CHAPTER

17 Moderation

Herbert W. Marsh, Kit-Tai Hau, Zhonglin Wen, Benjamin Nagengast, *and* Alexandre J.S. Morin

Abstract

Moderation (or interaction) occurs when the strength or direction of the effect of a predictor variable on an outcome variable varies as a function of the values of another variable, called a moderator. Moderation effects address critical questions, such as under what circumstances, or for what sort of individuals, does an intervention have a stronger or weaker effect? Moderation can have important theoretical, substantive, and policy implications. Especially in psychology with its emphasis on individual differences, many theoretical models explicitly posit interaction effects. Nevertheless, particularly in applied research, even interactions hypothesized on the basis of strong theory and good intuition are typically small, nonsignificant, or not easily replicated. Part of the problem is that applied researchers often do not know how to test interaction effects, as statistical best practice is still evolving and often not followed. Also, tests of interactions frequently lack power so that meaningfully large interaction effects are not statistically significant. In this chapter we provide an intuitive overview to the issues involved, recent developments in how best to test for interactions, and some directions that further research is likely to take.

Key Words: Interaction effect; moderator, moderated multiple regression; mediation; latent interaction; product indicator; structural equation model

Introduction

Moderation and interactions between variables are important concerns in psychology and the social sciences more generally (here we use moderation and interaction interchangeably). In educational psychology, for example, it is often hypothesized that the effect of an instructional technique will interact with characteristics of individual students, an aptitude-treatment interaction (Cronbach & Snow, 1979). For example, a special remediation program developed for slow learners may not be an effective instructional strategy for bright students (i.e., the effect of the special remediation "treatment" is moderated by the ability "aptitude" of the student). Developmental psychologists are frequently interested in how the effects of a given variable are moderated by age in longitudinal or cross-sectional studies (i.e., effects interact with age or developmental status). Developmental psychopathologists may also be interested to know whether a predictor variables is, in fact, a risk factor, predicting the emergence of new symptoms and useful in preventive efforts, or an aggravation factor, mostly useful in curative efforts (i.e. effects interact with the baseline level on the outcome in longitudinal studies; Morin, Janoz, & Larivée, 2009). Social psychologists and sociologists are concerned with how the effects of individual characteristics are moderated by groups in which people interact with others. Organizational psychologists study how the effects of individual employee characteristics interact with workplace characteristics. Personnel psychologists want to know whether a selection test is equally valid at predicting work performance for different demographic groups. Fundamental to the rationale of differential psychology is the assumption that people differ in the way that they respond to all sorts of external stimuli.

Many psychological theories explicitly hypothesize interaction effects. Thus, for example, some forms of expectancy-value theory hypothesize that resultant motivation is based on the interaction between expectancy of success and the value placed on success by the individual (e.g., motivation is high only if both probability of success and the value placed on the outcome are high). In self-concept research, the relation between an individual component of self-concept (e.g., academic, social, physical) and global self-esteem is sometimes hypothesized to interact with the importance placed on a specific component of self-concept (e.g., if a person places no importance on physical accomplishments, then these physical accomplishments—substantial or minimal—are not expected to be substantially correlated with self-esteem). More generally, a variety of weighted-average models posit—at least implicitly—that the effects of each of a given set of variables will depend on the weight assigned to each variable in the set (i.e., the weight assigned to a given variable interacts with the variable to determine the contribution of that variable to the total effect).

Interaction or moderation can be seen as the opposite of generalizabilty. For example, if the effect of an intervention is the same for males and females, it is said to generalize across gender. However, if the effect of the intervention differs for men and women, then it is said to interact with gender.

Classic Definition of Moderation

In their classic presentation of moderation, Baron and Kenny (1986, p. 1174) defined a moderator variable to be a "variable that affects the direction and/or strength of the relationship between an independent or predictor variable and a dependent or criterion variable." An interaction occurs when the effect of at least one predictor variable and an outcome variable is moderated (i.e., depends on or varies as a function of) by at least one other predictor. Moderation studies address issues like "when (under what conditions/situations)" or "for whom" X has a stronger/weaker (positive/negative) relation with or effect on Y.

Consider the effect of a variable X_1 on an outcome Y: If the effect of X_1 on Y is affected by another variable X_2, then we say X_2 is a moderator; the relation between X_1 and Y is moderated by X_2. Under this circumstance, the size or direction of the effect of X_1 on Y varies with the value of the moderator X_2. As used here, X_1 and X_2 are symmetrical and can be interchanged, such that either of them can moderate the effect of the other. However, depending on the design of the study, the goals of the research, or the specific research questions, it might be reasonable to designate one of the predictor variables to be a moderator variable. Implicit in the discussion of interaction effects is the assumption that the outcome variable is determined, at least in part, by a combination of the main effects of the two predictor variables and their interaction.

Moderators can be categorical variables (e.g., gender, ethnicty, school type) or continuous variables (e.g., age, years of education, self-concept, test scores, reaction time). They can be a manifest observed variable (e.g., gender, race) or a latent variable measured with multiple indicators (e.g., self-concept, test scores). Different analytic methods of testing interactions are associated with the different types of moderators. In tests of interaction effects, the interaction term is typically the product of two variables and treated as a separate variable (e.g., the X_1X_2 interaction is often denoted as X_1-by-X_2 or $X_1 \times X_2$). A significant interaction effect indicates that the simple slopes of the predictor vary when the moderator takes on different values. We begin by discussing methods for analyzing interactions between observed variables and then discuss alternative approaches to probing the meaning of these interactions. We then give a brief account of the development of more sophisticated but statistically (and theoretically) stronger models in the estimation of latent interactions.

Particularly for categorical independent variables, manifest variables as outcome, and experimental designs with random assignment to groups, ANOVA is commonly used to evaluate interactions. The initial tests of these interactions are performed almost completely automatically by statistical packages with little intervention by the researcher. Even here, however, probing the appropriate interpretation and meaning of statistically significant interactions requires careful consideration. In contrast to ANOVAs, interaction terms in regression analyses are usually based on *a priori* theoretical predictions. Although statistically all ANOVA models can be respecified as multiple regression models (e.g., by using dummy variables), ANOVA is appropriate when all the independent variables are categorical with a relatively small number of levels. In the present chapter, we concentrate on the detection and estimation of interactions in regression analyses with the understanding that with appropriate coding of the different design factors, these analytical techniques can be applied equally effectively to experimental designs.

Graphs of Interaction Effects

Whatever the statistical approach used to test interaction effects, it is always helpful to graph statistically significant interaction effects. For example, suppose you want to test the prediction that there is no relation between X_1 and Y for boys but that X_1 and Y are positively related for girls. The finding that gender interacts significantly with X_1 only says that the relation between X_1 and the outcome Y is different for boys and girls but not whether the nature of this interaction is consistent with a specific *a priori* prediction. The starting point for probing the nature of this interaction is typically to graph the results.

Although we discuss more statistically sophisticated ways to probe significant interaction effects, a graph is always helpful in understanding the nature of the interaction effect. In Figure 17.1 we illustrate a number of different graphs of paradigmatic interaction effects, which are sometimes given specific labels in the literature. For purposes of the example, we can consider these as interactions in which one of the predictor variables is dichotomous (e.g., male/female; experimental/control), whereas the other predictor and the outcome variable are continuous. However, later we will describe how such graphs can be constructed when all variables are continuous.

Even with relatively simple models, interaction effects can be very diverse (Fig. 17.1). In Figure 17.1 we have plotted different forms of interactions in which X_1 is a continuous predictor variable and X_2 (the other predictor, the moderating variable) is a dichotomous grouping variable (here labelled as boys and girls). The key distinguishing feature is that the lines (regression plots) for boys and girls are parallel in the first graph (indicating no interaction), whereas they are significantly non-parallel for all the other graphs.

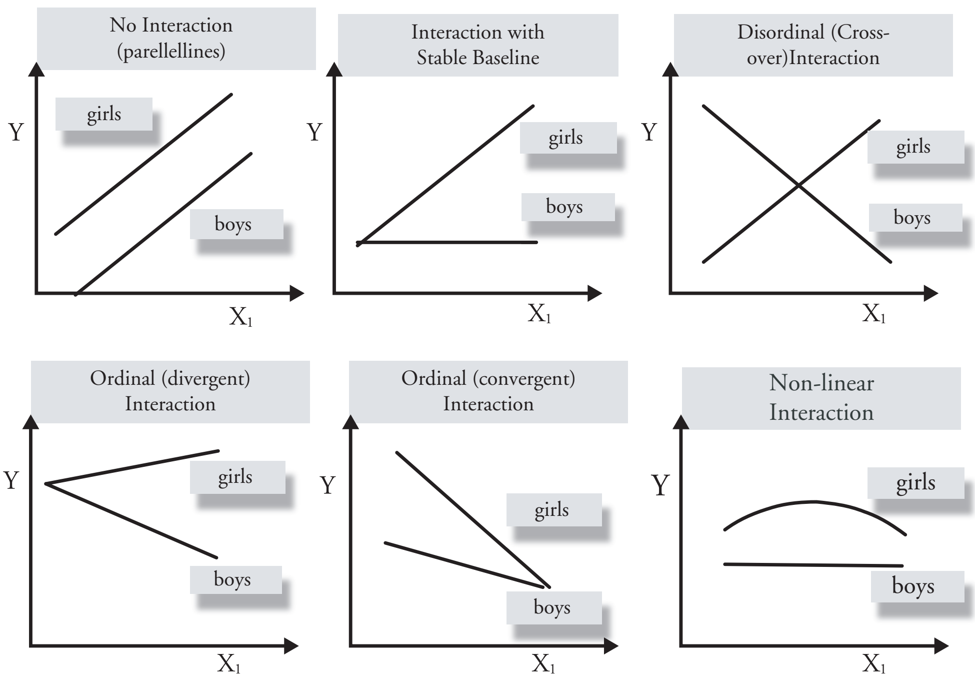

Figure 17.1 Diverse hypothetical outcomes testing whether the relation between a continuous predictor variable (X_1) and a continuous outcome variable (Y) varies as a function of the dichotomous grouping variable, gender (X_2, the moderating variable). The distinguishing feature is that the line (regression plots) for boys in girls are parallel in the first graph (indicating no interaction), whereas they are not parallel for any of the other graphs. Regression plots that contain only linear terms necessarily result in graphs that are strictly linear. However, the final graph illustrates an interaction that is nonlinear in relation to X_1 (i.e., the differences between the two lines is small when X_1 is large or small, but larger when X_1 is intermediate).

Specific forms of interactions are sometimes referred to by different names. In particular, it is typical to distinguish between disordinal graphs, where the lines cross, and ordinal graphs, where the lines do not cross, for the range of possible or plausible values that the predictor variables can take on. It is important to note that these graphs represent the predicted values from the regression model used to test the interaction effect (and should be limited to a range of values for X_1 and X_2 that are plausible and actually considered in the analyses). Thus, for example, all but the last graph are strictly linear in that only the linear effects of X_1 were included in the model (because X_2 is dichotomous, it can only have a linear effect). However, the final model includes a nonlinear (quadratic) component of X_1; the difference between the two lines is small when X_1 is small, large when X_1 is intermediate, and small when X_1 is large. Because of the nature of this interaction in this last model, it is likely that there would have been no statistically significant interaction if only the linear effects of X_1 were considered. It would also be possible to control for one or more covariates in any of these models, and this typically would change the form of the interaction.

Plotting regression equations like those in Figure 17.1 is a good starting point in the interpretation of statistically significant interactions, but a casual visual inspection of the graphs is not sufficient. Particularly for studies based on modest sample sizes, researchers are likely to overinterpret the results, leading to false–positive errors. We now turn to appropriate strategies to test the statistical significance of interactions and probe their meaning.

Traditional (Non-Latent) Approaches for Observed Variables

In this section, we introduce analytical methods for interaction models, depending on the nature of the variables. The critical feature common to all these approaches is that all the constructs are presented by a single indicator. Hence, these are not latent variable models in which the constructs are represented by multiple indicators (which we will discuss latter).

Interactions between Categorical Variables: Analysis of Variance

When the independent variables X_1 and X_2 are both categorical variables that can take on a relatively small number of levels and the dependent variable is a continuous variable, interaction effects can be easily estimated with traditional analysis of variance (ANOVA) procedures (for more general discussions of ANOVA, *see* classic textbooks such as Kirk, 1982; *see also* Jaccard, 1998 and Chapter XXX in the present Handbook). In the simplest factorial design, both X_1 and X_2 have two levels (i.e., a 2×2 design). In addition to the main effects of X_1 and X_2, this factorial ANOVA provides a test of the statistical significance of the interaction between X_1 and X_2.

AQ: Please provide chapter number

The null hypothesis for the interaction effect is that the effect of neither predictor variable (X_1 or X_2) depends on the value of the other. More complex factorial designs can have more than two levels of each variable and more than two predictor variables. There can also be higher-order interactions involving more than two variables (e.g., the nature of the X_1X_2 interaction depends on a third variable, X_3).

Although ANOVAs are typically used in experimental studies in which participants are randomly assigned to different levels of a grouping variable (e.g., experimental and control), this design can also be evaluated with more general regression approaches, representing the factors with dummy or indicator variables. Likewise, nonexperimental studies with only categorical predictor variables (e.g., gender) can be analyzed with either ANOVA or regression approaches.

The deceptive ease with which ANOVAs can be conducted has tempted researchers to transform reasonably continuous variables into a few discrete categories so that they can be evaluated with traditional ANOVA approaches. For example, with independent variables that are originally reasonably continuous, it might be possible to divide one or both of the variables into two (e.g., using a mean or median split) or a small number of categories (e.g., low, medium, high groups) so that they can be evaluated with ANOVAs. There are, however, potentially serious problems that dictate against this strategy, including (1) a reduction in reliability of the original variables, resulting in a loss of power in detecting main and particularly the interaction effects; (2) a reduction in variance explained by the original variables and particularly the interaction terms; (3) absence of a commonly used summary estimate of the strength of the interaction effect (using categories, we have t-values in different groups only; Jaccard, Turrisi, & Wan, 1990); and (4) difficulty in determining the nature of potential nonlinear relations (particularly when a continuous variable is represented by only two categories). A possible exception is that if the categorization is a natural cut-off of particular interest (e.g., minimum test scores to qualify for acceptance into a program or classification schemes). For example, Baron and Kenny (1986) discuss a threshold form of moderation such that a predictor variable has no effect when the moderator value is low, but has a positive effect when the moderator takes on a value above a certain threshold. In this case, if the threshold value is know *a priori*, it might be reasonable to dichotomize the moderator at the level of the threshold. Even here, however, there are typically stronger models that test, for example, whether the effects of the predictor variable are unaffected by variation in the moderator below the threshold or variation above the threshold. Nevertheless, the general strategy forming a small number of categories from a reasonably continuous variable should usually be avoided (for further discussion, *see* MacCallum, Zhang, Preacher, & Rucker, 2002). Although researchers have been warned of the inappropriateness ~~of such practices~~ for more than a quarter of a century (e.g., Cohen, 1978), it still persists. In summary, researchers should (almost) never transform continuous variables into discrete categories.

Interactions With One Categorical Variable: Separate Group Multiple Regression

When one independent variable is a categorical variable (e.g., X_2)—particularly with only a few naturally occurring levels (e.g., gender as in Fig. 17.1, or ethnic groups)—and the other one is a continuous variable (e.g., X_1), a possible approach is to conduct a separate regression for each group (*see also* subsequent discussion about multiple-group tests in the section on latent variable interactions). The interaction effect is represented by the differences between unstandardized regression coefficients obtained with the separate groups (Aiken & West, 1991; Cohen & Cohen, 1983; Cohen, Cohen, Aiken, & West, 2003). Assume we draw these regression equation lines: If these lines from different groups are parallel, then we conclude that there is no interaction (*see* the first graph in Fig. 17.1).

If one of the predictor variables (X_2) is dichotomous, then it is possible to test for the statistical significance of the difference between to regression coefficients relating $X_1$and Y (Cohen & Cohen, 1983, p. 56). If the hypothesis is rejected, then the interaction is significant. This approach is particularly useful if there are differences in the error variances at different levels of the moderator (*see also* subsequent discussion of latent variable tests of invariance across multiple groups). Although useful in some special cases, this multiple-group approach is typically more limited in terms of facilitating the interpretation of interaction effects, reducing power because of lower sample size in each group considered separately, and, of course, because it requires at least one of the predictor variables to be a true categorical variable. Therefore, we now present a more general approach that is appropriate when predictor variables are categorical, continuous, or a mixture of the two.

Interactions With Continuous Variables: Moderated Multiple Regression Approaches

Now we move from analyses of moderation involving categorical observed independent variables to those using continuous observed independent variables. Consider the familiar regression equation involving two predictors, X_1 and X_2:

AQ: Please clarify missing symbols throughout.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e,$$

which assumes no interaction; the effects of X_1 and X_2 are additive (Judd, McClelland, & Ryan, 2009; Klein, Schermelleh-Engel, Moosbrugger, & Kelava, 2009). That means the effect of a predictor (e.g., X_1) does not depend the value of the other (i.e., X_2) and the effects of the two predictors on Y can simply be added. The first graph in Figure 17.1 (with parallel lines) was based on a model of this form.

However, this assumption of strictly additive effects might be false. Irrespective of whether X_1 or X_2 has a main effect on Y, an interaction effect might exist in that the effect of X_1 on Y depends on the value of X_2. The mathematical equation representing this can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + e, \qquad (1)$$

where β_1 and β_2 represent the main effects of with X_1 and X_2, β_3 represents the interaction effect, X_1X_2 is the product of X_1 and X_2 (the interaction term), and e a random disturbance term with zero mean that is uncorrelated with X_1 and X_2.

Suppose that

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2$$

represents the estimated values of Equation 1. When X_1 is treated as the moderator, it can also be expressed as:

$$Y = (b_0 + b_1 X_1) + (b_2 + b_3 X_1) X_2,$$

where the intercept $(b_0 + b_1X_1)$ as well as the slope $(b_2 + b_3X_1)$ of X_2 are a linear functions of X_1. Similarly, the equation can be represented as:

$$Y = (b_0 + b_2X_2) + (b_1 + b_3X_2)X_1,$$

with the intercept and effect (slope) of X_1 on Y being moderated by X_2. With simple derivations using either X_1 or X_2 as the moderator, it can be shown that the interactive effect can be operationalized by adding a product term X_1X_2 into the regression equation (for details, *see*, for example, Jaccard et al., 1990; Judd, McClelland, & Ryan, 2009). This shows that the moderating relation is symmetric in that if the effect of X_1 on Y depends on the value of X_2, then the effect of X_2 on Y also depends on the value of X_1. Although this equation only represents the linear effects of X_1, X_2, and X_1X_2, it could easily be expanded to include nonlinear components of X_1 and X_2 as well as additional covariates. Although this means that statistically either one of the variables forming the interaction effect can be treated as the moderator, this choice should be guided by the design of the study or substantive theory.

The variables in Equation 1 can also be centered ($M = 0$, as the variable mean is subtracted from each value; *see* Equation 2).

$$Y_C = \beta_{C0}+\beta_{C1}X_{1C}+\beta_{C2}X_{2C}+\beta_{C3}X_{1C}X_{2C}+e_C \quad (2)$$

Standardized estimates are the parameter estimates that are obtained when all independent and dependent variables in the regression model are standardized (i.e., z-scores; subtracting the variables with their respective means and then divided by their respective standard deviations).

$$Z_Y = \beta_{Z0}+\beta_{Z1}Z_{X1}+\beta_{Z2}Z_{X2}+\beta_{Z3}Z_{X1}Z_{X2}+e_Z \quad (3)$$

The centered and standardized form of the regression equation that are based on centered and standardized variables, respectively, are given by where the subscripts Z and C on the regression weights are used to indicate that these parameter estimates based on centered scores (Equation 2) and standardized scores (Equation 3) are not the same as the corresponding parameter estimates based on raw scores (Equation 1).

For the raw score regression (Equation 1), β_0 is the *intercept* or the predicted value of Y when X_1 and X_2 equal 0. The intercept (β_0) might be meaningless if X_1 and X_2 never take on a value of 0. However, when the predictor variables are centered or standardized (Equation 2 or Equation 3), the intercept (β_{C0} or β_{Z0}) is the value of Y at the mean of X_1 and X_2, which is typically meaningful.

β_1 is the estimated change in Y associated with 1 unit of change in X_1 when $X_2 = 0$ (i.e., the slope of the relation between X_1 and Y when $X_2 = 0$). Again this may be meaningless in the raw score equation. However, with centered or standardized predictors, it is the association between X_1 and Y at the mean of X_2. With centered predictors, β_{C1} represents the change in the outcome if X_1 changes one unit in the raw metric. If the predictors are standardized, then β_{Z1} represents the change in Y in standard deviation (SD) units if X_1 changes one SD. Although main effects should always be interpreted cautiously when there is an interaction, this is typically a meaningful result in the standardized equation. Because the interaction is symmetric in relation to X_1 and X_2, β_2 is merely the change in Y associated with a 1-unit change in X_2 when $X_1 = 0$. For Equations 1 and 2, the 1-unit change is in the original (raw score) metric, whereas in Equation 3 it is in standard deviation units. If the units of the predictor variables are in a meaningful metric, it might be useful to use the centered predictor variables, as changes are then in the same metric as the original variables. However, particularly when the metric is arbitrary, it often is more meaningful to standardize the variables so that changes are in terms of standard deviation units. However, when one of the predictor variables is a categorical variable, it is traditional to use values of 0 and 1 (or appropriate dummy variables if there are more than two categories).

β_3 represents the interaction effect, the amount by which the effect of X_1 on Y changes with a 1-unit increase in X_2. Equivalently, because of the symmetry of the interaction effect, it is the amount by which the effect of X_2 on Y changes with a 1-unit increase in X_1. In the raw and centered form, this value may or may not be interpretable, depending on the nature of the metric of X_1 and X_2. In the standardized form, the interpretation of β_{Z3} is the same, with the exception that all changes are in terms of SD units.

It is important to understand the relation between the regression equation and graphs like those presented in Figure 17.1. Regression plots are constructed by substituting values for the X_1 and X_2 predictor variables into the regression equation. In this regard, the regression plots represent the ***predicted*** values based on the model that is being tested, not the raw data. Thus, for example, even if there is nonlinearity in the raw data, this will not be reflected in regression plots based on a model with no nonlinear terms. To address this issue, researchers sometimes include a scatterplot of the individual cases (or the raw mean values for categorical variables) as well as the regression plot.

The regression plot can be based on raw scores, centered scores, or standardized scores. In terms of plotting, it is, of course, critical that values plotted are the same as those used to estimate the regression parameters. Hence, if predictor variables are standardized in the regression, then the standardized values should be used to construct the plot. However, it is reasonable to construct different plots in relation to raw, centered, or standardized scores—whichever is most appropriate—as the shape of the plot will be similar except for the metric of the axes of the graph.

When there are more than two groups, typical practice is to use dummy coding (in which one group is "left out" and serves as a baseline of comparison for other groups). However, the whole range of orthogonal and nonorthogonal coding schemes (e.g., effect coding, polynomial coding, difference coding, as well as coding specific to the nature of the study; e.g., Cohen et al., 2003; Marsh & Grayson, 1994) are available to the researcher and may have strategic advantages depending on the nature of the study design, the goals of the researcher, the meaning of the variables, and so forth. Whatever coding scheme is used, it is useful that all of the predictor variables have a meaningful zero point to facilitate the interpretation of the regression equations.

When both the predictors (X_1 and X_2) are continuous variables, it is traditional to plot regression lines for at least two strategically chosen values and typically three or more values (although complicated interactions involving nonlinear components might require more than three ~~regression lines~~). The representative values are typically selected to be the mean, 1 or 2 SDs above the mean, and 1 or 2 SD below the mean. In each case, the graphs can be constructed by simply substituting these representative values into the regression equation.

In Figure 17.2 we represent some of the correspondences between the regression equation and the graphs. We have graphed only two regression lines, as would be appropriate if one of the predictor variables was dichotomous (it is also appropriate in a model with only linear effects of X_1 and X_2 but more conventional to include three lines). To facilitate interpretation, these are presented in standard deviation units (i.e., all variables are standardized before conducting the multiple regression). The regression equation is plotted for $X_1 = 0$ (the mean of X_1

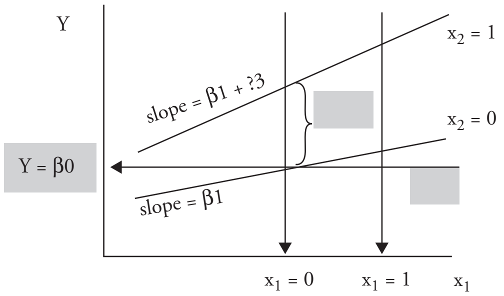

Figure 17.2 Graphical representation of the regression equation. X_1 and X_2 are predictor variables and Y is the outcome variable (to facilitate interpretation, these are presented in standard deviation units). The regression equation is plotted for $X_1 = 0$ (the mean of X_1) and $X_1 = 1$ (1 SD above the mean of 1). β_0 (the intercept) is the value that Y takes on when both $X_1 = 0$ and $X_2 = 0$. β_1 (the regression weight for X_1) is the change in Y associated with a 1-unit change in X_1 when $X_2 = 0$. β_2 (the regression weight for X_2) is the change in Y associated with a 1-unit change in X_2 when $X_1 = 0$. This form of the interaction term (the product of X_1 and X_2) is typical in social science research, but other forms of interaction could be considered.

when it is standardized) and $X_1 = 1$ (1 SD above the mean of X_1 when it is standardized). β_0 (the intercept) is the value that Y takes on when both $X_1 = 0$ and $X_2 = 0$. β_1 (the regression weight for X_1) is the change in Y associated with a 1-unit change in X_1 when $X_2 = 0$. β_2 (the regression weight for X_2) is the change in Y associated with a 1-unit change in X_2 when $X_1 = 0$. The relation between Y and X_2 when $X_1 = 1$ is given by the sum of β_2 and β_3.

There are, however, important caveats in the interpretation of the results that are highlighted by this graph. In particular, the effects of X_1 and X_2 on Y are conditional upon the value that the other variable takes on. The regression parameters β_1 and β_2 cannot be unconditionally interpreted as main effects of X_1 and X_2, respectively. This would only be possible if there were no interaction effect—that is, if $\beta_3 = 0$. The values of these regression weights represent the effect of their predictor at the point where the other predictor takes on a value of 0. In most cases in social science research, neither X_1 nor X_2 actually take on values of 0, or the value of 0 is arbitrary (i.e., ratio scales with an absolute 0 are rare). In these circumstances, it makes no sense to base substantive interpretations on the values of the regression weights unless the variables are centered or standardized. In this case, the zero point has a meaningful interpretation as the mean of the variable considered.

Some authors (e.g., Carte & Russell, 2003) are so adamant about this point that they have argued that regression coefficients should never be interpreted unless continuous X_1 and X_2 predictors are measured on a ratio scale that has an absolute 0 (for discussion of ratio scales, *see* Chapter XXX in the Handbook). However, we emphasize that regression weights can be meaningfully interpreted when predictor variables are centered or when all variables in the regression equation are standardized or, perhaps, when zero is meaningful. Thus, for example, β_2 (the regression weight for X_2) is the change in Y associated with a 1-unit change in X_2 when $X_1 = 0$. When values are standardized, this means the change in Y (in SD units) associated with a change of 1 SD in X_2 at the mean of X_1. Hence, in relation to standardized or centered values, this is an interpretable result. It is, of course, important to emphasize that the change in Y will differ when X_1 takes on different values. Similarly, for standardized values it is interpretable to report β_2, the marginal change in Y associated with a 1 SD change in X_2 at the mean of X_1. It is important that the values of the predictor variables are scaled so that 0 is a meaningful value—whether a 0 ~~to~~ 1 coding of experimental groups, a zero-centered coding of continuous predictor variables, or appropriately standardized values so that all predictor variables have $M = 0$ and SD $= 1$. However, although we argue that it is meaningful to evaluate interaction effects in relation to scores transformed to have a meaningful zero value, it is important to evaluate simple effects of the predictor variable at different levels of the moderator variable.

AQ: Please provide chapter number.

Standardized Solutions for Models With Interactions Terms Standardized estimates (*see* Equation 3) are useful for comparing results based on different variables in a standardized metric, even when the original variables are based on different, possibly arbitrary metrics. Standardizing a variable can be treated as two steps: centering (subtracting the variable with its mean) and rescaling (multiplying by a constant, e.g., meter is replaced by centimeter). Although the main effects of predictors on outcome variables are unaffected by centering in analyses for models without an interaction term, they may be affected substantially in models with interaction terms (Cohen, 1978; Cohen et al., 2003). However, centering does not affect the coefficient for the interaction term (*see* Cohen et al., 2003) or the nature (e.g., ordinal vs. disordinal) of the interaction. We also note that when predictors are not centered, product or nonlinear (e.g., X^2) terms will typically be highly correlated with the original variables, leading to multicollinearity problems such as large standard errors for the regression coefficients (Aiken & West, 1991; Cohen et al., 2003).

The computation of the appropriate standardized effects involving interaction terms in regression analyses is not straightforward with most commercial statistical packages (*see* Aiken & West, 1991; Cohen et al., 2003). In particular, the standardized regression coefficients are typically *not* correct for the interaction term. To obtain correct standardized regression coefficients (Friedrich, 1982):

1. standardize (z-score) all variables to become Z_Y, Z_{X1}, Z_{X2};
2. form the interaction term by multiplying the two standardized variables $Z_{X1}Z_{X2}$ (but do not re-standardize the product term);
3. use Z_{X1}, Z_{X2}, and Z_{X1} Z_{X2} in a regression analysis to predict Z_Y, determine the statistical significance of all predictors using their respective t-values, and report the *unstandardized* regression coefficients as the appropriate standardized coefficients. That is, the coefficients of the Equation 2. (Note: The intercept term β_0 is necessary and generally not equal to 0 because the product term $Z_{X1}Z_{X2}$ typically does not have a mean of 0).

Tests of Statistical Significance of Interaction Effects. In the application of a multiple regression model, it is straightforward to test the statistical significance of main effects and interactions by using the so-called hierarchical regression approach. This can be done in relation to raw, centered, or standardized variables:

1. For the additive model (with no interaction terms) predict Y with X_1 and X_2 and obtain the squared multiple correlation (percentage of variance explained) R_1^2;
2. For the interaction model, predict Y with X_1, X_2 and X_1X_2 using Equation 1, obtain the squared multiple correlation R_2^2 and compute the change in the squared multiple correlation, $\cancel{R_1 F_2}$

Based on Equation 1, to test whether the interaction effect is statistically significant, the hypothesis $H_0 : \beta_3 = 0$ the test statistic is

$$t = \frac{\hat{\beta}_3}{\mathrm{SE}(\hat{\beta}_3)},$$

where $\mathrm{SE}(\hat{\beta}_3)$ is the standard error of estimated β_3. More generally, when there is only one interaction term in the regression model, an equivalent test is whether R_2^2 is significantly higher than R_1^2 with the following F-test,

$$F = [(R_2^2 - R_1^2)/(k_2 - k_1)]/[(1 - R_2^2)/(N - k_2 - 1)],$$

where R_2^2 is from the equation involving the interaction term of k_2 predictors, R_1^2 is from the original equation with k_1 predictors, and N is the sample size. This is evident in that the squared t-statistic is equal to the F-statistic with $df = 1$ in the numerator. This value will be the same regardless of whether X_1 and X_2 have been centered or standardized.

The greater the partial regression coefficient β_3, and the change in R^2 (i.e., $R_2^2 - R_1^2$) resulting from the introduction of the interaction terms, the greater the moderating effect of X_1 on the relation of X_2 on Y (or symmetrically X_2 on the relationship of X_1 on Y). As pointed out by various researchers (e.g., Aiken & West, 1991; McClelland & Judd, 1993), it is important that the hierarchical regression be conducted with a model in which the interaction term is tested after controlling for both X_1 and X_2. That is, regression models testing the interaction term (X_1X_2) must also contain the corresponding predictor variables (X_1 and X_2). Otherwise the effect of the interaction is confounded with the main effects of X_1 and X_2. For models involving categorical variables in which more than one dummy variable is needed to represent one of the predictor variables (i.e., there are more than two categories), then all the related product terms must be entered simultaneously in the same regression step. Whether analyses are done in one step (including all predictor and interaction variables) or two steps (testing additive effects first and then including interaction terms), it is critical that interaction terms are evaluated in a model that contains all the additive terms for all variables in the interaction.

If additional covariates are included as control variables, they should be entered as the very first set of variables in the hierarchical regressions, followed by those involved in the interaction terms (Frazier, Tix, & Barron., 2004). However, this recommendation is based on the assumption that covariates come before the predictor and outcome variables in relation to their causal ordering. For some covariates, this is reasonable (e.g., gender, ethnicity, pretest variables), but in others it might not (*see* further discussion of mediated effects below). It is also relevant to evaluate whether the covariates really have similar effects on the outcome variable at different values of the other predictor variables by testing interactions between covariates and other variables (Cohen et al., 2003; Frazier et al., 2004).

Post hoc ***Examination of the Interactions With Continuous Observed Variables.*** Even when there is a statistically significant interaction effect, the interpretation of the values of the regression weights is hazardous, particularly when based on raw scores in the original metric. Hence, we recommend that researchers should always graph the regression equation at representative values of the predictor variables. Although this can be done in relation to the original (raw score) metric, there are sometimes interpretational advantages in constructing these graphs based on centered or standardized values. Logical questions might include (1) what is the pattern of changes in the slope (e.g., does the slope increase or decrease with increasing values of X_2)? and (2) is the regression of Y on X_1 significant at a particular value of X_2 or for a range of X_2 values?

Graphically this simple approach can be depicted by plots like those in Figure 17.2. When one of the predictor variables is a dichotomous grouping variable and the other is a continuous variable, the group-by-linear interaction can be depicted by two straight lines. Simple slopes refer to the fact that for each value of one predictor variable, it is possible to plot the relation of the other predictor variable to the outcome. This approach can still be used even when the effect of the continuous variable is nonlinear. Even when both of the variables are continuous, such plots of the effects of predictor variable at representative values of the other predictor variable provide a heuristic pictorial representation of an interaction. We generally recommend that all interactions should be represented graphically to better understand and communicate the nature of the interaction.

More formally, to explore the pattern of changing slopes in the regression of Y on X_1 at different values of X_2, it is customary to plot at least two or more of these lines (Y against X_1—i.e., simple slopes) at specific values of X_2 (e.g., the mean of X_2 and $X_2 = +1$SD and – 1SD; Aiken & West, 1991; Cohen et al., 2003). Mathematically, we can substitute X_2 (or symmetrically X_1) with values at –1SD, 0, $+1$SD from its mean into the regression equation and manually calculate the appropriate t-values for the respective regression weights for X_1 (*see* Aiken & West, 1991, for an illustrated example). It is also possible to plot simple slopes either in terms of the original regression equation or the completely standardized regression equation, whichever is easier and most interpretable. Based on values provided from most computer packages, it is also possible to compute standard errors and appropriate t-values to test the statistical significance of slopes for a predictor at a given value of the moderator or to test the difference in slopes at two different values of the moderator variable (for further discussion, *see* Aiken & West, 1991; Darlington, 1990; Jaccard et al., 1990). However, a possibly more expedient approach is to center the value of the moderator at the desired value and re-run the regression model with the new terms (i.e., the newly centered moderator variable and corresponding new cross-product terms based on the newly centered moderator variable). As is always the case, the effect of X_1 can be interpreted as its effect when $X_2 = 0$. Hence, if X_2 is centered on a particular value of interest, the test of statistical significance of X_1 provides a test of the simple slope of X_1 at that value of X_2. A similar logic can be used with categorical variables by choosing different categories as the reference (or "left-out") category that is assigned a value of zero. Thus, for example, when X_2 is gender, with boys $= 0$ and girls $= 1$, the test of statistical significance of X_1 provides a test of the simple slope of X_1 for boys. However, redoing the analysis with boys $= 1$ and girls $= 0$ provides a test of the simple slope for girls. (For further discussion, *see* Cohen et al., 2003).

The Johnson-Neyman approach is a more general alternative to the simple slopes approach that is particularly relevant when both predictor variables are continuous (Johnson & Neyman, 1936; Potthoff, 1964). This approach is used to define *regions of significance* to represent the range (or ranges) of values of one predictor variable for which the slope of the other predictor variable is significantly different than zero. Thus, a *region of significance* is the range of X_2 (moderator) values for which the relation between X_1 (predictor) and Y are statistically significant. For each region of significance, there is at most one upper and one lower bound, although one or the other of these values might be outside of the range of possible values of the moderator (or even take on the value of infinity) so that there is effectively only one bound. Alternatively, regions of nonsignificance are the ranges of values of one predictor variable for which the slope relating the other predictor variable is not significantly different from 0. As noted by Preacher, Curran, and Bauer (2006; *see also* Dearing & Hamilton, 2006), the logic of regions of significance is the converse of that used to interpret simple slopes. Simple slopes identify a slope coefficient (and its standard error, confidence interval, and statistical significance) at chosen values of the moderator. In contrast, regions of significance provide the range of values of the moderator for which the simple slopes are significant.

The manual calculation and plotting involved in probing significant interactions (e.g., simple slopes and regions of significance) is tedious and prone to error. Fortunately, Preacher et al. (2006) have provided a useful online program that helps the examination of two- and three-way interactions in multiple regression, multilevel modeling, and latent curve analysis (*see* Appendix of this chapter for more detail). Users just input regression coefficients, the variances/covariances of these coefficients, degrees of freedom, and α levels. Simple slopes and intercepts for conditional values of the moderator at predetermined points of interest, their standard errors, critical ratios, p-values, as well as confidence bands around them are either provided or can be easily generated (*see* Appendix for an example of this presented in detail).

Based on the tools introduced by Preacher et al. (2006; *see also* Appendix of this chapter for more detail) we constructed a graph of simple slopes (Fig. 17.3A) and regions of significance (Fig. 17.3B). The graph of relations between X_1 (a predictor variable) and Y (the outcome variable) at different values of X_2 (the moderator variable) shows that this is a disordinal interaction (Fig. 17.3A) in that the three regression plots cross (at the "point of intersection") within the range of X_1 and X_2 values that were considered. The effect of X_1 on Y is very positive for $X_2 = -1$, less positive for $X_2 = 0$ (the mean of the moderator variable X_2), and close to 0 for $X_2 = +1$. Figure 17.3B represents the relation between the moderator (X_2) and the simple main slope of the relation of X_1 on Y—the regression plot that is a solid dark-gray line. The relation between the simple slope and the moderator is negative; the simple slope is positive for sufficiently low values of X_2, 0 for some intermediate value of X_2, and negative for sufficiently high values of X_2. The curved lines on either side of the regression line are the 95% confidence bands for the simple slope at each value of X_2. The solid dark line for the simple slope of 0 represents the values of X_2 for which X_1 is not related to Y; this occurs at a value of $X_2 = 0.87$. The 95% confidence interval around the value of X_2 where the simple slope is 0 is the "region of nonsignificance" (represented by the two horizontal dashed lines—that is, the values for X_2 for which the simple slope of X_1 on Y is not significantly different from 0; $X_2 = +0.55$ to $X_2 = +1.40$ in our example). There are two regions of significance; the effect of X_1 on Y is positive for

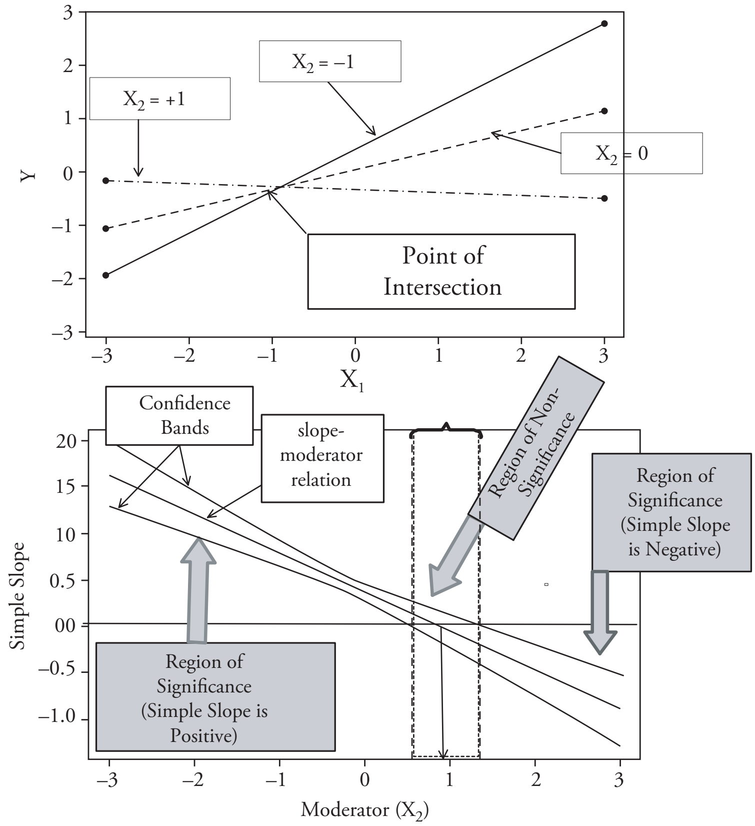

Figure 17.3 Plot of a Two-way interaction (**A**) and regions of significance and nonsignificance (**B**). Y = outcome variable, X_1 = predictor variable, X_2 = moderator. In (**A**), simple slopes are represented as regression plots of the relation between X_1 and Y at different values of X_2. (**B**) shows the relation between the simple slope (relating X_1 to Y) and the moderator X_2. Regions of significance show the ranges of values of X_2 for which the simple slope is statistically significant (i.e., X_1 is significantly related to Y), whereas regions of nonsignificance show the range for the values of X_2 for which the simple slope is not statistically significant.

values of X_2 less than $+0.55$ but negative for values of X_2 greater than 1.40.

Point of Intersection for Disordinal Interactions. For a regression equation with a significant interaction, the simple slopes may cross within or outside the possible (or plausible) range of the predicting variables (e.g., the disordinal interaction in Fig. 17.1). In an ordinal interaction the set of lines merely fan out or fan in with the nominal crossover point outside of the range of plausible values of X_1 (*see* ordinal interaction in Fig. 17.1). Thus, for example, for the disordinal interaction in Figure 17.1, the point of intersection represents the value for X_1 at which gender (X_2) has no effect on the outcome variable (i.e., where the lines cross so that scores for boys and girls are the same). In this graph, for all values of X_1 to the right of the intersection (i.e., more positive values of X_1), girls score higher than boys. Depending on the size of the confidence intervals, for some value of X_1 sufficiently above the intersection, girls have outcomes significantly higher than boys. Likewise, for values X_1 below the intersection, boys have higher values than girls, and this difference is statistically significant for values of X_1 less than the lower bound of the region of nonsignificance.

For disordinal interactions where all the effects are linear, there can be only one region of nonsignificance and at most two regions of significance. For ordinal interactions based on linear effects, it is possible to have no regions of nonsignificance and no more than one region of significance. However, for nonlinear effects, it is possible to have disordinal interactions with more than one point of intersection and multiple regions of significance and nonsignificance.

For a disordinal interaction based on linear effects, we can determine mathematically the intersection or crossing point. With some simple algebra based on the appropriate regression equation (Aiken & West, 1991), it can be shown that when X_1 is the moderator, $X_2 = -b_1/b_3$ is the intersection and that when X_2 is the moderator, $X_1 = -b_2/b_3$ is the intersection. For example, in Figure 17.3A, $b_2 = -0.355$ and $b_3 = -0.397$ so that the point of interaction is .89 ($-b2/b3 = -(-0.355/-0.397) = -0.89$

Power in Detecting Interactions. Using standard multiple regression procedures, it is not difficult to test the statistical significance of interaction effects. Nevertheless, McClelland and Judd (1993) emphasized that researchers have had great difficulty in identifying substantively meaningful, statistically significant interactions that are replicable—particularly in nonexperimental research. Reasons for this situation include (1) overall model error is generally smaller in controlled experiments than nonexperimental studies; (2) measurement errors are exacerbated when X_1 and X_2 are multiplied to form the product term X_1X_2 (*see also* Frazier et al., 2004) and are likely to be larger in nonexperiments than in experimental studies when values are manipulated to take on a few fixed values at predetermined levels; (3) the magnitude of interactions is typically constrained in field studies in which researchers cannot assign participants to optional levels of the predictor variables; and (4) power to detect interactions is compromised because of nonlinearities of the effects of X_1 and X_2 and their interaction (e.g., products of higher order terms), which is more problematic for field studies with more levels of measurements. Finally, the values that factors in experimental studies are assigned are often selected to maximize the sizes of effects (e.g., extreme values) and to be optimal in relation to research predictions—even if not representative of values typically observed, whereas the values in nonexperimental studies are more likely to be representative of the population from which the sample was drawn.

Particularly relevant to this chapter, McClelland and Judd (1993) demonstrated that because of the reduced residual variance of the product X_1X_2 after controlling for X_1 and X_2 in typical field studies, the efficiency of the interaction coefficient estimates and its associated power are much lower. This is likely to dramatically reduce the power to detect interactions in field studies as compared to more optimal design in true experiments. Thus, for example, Aguinis (2004; McClelland & Judd, 1993) has shown that the power of the test of an interaction effect is typically less than 50% so that large sample sizes might be needed to evaluate interaction effects. The situation becomes worse when the variances of X_1 and X_2 are limited, truncated, or reduced in a particular study. For these reasons, it is reasonable to conduct a preliminary power analysis before collecting data to determine how large an N is needed to have a reasonably high probability of finding a moderately sized interaction to be statistically significant (for further discussion, *see* Cohen et al., 2003).

Before conducting research, Frazier et al. (2004) suggested that researchers ensure that they have sufficient power to detect hypothesized interactions. In particular, sample size should be sufficiently large, especially when interaction effects are likely to be small, as is typically the case. For categorical variables, it is best to have approximately similar sample sizes across subgroups. Further, if error variances within subgroups are not similar, then alternative tests of significance are needed (e.g., a multiple-group approach with heterogeneous error terms; *see* subsequent discussion). For continuous predictor and outcome measures, reliabilities should be high and samples leading to restriction of the range should be avoided. Several researchers (e.g., Frazier et al., 2004; Marsh, Wen, & Hau, 2004) have also suggested the use of latent factors in structural equation modeling (SEM) to control for measurement errors (*see* subsequent discussion of latent-variable approaches to interaction models).

Multicollinearity Involved With Product Terms. Multicollinearity results when multiple predictor variables are correlated with each other such that the values of one variable systematically vary with the other variable (Marsh, Dowson, Pietsch & Walker, 2004). Hence, multicollinearity is a function of relations among predictor variables and their relation to the outcome variable. There are three quite different meanings of multicollinearity that are often confused. The first, and the focus of this section, is when the two or more predictor variables are correlated, thus complicating the interpretation of each considered separately. This form of multicollinearity occurs in almost all studies in which the predictor variables are not experimentally manipulated so as to be orthogonal. The second meaning of multicollinearity is when the predictor variables are so highly correlated that the typical tests of statistical significance are distorted or cannot be done (e.g., matrices become nonpositive definite so they cannot be inverted). This almost never happens and so is not a major concern. [Actually this sometimes happens when researchers inadvertently use different variables that are a linear combination of each other but this should be easy to recognize.] The third meaning, perhaps, is when multicollinearity is sufficiently large that SEs of coefficients are so large that meaningful interpretations of the parameter estimates cannot be made. This is not a statistical problem *per se* but, rather, the inability to disentangle the effects of each separate variable when they are highly correlated. Although there are numerous guidelines as to how large multicollinearity has to be before it is a serious problem, these are rough rules of thumb typically based on arbitrary cut-off values (*see* discussion by Cohen et al., 2003) and not "golden rules" (Marsh, Hau, & Wen, 2004) to be followed blindly in all instances. Importantly, any non-zero correlation between predictor variables results in multicollinearity. Nevertheless, reducing multicollinearity can substantially enhance the interpretability of regression weights and, in extreme cases, affect statistical tests of these effects.

Whenever product terms are computed to represent interaction effects (e.g., X_1X_2)—and particularly nonlinear effects (e.g., X^2 to represent a quadratic effect)—there is likely to be substantial amounts of multicollinearity. That is, the product terms are likely to be substantially correlated with the variables used to construct the product terms. Cohen et al. (2003; *see also* Marquardt, 1980) have distinguished between what they refer to as essential and nonessential multicollinearity. Nonessential multicollinearity can be seen as artificial effects resulting from the scaling of predictor variables in relation to the means of the predictor variables, whereas essential multicollinearity is a function of the correlation between the predictor variables (and would be 0 if X_1 and X_2 are uncorrelated). Cohen et al. have noted that nonessential multicollinearity can be reduced substantially by centering the predictor variables (or by standardizing them).

Even when there is substantial levels of "essential" multicollinearity resulting in substantial SEs that undermine interpretability, alternative models can be specified to circumvent these problems. Thus, for example, Marsh, Downson, Pietch, and Walker (2004) reanalyzed data apparently showing that paths leading from self-efficacy to achievement (0.55, $p < 0.05$) were apparently much larger than those from self-concept (–0.05, *ns*). However, the interpretation was problematic because the correlation between self-concept and self-efficacy (0.93) was so high, resulting in very large SEs for both paths. In an alternative model, the two paths were constrained to be equal. This constraint did not result in a statistically significant decline in fit, provided a better more parsimonious fit to the data, and also substantially reduced the sizes of SEs (from 0.50 to 0.03).

Multicollinearity can be reduced even more by applying the more general, hierarchical (sequential) approach analogous to residualizing procedure like that proposed by Landram and Alidaee (1997; Lance, 1988; Little, Bovaird, & Widaman, 2006; but *see also* Lin, Wen, Marsh, & Lin, ~~in press~~; Marsh, Wen, Hau, Little, Bovaird, & Widaman, 2007) for polynomial components. With this approach, variance attributable to the predictor (X_1 and X_2) variables is partialed out of the X_1X_2interaction (product) term, main effect and first-order interactions are partialed out of second-order interactions, and so forth. This strategy is consistent with the hierarchical approach used to test interactions (based on the change in R^2) but has the advantage of substantially reducing levels of multicollinearity. However, it also introduces interpretational problems because the coefficients of the interaction terms are no longer in the same metric as the predictor variables and so it is not so easy to graph the interaction effect (*see* King, 1986, on problems with analyses of residuals). Hence, we recommend that this approach should only be used with appropriate caution and this it is generally better to use centering and standardizing strategies). Importantly, multicollinearity may sometimes indicate that the model has been in some ways misspecified (e.g., when the predictor and the moderator should, in fact, be considered as indicators of the same construct) and should not be simply partialled out of the model.

AQ: Is there an update to Lin, Wen, Marsh, & Lin?

Summary of Traditional (Non-Latent) Approaches to Interaction Effects

In this section, we briefly reviewed approaches to testing interaction effects based on observed (non-latent) variables. We began with a brief discussion of ANOVA when all predictor variables are categorical, noting in particular the dangers in transforming continuous predictors into categorical variables that are appropriate for ANOVA. We then moved to the more general moderated multiple regression approach that can be applied to continuous or categorical predictors (and thus include ANOVA models as a special case). We emphasized that graphs of interactions are typically a useful starting place in the interpretation of results, but these should always be supplemented with tests of statistical significance of interaction effects overall and supplemental analyses such as tests of simple slopes and regions of significance. We recommended consideration of, and discussed issues related to, a number of transformations of the original data (centering predictor variables or standardizing all variables). We concluded with a discussion of the related issues of power and multicollinearity (and the little-used residualizing procedure). A critical limitation in most psychological research is that when tests of interactions are appropriately constructed, *a priori* interaction effects that are intuitive and based on a strong theoretical rationale are typically small, non-significant, or not replicable. The implication is that most research lacks the statistical power to detect interaction effects—particularly in non-experimental studies where interaction effects are typically small. A critical issue related to power is the substantial measurement error associated particularly in interaction terms. Latent variable models of interaction effects that control for measurement error might substantially reduce measurement error as well as having more general strategic advantages. Hence, we now turn to an overview of these approaches.

Latent Variable Approaches to Tests of Interaction Effects

The critical feature common to each of the above non-latent approaches is that all of the dependent and independent variables are *observed* variables inferred on the basis of a single indicator rather than latent variables inferred by multiple indicators. Particularly when there are multiple indicators of these variables (e.g., multiple items in rating scales or achievement tests), latent variable approaches provide a much stronger basis for evaluating the underlying factor structure relating multiple indicators to their factors, controlling for measurement error, increasing power, testing the implicit factor structure used to create scale scores, and, ultimately, providing more defensible interpretations of the interaction effects.

Despite the ongoing emphasis on interaction effects, empirical support for predicted interactions has been disappointingly limited. One reason might be that the independent variables are contaminated by measurement error and do not provide accurate estimates of true interaction effects (Moulder & Algina, 2002). Indeed, because the measurement error of each of the main effect variables combines multiplicatively in the formation of the interaction term, measurement error in the interaction is likely to be substantially larger than in either of the main effects variables (*see* earlier discussion). When each of the independent variables is a latent variable inferred from multiple indicators, SEM provides many advantages over the use of analyses based on observed variables.

Nevertheless, despite the widespread use of SEMs for the purposes of estimating relations among latent variables and the importance of interaction effects, there have been very few substantive applications of SEMs to estimating interactions between two latent variables. Obviously, the paucity of such interaction applications does not result from a lack of relevant substantive applications that require interaction terms. Rather, as noted by Rigdon, Schumacker, and Wothke (1998), inherent problems in the specification of SEMs with interactions between latent variables have led researchers to pursue other approaches. Similarly, Jöreskog (1998) warned that latent variable approaches require the researcher to understand how to specify complicated nonlinear constraints. Although a variety of different approaches to estimating latent interactions were described in the book edited by Schumacker and Marcoulides (1998), and some new approaches have been developed by Algina and Moulder (2001), Wall and Amemiya (2001), and Klein and Moosbrugger (2000), there has been no consensus that any one of these approaches was optimal. Further, most of these approaches are not practically useful for the applied researcher because of the difficulties in specifying the nonlinear constraints. Another possible reason for the infrequent use of the latent approach is the difficulty in deciding how to construct or select the multiple indicators to form the latent factors, as typically required by all traditional latent approaches.

Latent interactions fall into two broad categories. In the first and generally simpler situation, at least one of the variables involved in the interaction is a categorical variable with only a few categories (e.g., gender: male and female). In this situation, a multiple-group SEM is appropriate in which the different categories are treated as separate groups within the same model. Even when both variables involved in the interaction are categorical, they can be used to form groups reflecting the main and interaction effects (e.g., two dichotomous variables can be used to form four groups reflecting the two main effects and one interaction effect; which is like an ANOVA within a SEM framework but providing control for measurement error). Although it might be argued that the interaction in this case is not really latent, it is still a latent variable model as long as the dependent (or other predictor) variables are inferred on the basis of multiple indicators.

In the second situation, both independent variables involved in the interaction are latent and continuous. Here there are various approaches to estimating their interaction effects, and "best practice" is still evolving. Marsh, Wen, and Hau 2004; (*see also* Marsh, Wen & Hau, 2006) reviewed various alternative approaches for estimating these latent interaction effects, including the unconstrained (Marsh et al., 2004), constrained (Algina & Moulder, 2001), generalized appended product indicator (GAPI; Wall & Amemiya, 2001), and the distribution-analytic (Klein & Moosbrugger, 2000; Klein & Muthén, 2007; *see also* Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2009) approaches[1].

In the remainder, we will first discuss multiple-group SEMs that are appropriate to analyze interactions when one of the predictors is categorical. We will then turn to the discussion of approaches to study interaction between latent continuous variables.

Multiple Group Structural Equation Modeling Approach to Interaction

Consider an interaction between a latent variable (ξ_1) and an observed variable (X_2) on a latent variable (η), and assume that X_2 is a categorical variable with a few naturally existing categories. We can use a multiple-group SEM approach (e.g., Bagozzi & Yi, 1989; Byrne, 1998; Rigdon, Schumacker, & Wothke, 1998; Vandenberg & Lance, 2000) with the categorical variable (X_2) as the grouping variable. Once the sample is divided into a small number of groups according to the values of X_2, we conduct multiple-group SEM for the latent variable ξ_1 and compare the model with and without restraining the effect of ξ_1 on the dependent variable η to be equal across the groups. Letχ^2_1 (with df_1) andχ^2_2 (with df_2) be the chi-square test statistics, respectively, with and without restraining the effect of ξ_1 on η to be equal in all the groups. If there is substantial decline in the goodness of fit with the invariance constraint (i.e., $\chi^2_1 - \chi^2_2$ with $df_1 - df_2$ is significant and there is a substantial deterioration in selected goodness of fit indexes), then the effects in the different groups are not identical and there is an interaction between the categorical predictor X_2 and the latent variable ξ_1. Of course, the invariance of the corresponding loadings and factor variances (i.e., constraining them to be the same in the different groups) must have been examined earlier (for details, *see*, e.g., Marsh, Muthén et al., 2009; Marsh, Lüdtke et al., ~~in press~~; Vandenberg & Lance, 2000; *see also* subsequent discussion and further discussion in Chapter XXX).

AQ: Is there an update to Marsh, Ludt al?

AQ: Please specify chapter number.

Historically this approach has been used often and has some advantages—particularly in terms of ease of implementation. If one of the independent variables is an observed variable that can be used to divide the sample naturally into a small number of groups, then multiple-group SEM is a simple, direct, yet effective approach (Bagozzi & Yi, 1989; Rigdon, Schumacker, & Wothke, 1998), which can be easily implemented in most common commercial SEM softwares. However, as emphasized earlier, this approach is generally not appropriate when two or a small number of groups are formed from a reasonably continuous predictor variable in that it ignores measurement error in the variable used to divide the sample into multiple groups and actually increases the unreliability in the grouped variable relative to the original continuous variable. Strategically there are limitations in that it is not easy to estimate the size of the interaction effect and some of the groups can become unacceptably small. In general we recommend not using this approach unless one of the interacting variables is a true categorical variable with a small number of categories and at least moderate sample sizes.

Structural Equation Models with Product Indicators

Kenny and Judd (1984) first used a SEM model with an interaction term to estimate latent interaction effects. In their approach, the dependent variable y is an observed variable, whereas the independent variables ~~1,2~~ each has two observed indicators x_1, x_2 and x_3, x_4, respectively, with all variables being centered (mean = 0). If the latent variables are identified by fixing the loading of the first indicator to 1, then the measurement equations of the model are:

$$x_1 = \xi_1 + \delta_1, x_2 = \lambda_2\xi_1 + \delta_2; x_3 = \xi_2 + \delta_3;$$
$$x_4 = \lambda_4\xi_2 + \delta_4.$$

The structural equation is:

$$y = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta, \quad (4)$$

where $\xi_1\xi_2$ is the interaction term of ξ_1 and ξ_2 on y. In the model, it is assumed that latent variables and the error terms are all normally distributed and there is no correlation between the latent variables and the error terms or between any two error terms.

Equation 4 is not linear with respect to ξ_1 and ξ_2 and is different from the SEM equations generally used. If we consider $\xi_1\,\xi_2$ as a separate third latent variable, then having no indicators for this variable remains a problem. To solve this, Kenny and Judd used the products of all possible pairs of the centered indicators x_1x_3, x_1x_4, x_2x_3, $x_2x_4$as the indicators for $\xi_1\xi_2$ with a lot of constraints added for identification.

Kenny and Judd's (1984) ingenious work was heuristic, stimulating many published studies that present alternative approaches to the use of products of indicators to estimate latent interactions (e.g., Algina & Moulder, 2001; Coenders, Batista-Foguet & Saris, 2008; Hayduk, 1987; Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Marsh et al., 2004; Ping, 1996; Wall & Amemiya, 2001). However, their original approach was unduly cumbersome and overly restrictive in terms of the assumptions on which it was based (*see* Marsh et al., 2004), leading to the development of new approaches.

Unconstrained Approach for Latent Interaction. Compared to the traditional constrained approach (e.g., Algina & Moulder, 2001; Marsh et al., 2004) and the partially constrained approach (Wall & Amemiya, 2001; Marsh et al., 2004), the unconstrained approach is fundamentally different and impressively simple to implement in that many of the complicated constraints in the original Kenny and Judd (1984) approach are no longer necessary. Marsh et al. (2004) showed that the unconstrained approach performed nearly as well as the constrained approach when all assumptions of the constrained approach were met and substantially better under more realistic conditions.

For the sake of simplicity, suppose that the endogenous latent variable has three indicators: y_1, y_2, y_3; the exogenous latent variables ξ_1 and ξ_2 also have three indicators, respectively: x_1, x_2, x_3 and x_4, x_5, x_6. Lin et al. (~~in press~~) proposed a double-mean-centering strategy for the unconstrained model that does not require a mean structure. First they centered all indicators to their mean, then formed a more parsimonious set of product indicators, and finally centered the product again (double-mean-centering). When SEM software (such as LISREL) is employed in practice, however, the product indicators do not really need to be re-centered again. Because when the model does not consist of a mean structure, the estimation results are identical with and without re-centering the product indicators. In other words, SEM software routinely treats the data being mean-centered if there is no mean structure in the model. Noting this point, Wu, Wen, and Lin (2009) suggested that researchers can directly use single-mean-centered data to analyze a latent interaction model without a mean structure. We note, however, that some statistical packages (e.g., Mplus) include a mean structure as the default so that this potential advantage of being able to ignore the mean structure might not be so important. Thus, the steps for analyzing the latent interaction are as below:

AQ: Is there an update to Lin et al?

1. center all indicators to their mean, still denoted as y_1, y_2, y_3; x_1, x_2, x_3; x_4, x_5, x_6;
2. form the product indicators x_1x_4, x_2x_5, x_3x_6;
3. use y_1, y_2, y_3 as the indicators of η, x_1, x_2, x_3 as the indicators of ξ, x_4, x_5, x_6 as the indicators of ξ_2, x_1x_4, x_2x_5, x_3x_6 as the indicators of ξ_3 (ξ_3 is the centered interaction term $\xi_1\xi_2 - E(\xi_1\xi_2)$, *see* Lin et al., ~~in press~~; Wu et al, 2009). The structural equation has three exogenous latent variables:

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \zeta. \qquad (5)$$

ξ_1, ξ_2, and ξ_3 are allowed to be correlated with each other, but each is uncorrelated with measurement errors and the residual term ζ.

In Equation 4 γ_1 and γ_2 represent the conditional main effects, γ_3 represents the interaction effect (for the path diagram, *see* Fig. 17.4). A simple LISREL syntax is given in Appendix 3). Readers who are familiar with the usual LISREL syntax may easily revise it for their researches involved in latent interactions).

Construction of Product Indicators of the Latent Interaction. A potential problem with the indicator approach is how to form the product indicators. In contrast to earlier, *ad hoc* approaches to create produce indicators, Marsh, Hau, and Wen (2004, 2006) proposed the guiding principles: (1) use all the multiple indicators from both interacting variables in the formation of the indicators of the latent interaction factor, and (2) do not re-use the same indicator in forming the indicators for the latent interaction factor. Thus, each indicator in ξ_1 *and* $\xi_2$should be used once and only once in forming the indicators for the latent interaction factor.

In some situations there is a natural matching that should be used to form product indicators (e.g., the items used to infer ξ_1 and ξ_2 have parallel wording). More generally, when the two first-order effect factors ξ_1 and ξ_2 have the same number of indicators, Marsh et al. (2004; *see also* Saris, Batista-Foguet, & Coenders, 2007) suggested that it would be better to match indicators in terms of the reliabilities of the indicators (i.e., the best item from the first factor with the best item from the second factor, etc.).

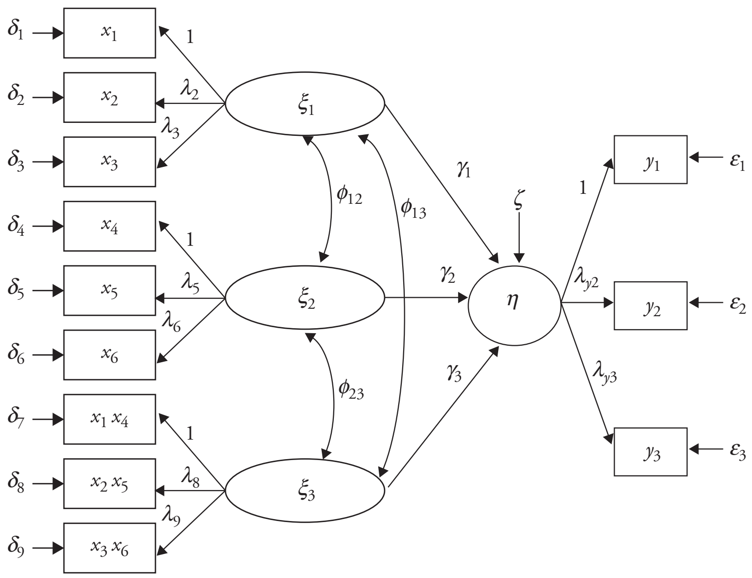

Figure 17.4 The notional path diagram of the latent interaction model. ξ_3 is the centered interaction term $\xi_1\xi_2 - E(\xi_1\xi_2)$.

However, if the number of indicators differs for the two first-order effect factors, then a simple matching strategy does not work. Assume, for example, that there were 5 indicators for the first factor and 10 for the second. One approach would be to use the 10 items from the second factor to form five (item pair) parcels by taking the average of the first 2 items to form the first item parcel, the average of the second 2 items to form the second parcel, and so forth. In this way, the first factor would be defined in terms of 5 (single-indicator) indicators, the second factor would be defined by ~~5 (item pair parcel)~~ indicators, and the latent interaction factor would be defined in terms of 5 matched-product indicators.

Appropriate Standardized Solution for Latent Interactions Standardized parameter estimates are important because they facilitate the comparison of the effects of different predictor variables in the equation. Analogous to the corresponding problem in multiple regression with the manifest interaction described earlier, the usual standardized coefficients of SEMs with latent interaction are not appropriate. Wen, Marsh, and Hau 2010; (*see also* Wen, Hau & Marsh, 2008) derived an appropriate standardized solution for latent interaction. They proved that these appropriate standardized estimates of the main and interaction effects as well as their standard errors and$t$-value are all scale-free (for further discussion, *see* Wen et al., 2010). Although specifically designed for the latent interaction model, this one-step approach might also be appropriate to manifest models based on single indicators of predictor variables.

Distribution-Analytic Approaches

In contrast to the product-indicator approaches that model the latent interaction by specifying a separate latent interaction variable, the distribution-analytic approaches (Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2009) explicitly model the distribution of the latent outcome variables and their manifest indicators in the presence of latent nonlinear effects and provide a promising alternative for the estimation of nonlinear SEM. Two methods are currently available: the Latent Moderated Structural Equations (LMS) approach (Klein & Moosbrugger, 2000) that is implemented in Mplus (Muthén & Muthén, 1997–2008) and the Quasi-Maximum Likelihood (QML) approach (Klein & Muthén, 2007) that has not yet been implemented in a readily available software package (but is available from its author Andreas Klein).

Both QML and LMS directly estimate the parameters of the latent interaction model given in Equation 4 (and are flexible enough to handle quadratic effects of latent variables as well) without having to resort to the use of product-indicators. They differ in the distributional assumptions made about the latent dependent variable ? and its indicators and in the estimation method used to obtain the parameter estimates (for more technical description, *see* Klein & Moosbrugger, 2000; Klein & Muthén, 2007).

In contrast to product-indicator approaches following from Kenny and Judd's (1984) research, it is not necessary to have indicators of the latent interaction variable in the distribution-analytic approaches. The product-indicator approaches assume normality of indicators and latent constructs to estimate parameter and standard errors. In general, both assumptions are violated in models with latent interactions, although results seem to be robust in relation to these violations. In addition, alternative approaches such as bootstrapping (*see* Wen et al., 2010) can be used. Bootstrapping is becoming more readily available in standard statistical packages and provides a more accurate estimate of SEs—particularly when N is small. The distribution-analytic approaches, on the other hand, maximize special fitting functions that take the non-normality of the indicators of the dependent latent variable explicitly into account (but still rely on normality assumptions about the indicators of the latent predictor variables). However, these advantages are offset by the need to use specialized software to estimate these models (Mplus and the QML-program) and the high computational demands of the LMS approach that limit its applicability.

Simulation studies that compared distribution-analytic approaches (Klein & Muthén, 2007; Marsh et al., 2004) to the unconstrained approach showed that QML appears to be sufficiently robust in relation to non-normal data. LMS, on the other hand, can yield biased standard errors when distributional assumptions about the indicators of the predictor variables are violated. The QML program provides a fully standardized solution that assumes that all manifest and latent variables are standardized but no standard errors for the standardized effects. These estimates are equal to the parameters of the appropriately standardized solution. Standardized effects for LMS are harder to obtain, as the current implementation does neither provide a standardized solution nor an estimator for the variance of the latent product variable $\xi_1\xi_2$ to calculate the standardized solution.

Comparisons of LMS and QML (Klein & Muthén, 2007) indicate that LMS is slightly more efficient when its distributional assumptions are fulfilled, but QML is comparatively more robust to violations of normality. It appears that the QML approach will be more useful for applied researchers, but there have been too few studies with real data to fully evaluate its potential in actual practice. Indeed, this concern can be applied to all of the various approaches to latent interaction.

Summary of Latent-Variable Approaches to Interaction Effects

Non-latent variable tests of interaction effects traditionally lack power because of the substantial amount of measurement error, particularly in the interaction component. In this section we briefly described new and evolving approaches to testing latent interaction effects. When one of the predictor variables is a manifest grouping variable with a small number of categories (e.g., male/female) the traditional approach to multigroup SEM can be applied. This approach is well established, easily implemented, and facilitates a detailed evaluation of invariance assumptions (e.g., invariance of factor loadings over groups) that are largely ignored in other approaches. However, the multigroup approach is generally not recommended when all the predictor variables are continuous or based on multiple indicators. Historically, the product indicator approaches have dominated latent interaction research. Although these approaches are still evolving, there is good evidence in support of the unconstrained approach in terms of robustness and ease of implementation in terms of all commercial SEM packages. More recently, distribution-analytic approaches (LMS and QML) hold considerable promise and apparently have strategic advantages over the product-indicator approach. Although LMS is available in Mplus, the QML approach is not available in any major statistical package (but is available from its author Andreas Klein upon request). Despite a long history dating back to 1984, latent interaction effects are rarely used in applied research. Furthermore, even the limited numbers of demonstration and simulation articles have focused on statistical issues involved in estimating the models and have not adequately dealt with many of the issues faced by applied researchers (including some of those discussed here in relation to the moderated multiple regression approach).

Summary

Tests of interaction effects are central to psychology. The implications of being able to test and interpret interaction effects are critical for theory, substantive understanding, and applied practice. Yet, typical practice in the evaluation of interaction effect is surprisingly weak. In our chapter we have provided an overview of the topic of moderation based on both manifest and latent variable models. In doing so, we have attempted to summarize current best practice. The good news is that there has been important progress in both best practice and typical practice. Applied researchers are apparently becoming more aware of the basic issues in the testing and interpretation of interaction effects. The bad news is that many well-established methodological requirements continue to be ignored in applied research. Although there are some stunning new developments in the application of latent models to testing interaction effects, these new approaches have not yet had much impact on applied research. Further, these new developments have focused primarily on the substantial statistical issues involved in fitting data to the latent variable models and have not made much progress in resolving many of the complications that have been the focus of manifest applications. Hence, it is perhaps unsurprising that evolving latent variable approaches to moderation effects have had limited effect on typical practice in applied research.

In this chapter, we have tried to promote a substantive-methodological synergy (Marsh & Hau, 2007), bringing to bear new, strong, and evolving methodology to tackle complex substantive issues with important implications for theory and practice. Here, as in other areas of applied psychological research, theory, good measurement, research, and practice are inexorably related such that the neglect of one will undermine pursuit of the others. Marsh and Hau (2007) claimed: (1) some of the best methodological research is based on the development of creative methodological solutions to problems that stem from substantive research; (2) new methodologies provide important new approaches to current substantive issues; and (3) methodological-substantive synergies are particularly important in applied research. In the study of moderation, this synergy is particularly important.

Limitations and Directions for Further Research

We now turn to a set of issues that are beyond the scope of the present chapter. In some cases these are ongoing or evolving issues that have not been resolved in the literature and reflect our thoughts about directions for further research.

Quadratic Effects: Confounding Nonlinear and Interaction Effects

Examples with quadratic effect are common. For example, nonlinear effects may be hypothesized between strength of interventions (or dosage level) and outcome variables such that benefits increase up to an optimal level, and then level off or even decrease beyond this optimal point. At low levels of anxiety, increases in anxiety may facilitate performance but at higher levels of anxiety, further increases in anxiety may undermine performance. Self-concept may decrease with age for young children, level out in middle adolescence, and then increase with age into early adulthood. The level of workload demanded by teachers may be positively related to student evaluations of teaching effectiveness for low to moderate levels of workload but may have diminishing positive effects or even negative effects for possibly excessive levels of workload. Quadratic effects can be seen as a special case of nonlinearity effect that can be very complicated but are not discussed here in detail.

The existence of quadratic effects can complicate the analysis and interpretation of interaction effects. A strong quadratic effect may give the appearance of a spurious significant interaction effect, and it is sometimes difficult to distinguish them (Klein et al., 2009; Lubinski & Humphreys, 1990; MacCallum & Mar,1995). Thus, without the proper analysis of the potential quadratic effects, the investigator might easily misinterpret, overlook, or mistake a quadratic effect as an interaction effect. Particularly when both X_1 (the predictor) and the moderator are positively correlated, it is likely that the product (quadratic and interaction) terms are also correlated. That is, the quadratic function of X_1 (i.e., $X_1 \times X_1 = X_1^2$) is likely to be correlated with the interaction ($X_1 \times X_2$). Unless there is a well-established causal ordering of X_1 and X_2, then there is no easy solution as to how to disentangle the potential confounding between the quadratic and moderation terms. In most cases, the variance that can uniquely be explained by the interaction effect will be diminished by controlling for quadratic effects. However, this is the same sort of confounding that is typical in multiple regression analyses even when product terms are not considered; some variance can be uniquely attributed to one predictor, some to the other predictor, and some can be explained by either predictor. Hence, it is reasonable to include both the nonlinear and the interaction effects in the same model to more fully specify that pattern of relations among the variables.

When all variables are manifest, the inclusion of quadratic effects is a relatively straightforward extension of approaches outlined here. Indeed, the construction of the statistical model is likely to be much easier than the interpretation of the results. However, for latent variable models when all predictor variables are based on multiple indicators, the simultaneous inclusion of latent interaction and latent quadratic terms is likely to prove more complicated. Nevertheless, there has been some recent work on this issue involving both the multiple-indicator and distributional approaches to latent variable modeling (*see* Klein et al., 2009; Marsh et al., 2006; Kelava & Brandt, in press). Furthermore, as is the case with interaction effects more generally, the focus of latent variable models has been more on the statistical issues involved in fitting the model than the interpretational concerns that have been the focus of studies based on manifest variables (and our chapter).

AQ: Is there an update to Kelava & Brandt?

We also note that in applied research it is sometimes suggested that nonlinear relations should be specified *a priori* on the basis of theory or previous research. However, the linearity of the observed relationships is an implicit assumption of any form of model that does not include nonlinear terms. Hence, it is always reasonable to test the appropriateness of the linearity assumption through the inclusion of nonlinear terms.

Moderation versus Mediation and the Role of Causal Ordering

The concept of moderation (or interaction) is often confused with that of mediation (*see* Holmbeck, 1997; Baron & Kenny, 1986). As noted by Frazier et al.(2004, p. 116): "Whereas moderators address 'when' or 'for whom' a predictor is more strongly related to an outcome, mediators establish 'how' or 'why' one variable predicts or causes an outcome variable." Mediation refers to the mechanism that explains the relation between X_1 and Y.

Mediation occurs (*see* Fig. 17.5) when some of the effects of an independent variable (X) on the dependent variable (Y) can be explained in terms of another mediating variable (MED) that falls between X_1 and Y in terms of the causal ordering: X_1 ➔ MED ➔ Y. A mediator is an intervening variable that accounts for—at least in part—the relation between a predictor and an outcome such that the predictor influences an outcome indirectly through the mediator. Thus, for example, the effects of mathematical ability prior to the start of high school (X) on mathematics achievement test scores at the end of high school (Y) are likely to be mediated in part by mathematics coursework completed during high school (MED). In Figure 17.5A the model assumes that the effect of X_1 on Y is unmediated (X_1 ➔ Y).

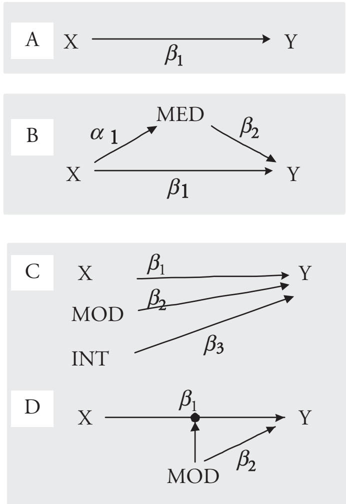

Figure 17.5 The distinction between mediation (MED) and moderation (MOD). X = predictor variable. Y = outcome variable. INT = interaction. In (**A**), X effects Y with no mediation or moderation. In (**B**), the effect of X on Y is mediated in part by MED. There is total mediation if the direct effect of X on Y is 0 ($\beta_1 = 0$) or partial medication ~~of~~ $\beta_1 \neq 0$. The indirect effect of X on Y thru that is mediated by MED is the product of α_1 and β_2. (**C**) and (**D**) offer two representations of the interaction effect. In (**C**), the interaction is depicted as a separate construct (typically defined as the product of X and MOD, but sometimes after both have been centered or standardized). (**D**) offers demonstrates more explicitly that MOD moderates the relation between X and Y.

In Figure 17.5B the model assumes that part of the effect of X_1 on Y is direct (X_1 ➔ Y, the path that goes directly from X_1 to Y) and that some of the effect is mediated (X_1 ➔ MED ➔ Y) by the indirect path that goes from X_1 to MED and then from MED to Y). If both the direct (X_1 ➔ Y) and indirect (X_1 ➔ MED ➔ Y) effects are non-zero, then the X–Y relation is said to be partially mediated, whereas in Figure 17.5B, the X_1–Y relation is said to be completely mediated by MED if the X_1 ➔ Y effect is 0.

In contrast, and as discussed extensively above, moderation is said to have taken place when the size or direction of the effect of X_1on Y varies with the value of a moderating variable (MOD). Thus, for example, the effect of a remedial course in mathematics rather than regular mathematics coursework (X) on subsequent mathematics achievement (Y) may vary systematically depending on the student's initial level of mathematics ability (MOD); the effect of the remedial course may be very positive for initially less able students, negligible for average-ability students, and even detrimental for high-ability students who would probably gain more from regular or advanced mathematics coursework. Figure 17.5C shows the interaction as a separate construct (INT) and the effect of the interaction on the outcome (INT $\rightarrow$ Y) as a separate regression coefficient (β_3). Alternatively the representation of the interaction effect in Figure 17.5D shows more clearly that the moderator influences the relation between two variables.

Although a full discussion of mediation is beyond the scope of this chapter, a critical requirement of mediation is that there is a clear causal ordering from $X \rightarrow$ MED $\rightarrow Y$. Particularly as mediation analyses are typically applied in nonexperimental studies, this is an important consideration. Frazier et al. (2004) have proposed that the most defensible strategy is a longitudinal study in which all three variables are tested on at least three occasions (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Marsh & O'Mara, 2008). Without clear support for the causal ordering of the X_1 and MED variables, traditional tests of mediation make no sense. Unless there is a clear causal ordering, it is not possible to rule out (empirically, theoretically, or commonsensically) the possibility that X_1 and MED are reciprocally related (i.e., both $X_1 \rightarrow$ MED, and MED $\rightarrow X_1$; *see* Marsh & Craven, 2006, for further discussion of reciprocal effects). This point is at the heart of the Judd and Kenny (2010) critical re-evaluation of the Baron and Kenny (1986; Judd & Kenny, 1981) assumptions of what are necessary and sufficient conditions to demonstrate mediation. The issue of causality in mediation requires strong theory, a good design, appropriate statistical analyses, and probably an ongoing research programme that addresses the same issues from multiple perspectives. It is our contention that the vast majority of studies purporting to test mediation—particularly when based on a single wave of data—are either wrong or cannot be defended in relation to the typically implicit, untested assumption of causal ordering. Without strong assumptions of causality, tests of mediation are typically uninterpretable.

For purposes of this chapter, we argue that there needs to be no causal ordering established between X_1 and MOD to test interaction effects. The statistical tests cannot distinguish between interpretations that X_1 moderates MOD and MOD moderates X_1; the two perspectives are equivalent in terms of statistical significance, variance explained, and so forth. Our recommendation is that unless there is a clearly established causal ordering, then both possibilities should be considered, although one might be more substantively or theoretically useful (Brambor, Clark, & Golder, 2006). However, if a strong basis of causal ordering can be established on the basis of theory or the design of the study (e.g., with longitudinal data), then the interpretability of the results is greatly enhanced. In the evaluation of interaction effects, there is, of course, the assumption that both X_1 and MOD precede Y in the causal ordering and tests of interactions typically are not symmetric in relation to X_1 and Y (or for MOD and Y). Thus, for example, if there is not a theoretical or substantive rationale for assuming $X_1 \rightarrow Y$ rather than $Y \rightarrow X_1$ (e.g., X_1 is an intervention based on random assignment) Traci and Russell (2003) have suggested that researchers should consider different models in which X_1 is considered the outcome rather than the predictor variable.

Although we have emphasized the distinction between mediation and moderation, they are obviously not mutually exclusive. It is quite conceivable to have moderated mediation or mediated moderation. Here the issues of direction of causality become even more important (for further discussion, *see* Judd & Kenny, 2010; Muller, Judd, & Yzerbyt, 2005).

Interactions with More Than Two Continuous Variables

The moderated regression approach described above can be extended to include more than two independent variables with multiple two-way or other higher-order interactions involving more than two variables. Should we include all possible interactions in the regression equation? In general, the inclusion of interaction terms should be theory driven. It is also tempting to exclude non-significant effects to simplify the regression equation, increase the degrees of freedom, and increase the power of the remaining statistical tests. Nevertheless, Jaccard et al. (1990) and many others have recommended that interaction terms supported by strong theory—as well as the predictor variables used to construct the interaction terms—should always be included irrespective of whether they are significant. In particular, failure to include main effects (or lower-order interaction) will typically bias coefficient estimates for the higher-order interactions (*see* earlier discussion). Also, for hypothesized interaction effects that are based on theory, the values of even non-significant effects can be important, for example, for meta-analyses that seek to combine the effects from many different studies.

When multiple interaction effects are examined in the same regression equation, one approach to safeguard against an inflated Type I error is to use an omnibus F-test to compare squared multiple correlations with and without the entire set of interactions terms. Interactions are individually inspected only when this ombibus F-test is significant (Aiken & West, 1981; Frazier et al., 2004). Analogous to analyses of higher-order interactions in ANOVA, the interpretation of the three-way (or other higher order) interaction in regression may not be straightforward and easily interpretable. When a certain three-way interaction is significant, for example, we can construct simple slopes similar to those for the two-way interactions. Operationally, to find the slope of X_1 at certain values of X_2 and X_3 (or symmetrically, slope of X_2 at certain values of X_1 and X_3), substitute these values of X_2 and X_3 into the full regression equations with known coefficients and the coefficient of X_1 will be the required slope. Corresponding t-tests on these coefficients can also be computed with the appropriate standard errors of the coefficients (*see* Jaccard et al., 1990) as in the case with two-way interactions.

Measurement error is likely to be an even larger problem for the evaluation of higher-order interactions. As noted earlier with two-way interactions, measurement errors associated with each of the separate predictor variables combines multiplicatively so that measurement error is typically much larger for the interaction terms. This issue is likely to be exacerbated even more for higher-order interactions. Hence, there is the need for latent variable models that control for unreliability. However, latent variable models of interaction have primarily focused on models of two-way interactions, and there has been little progress on extending these models to include higher-order interaction.

Tests of Measurement Invariance

Of particular substantive importance for applied psychological research are mean-level differences across multiple groups (e.g., male vs. female; age groups; single-sex vs. co-educational schools) or over time (i.e., observing the same group of participants at multiple occasions, perhaps before and after an intervention) as well as interactions between these variables. What have typically been ignored in such studies are tests of whether the variables have the same meaning in the different groups or for multiple occasions. When there are multiple indicators of the constructs, typical approach is to test the invariance over groups or time of the factor structure and the item intercepts. An important assumption underlying tests of mean differences is that group differences in the latent mean are reflected in each of the multiple indictors of the latent construct. For example, if the underlying factor structure for the outcome variable Y differs for boys and girls, or over time, then interpretations of mean level differences are problematic. Also, if mean level differences are not consistent across items used to infer the outcome, given comparable levels on the estimated outcome (i.e., there is differential item functioning), then the observed differences might be idiosyncratic to the particular items used. Obviously these issues are critical in the interpretation of interaction effects. If the meaning of a predictor variable is qualitatively different for boys and girls, then it makes no sense to test the effect of the interaction between the predictor variable and gender

The evaluation of model invariance over different groups (e.g., gender) or over time for the same group is widely applied in SEM studies (Jöreskog & Sörbom, 1988; Meredith & Teresi, 2006; Vandenberg & Lance, 2000). Indeed, such tests of invariance might be seen as a fundamental advantage of CFA over traditional approaches to EFA (but *see also* recent applications of exploratory structural equation modeling that integrate EFA and CFA approaches in the evaluation of measurement invariance; *see* Marsh, Muthén et al., 2009, Marsh, Lüdtke, et al., ~~in press).~~ Tests of measurement invariance begin with a model with no invariance of any parameters (*configural invariance*) followed by tests of whether factor loadings are invariant over groups (*weak invariance*). *Strong measurement invariance* requires that the indicator intercepts and factor loadings are invariant over groups or time and is an assumption for comparison of latent means. *Strict measurement invariance* requires invariance of item uniquenesses (i.e., each item's residual variance) in addition to invariant factor loadings and intercepts and is an assumption for the comparison of manifest means over groups or time. Unless there is support for at least strong measurement invariance, then the comparison of latent means is not justified (because of the problem of differential item functioning). The comparison of manifest means requires the much stronger assumption of strict measurement invariance (as a necessary, but not a sufficient condition).

The multiple-group approach to measurement invariance is easily formulated and tested when in relation to a predictor that is a true grouping variable with a relatively small number of discrete categories. However, for continuous variables, the multiple-group approach would require researchers to transform continuous variables into a relatively small number of categories that constitute the multiple groups—a practice we have criticized earlier. Marsh, Tracey, and Craven (2006; *see also* Marsh, Lüdtke et al., ~~in press~~) have proposed a hybrid approach involving an integration of interpretations based on both multiple-indicator-multiple-cause (MIMIC) and multiple group approaches. In the MIMIC approach, the predictor variables (X_1 and X_2) and their interaction are included in the model based on the total sample; they can be categorical or continuous, latent or manifest. In this respect, the distinction between the MIMIC and multiple-group approaches is analogous to the multiple-group and single (total) group distinction already discussed in relation to the moderated regression approach. The main difference is that in the MIMIC and multiple-group approach as used here, some of the variables are latent variables based on multiple indicators.

Moderated multiple regression models like those considered in the first half of this chapter are based on manifest variables that might not have multiple indicators. Although some constructs might be appropriately measured with a single indicator that has no measurement error (e.g., gender), most cannot. However, tests of single-indicator (manifest) variables make the same assumptions of measurement invariance, as do latent variables. Indeed, the assumptions are even more stringent (strict measurement invariance is needed rather than strong measurement invariance). The problem is that typical approaches to measurement invariance cannot be used to test the appropriateness of these assumptions for single-indicator manifest constructs. Although issues of measurement invariance have been largely ignored in tests of interactions, this situation is likely to change with the increased focus on latent variable models in general and latent interactions in particular.

Multilevel Designs and Clustered Samples

There is a special type of interactions in that data points are related as clusters and are not collected totally independently (e.g., students' questionnaire responses from 100 schools, with data from the same school sharing some commonalities). The effects of the independent variables on dependent variables (or their relations) may change according to their subgroup units (e.g., school) or their characteristics. For example, parental influence on students' achievement depends on the characteristics of the schools (e.g., whether the student is studying in a school with low or high average socioeconomic status). Indeed, a key question in many multilevel studies is how the effect of an individual variable varies from group-to-group, and whether there are group-level variables that can explain the group-level variation. Although the groups are typically considered as a random effect, rather than a fixed effect, the issue is still an interaction question—is the effect of one variable moderated by another variable (e.g., the particular school that a student attends).

Historically, multilevel researchers have tended to work with manifest variables that ignore measurement error, whereas SEM researchers have tended to work with latent variable models that ignore multilevel structures in their data. However, these two dominant analytic approaches are increasingly being integrated within more comprehensive multilevel SEMs (e.g., Lüdtke, Marsh et al., 2008; Marsh, Lüdtke et al., 2009). Inevitably, this integration will lead to increased sophistication in the study of latent interactions, a better understanding of the cross-level interactions between individual-level and group-level variables, and a clearer distinction between fixed-effects that have been emphasized in our chapter and random effects that are emphasized in multilevel analyses. Although clearly beyond the scope of our chapter (*but see* chapter "Multilevel regression and Multilevel SEM" in this monograph), this is an important emerging area in quantitative analysis.

AQ: Please specify chapter number.

Author Notes

The authors would like to thank the following colleagues for input in relation to earlier drafts of this chapter: Eric Dearing, Kathleen McCartney; David Kenny, Matthew R Golder; and Thomas Brambor.

Note

1. As in many other statistical models, there are Bayesian approaches and non -Bayesian approaches in estimating latent interaction models. Although well developed (see, e.g., Arminger & Muthén, 1998; Lee, Song & Poon, 2004), Bayesian approaches and their calculation algorithms are relatively difficult for general applied researchers and thus although available in the special WinBUGS software, these approaches have not been adopted for the most popular commercial SEM software (but also see the Bayesian approach recently introduced in Mplus).

AQ: Please note that the reference for the citation "Arminger & Muthén, 1998" has not been given in the references list.

References

Aguinis, H. (2004). *Moderated regression*. New York: Guilford.
Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, 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.

Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental designs. *Journal of Marketing Research*, *26*, 271–284.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis, 14,* 63–82.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications and programming*. Mahwah, NJ: Erlbaum.

Carte, T. A. & Russell, C. J. (2003). In Pursuit of Moderation: Nine Common Errors and Their Solutions. *Management Information Systems Quarterly*, *27*, 479–501.

Coenders, G., Batista-Foguet, J.M. & Saris. W. E. (2008). Simple, efficient and distribution-free approach to interaction effects in complex structural equation models. *Quality & Quantity, 42*, 369–396.

Cohen, J. (1968). Multiple regression as a general data analytic system. *Psychological Bulletin 70*(6), 426–443.

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

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

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

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Oxford, England: Irvington.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Dearing, E., & Hamilton, L. C. (2006). Contemporary approaches and classic advice for analyzing mediating and moderating variables. *Monographs of the Society for Research in Child Development*, 71 (Serial No. 285).

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology. *Journal of Counseling Psychology*, *51*, 115–334.

Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, *26*, 797–833.

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

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65,* 599–610.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance.* Thousand Oaks, CA, US, Sage Publications.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.

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.

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

Jöreskog, K. G. (1998). Interaction and nonlinear modeling: issue and approaches. In R.E. Schumacker, G.A. Marcoulides, (Eds.), *Interaction and nonlinear effects in structural equation modelling* (pp. 239–250). Mahwah, NJ: Erlbaum.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7—A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS.

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: Issued and techniques* (pp. 57–88). Mahwah, NJ: Erlbaum.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602–619.

Judd, C. M., & Kenny, D. A. (2010). Data analysis in Social Psychology: Recent and recurring issues (pp. 115–139). In S. T. Fiske, D. T. Gilbert & G. Lindzey. *Handbook of Social Psychology* (5th ed). Hoboken NJ: Wiley.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2009). *Data analysis: A model comparison approach* (2nd ed.). New York: Routledge.

Kelava, A., & Brandt, H. (in press). Estimation of Nonlinear Latent Structural Equation Models using the Extended Unconstrained Approach. *Review of Psychology.*

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 9,* 201–210.

King, G. (1986) How not to lie with statistics: Avoiding common mistakes in political science. *American Journal of Political Science, 30,* 666–687.

Kirk, R. (1982). *Experimental Design: Procedures for the Behavioral Sciences*. Belmont CA: Brooks/Cole.

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. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, 42, 647–674.

Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (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, The Netherlands: Sense Publishers.

Lance, Charles E. (1988). Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions.

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.

Landram, F. G. & Alidaee, B. (1997). Computing orthogonal polynomials. *Computers Operations Research*, *24*, 473–476.

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*, *29*, 37–67.

AQ: Please update Kelava & Brandt.

AQ: Is there an update to Lin, Wen, Marsh, & Lin?

~~Lin, G. C., Wen, Z., Marsh, H. W., & Lin, H. S. (in press). Structural equation models of latent interactions: clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling.*~~

Little, Todd D.; Bovaird, James A.; Widaman, Keith F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables.

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, 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. *Psychological Bulletin, 107,* 385–393.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*, 203–229.

MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118*, 405–421.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

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. & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163.

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A re-examination of relations between self-efficacy, self-concept and achievement. *Journal of Educational Psychology, 96,* 518–522.

Marsh, H. W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling, 1,* 317–359.

Marsh, H. W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology, 32,* 151–171.

Marsh, H. W., Hau, K.T. & Wen, Z., (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modelling, 11,* 320–341.

AQ: Is there an update to Marsh, Ludtke, Muthen, Asparouhov, Morin, Trautwein, & Nagengast

~~Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U. & Nagengast, B. (in press). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*.~~

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling,16*, 439–476.

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*, 542–552.

Marsh, H. W., Tracey, D. K., & Craven, R. G. (2006). Multidimensional self-concept structure for preadolescents with mild intellectual disabilities: A hybrid multigroup-mimic approach to factorial invariance and latent mean differences. *Educational and Psychological Measurement, 66,* 795–818.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005) Academic self-concept, interest, grades and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*, 297–416.

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. Hancock & Mueller (Eds). *Structural equation modeling*: *A second course* (pp. 225–265). Greenwich, CT: Information Age Publishing.

Marsh, H. W., Wen, Z. L., Hau, K. T, Little, T. D. Bovaird, J. A. & Widaman, K. F. (2007). Unconstrained structural equation models of latent interactions: Contrasting residual- and mean-centered approaches. *Structural Equation Modelling, 14,* 570–580.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.

Meredith, W., & Teresi, J. A. (2006). n Essay on Measurement and Factorial Invariance. *edical Care, 44*, S69–S77.

Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp.103–136). Rotterdam, The Netherlands: Sense Publishers.

Morin, A. J. S., Janoz, M. & Larivée, S. (2009). The Montreal adolescent depression development project (MADDP): School life and depression following high school transition. *Psychiatry Research Journal, 1*(3), 1–50.

Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, *9*, 1–19.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863.

Muthén, L. K., & Muthén, B. O. (1997–2008). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Ping, R. A., Jr. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin, 119*, 166–175.

Potthoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, *29*, 241–256.

Preacher, K. J., Curran, P. J., & Bauer, D. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437–448.

Rigdon, E. E., Schumacker, R. E. & Wothke, W. (1998). A comparative review of interaction and nonlinear modeling. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 1–16). Mahwah, NJ: Lawrence Erlbaum Associates.

Saris. W. E., Batista-Foguet, J.M. & Coenders, G. (2007). Selection of indicators for the interaction term in structural equation models with interaction. *Quality & Quantity*, *41*, 55–72.

Schumacker, R. E., & Marcoulides, G. A. (Eds.) (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Erlbaum.

Traci, A. C., & Russell, C. J. (2003). *In pursuit of moderation: Nine common errors and their solutions. MIS Quaterly, 27,* 479–501.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practi ces, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–69.

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.

Wen, Z., Hau, K. T., & Marsh, H. W. (2008). Appropriate standardized estimates for moderating effects in structural equation models. *Acta Psychologica Sinica, 40*(6), 729–736.

Wen, Z., Marsh, H. W., & Hau, K. T. (2010). Structural equation models of latent interactions: An appropriate standardized solution and its scale-free properties. *Structural Equation Modeling, 17*, 1–22.

Wu, Y., Wen, Z., & Lin, G. C. (2009). Structural equation modeling of latent interactions without using the mean structure. *Acta Psychologica Sinica*, *41*(12), 1252–1259.