC;
  Model x1z2 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x1z2; RUN;
PROC REG noprint data=RC;
  Model x1z3 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x1z3; RUN;
PROC REG noprint data=RC;
  Model x2z1 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x2z1; RUN;
PROC REG noprint data=RC;
  Model x2z2 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x2z2; RUN;
PROC REG noprint data=RC;
  Model x2z3 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x2z3; RUN;
PROC REG noprint data=RC;
  Model x3z1 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x3z1; RUN;
PROC REG noprint data=RC;
  Model x3z2 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x3z2; RUN;
PROC REG noprint data=RC;
  Model x3z3 = x1 x2 x3 z1 z2 z3;
  Output out= RC r=o_x3z3; RUN;
```

APPENDIX B
LISREL 8.X SYNTAX FOR ESTIMATING LATENT VARIABLE INTERACTION TERMS

Orthogonalizing Approach

This model was specified using all Y-side matrices because the distinction between X and Y is simply a theoretical distinction that does not impact the estimates if the variables are specified in LISREL as Y variables. Conceptually, the first three latent variables can be thought of as exogenous and are treated as such because no other variables predict them even though they are listed on the Y side. However, it is just as easy to consider the first three latent variables as endogenous to some unknown exogenous factors, and therefore, they can be specified on the Y side with the idea that future research may identify and include endogenous factors or potential covariate effects as X-side, exogenous factors.

```
DA NI=18 NO=1503 MA=CM
LA
   x1 x2 x3 z1 z2 z3
   o_x1z1 o_x1z2 o_x1z3 o_x2z1 o_x2z2 o_x2z3 o_x3z1 o_x3z2 o_x3z3
   y1 y2 y3
me=means.dat
sd=stdev.dat
km=matrix.dat
MO NY=18 NE=4 LY=FU,FI PS=SY,FI BE=FU,FI TE=SY,FI
FR LY(1,1) LY(2,1) LY(3,1)
FR LY(4,2) LY(5,2) LY(6,2)
FR LY(7,3) LY(8,3) LY(9,3) LY(10,3) LY(11,3) LY(12,3) LY(13,3) LY(14,3) LY(15,3)
FR LY(16,4) LY(17,4) LY(18,4)
VA 1 PS(1,1) PS(2,2) PS(3,3) PS(4,4) !identifying and establishing scale
FR PS(2,1)
FR BE(4,1) BE(4,2) BE(4,3)
FR TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6) TE(7,7)
FR TE(8,8) TE(9,9) TE(10,10) TE(11,11) TE(12,12) te(13,13)
FR TE(14,14) TE(15,15) TE(16,16) TE(17,17) TE(18,18)
FR TE(8,7) TE(9,7) TE(9,8)
FR TE(11,10) TE(12,10) TE(12,11)
FR TE(14,13) TE(15,13) TE(15,14)
FR TE(10,7) TE(13,10) TE(13,7)
FR TE(11,8) TE(14,11) TE(14,8)
FR TE(12,9) TE(15,12) TE(15,9)
LE
   X Z XZ Y
OU ND=3 SO SC
```

Mean-Centered Approach

The basic syntax for estimating a latent variable interaction using either the orthogonalizing approach proposed in this article or the mean-centered approach used for comparison purposes is identical. The difference between the two procedures lies in the data that are used for input. The data for the orthogonalized approach would include product indicator variables for the interaction factor that have been residual centered as described in Appendix A. The data for the mean centered approach would include interaction indicators that have been mean centered.

LMS Approach

As of this writing, it is not possible to implement the LMS approach using LISREL.

APPENDIX C
M*plus* 3.12 SYNTAX FOR ESTIMATING LATENT VARIABLE INTERACTION TERMS

Orthogonalizing Approach

The following M*plus* syntax examples begin with the ANALYSIS command. It is assumed that the reader is familiar with specifying the TITLE, DATA, and VARIABLE commands in M*plus*.

```
ANALYSIS: TYPE = MEANSTRUCTURE;
          ESTIMATOR = ML;
MODEL:    X BY x1* x2* x3*;
          Z BY z1* z2* z3*;
          Y BY y1* y2* y3*;
          INT BY x1z1* x1z2* x1z3* x2z1* x2z2* x2z3* x3z1* x3z2* x3z3*;
          X@1;Z@1;Y@1;INT@1;
          X WITH Z*;
          X WITH INT@0;
          Z WITH INT@0;
          Y ON X Z INT;
          x1z1 WITH x1z2* x1z3* x2z1* x2z2@0 x2z3@0 x3z1* x3z2@0 x3z3@0;
          x1z2 WITH x1z3* x2z1@0 x2z2* x2z3@0 x3z1@0 x3z2* x3z3@0;
          x1z3 WITH x2z1@0 x2z2@0 x2z3* x3z1@0 x3z2@0 x3z3*;
          x2z1 WITH x2z2* x2z3* x3z1* x3z2@0 x3z3@0;
          x2z2 WITH x2z3* x3z1@0 x3z2* x3z3@0;
          x2z3 WITH x3z1@0 x3z2@0 x3z3*;
          x3z1 WITH x3z2* x3z3*;
          x3z2 WITH x3z3*;
          [y1 y2 y3 x1 x2 x3 z1 z2 z3 x1z1 x1z2 x1z3 x2z1 x2z2 x2z3
          x3z1 x3z2 x3z3];
```

Mean-Centered Approach

As was the case in Appendix B, the only functional difference between implementing the orthogonalizing approach and the mean-centered approach is in the input data file(s). The orthogonalizing approach requires residual-centered product indicators for the interaction term, whereas the mean-centered approach requires mean-centered product indicators.

LMS Approach

The following M*plus* syntax examples begin with the ANALYSIS command. It is again assumed that the reader is familiar with specifying the TITLE, DATA, and VARIABLE commands in M*plus*.

```
ANALYSIS: TYPE = RANDOM;
          ALGORITHM = INTEGRATION;
MODEL:    X BY x1* x2* x3*;
          Z BY z1* z2* z3*;
          Y BY y1* y2* y3*;
          XZ | X XWITH Z;
          Y ON X Z XZ;
          X@1; Y@1; Z@1;
```

APPENDIX D
M*plus* 3.12 SYNTAX FOR MONTE CARLO DATA GENERATION

Data Generation

```
TITLE:    Data generation syntax.
MONTECARLO:
          NAMES ARE y1 y2 y3 x1 x2 x3 z1 z2 z3;
          NOBSERVATIONS = 1500;
          NREPS = 1000;
          SEED = 53487;
          REPSAVE = ALL;
          SAVE = C:\data*.txt;

ANALYSIS: TYPE = RANDOM;
          ALGORITHM = INTEGRATION;
MODEL MONTECARLO:
          [x1-x3*3 z1-z3*3 y1-y3*3];
          x1-x3*.3; z1-z3*.3; y1-y3*.3;
```

```
X BY x1*.7 x2*.7 x3*.7;
Z BY z1*.7 z2*.7 z3*.7;
Y BY y1*.7 y2*.7 y3*.7;
X@1; Y@1; Z@1;
XZ | X XWITH Z;
Y ON X*.4 Z*.4 XZ*.2;
Z WITH X*.3;
```

Data Manipulation

A SAS macro was then used to modify each data set to match the input requirements for the particular comparison approach. Thus, two additional versions of each of the 1,000 replications were created. The original simulated data were used for the LMS approach. A second set of replications were created where all X and Z variables were mean centered prior to forming all possible product indicators for the mean-centered approach. Finally, a third set of replications were created for the orthogonalizing approach where all possible X-Z product indicators were created and then residual centered as demonstrated in Appendix A.

Monte Carlo Data Analysis

Once replication data were simulated and manipulated, the following DATA command was used in conjunction with the M*plus* syntax presented in Appendix C for automated analysis of the multiple data sets and summarization of the replication results as detailed in the M*plus* manual (Muthén & Muthén, 1998–2004).

```
DATA:
FILE IS C:\datalist.txt;
      TYPE = MONTECARLO;
```