Statistical Power Problems with Moderated Multiple Regression in Management Research

Herman Aguinis
University of Colorado at Denver

Due to the increasing importance of moderating (i.e., interaction) effects, the use of moderated multiple regression (MMR) has become pervasive in numerous management specialties such as organizational behavior, human resources management, and strategy, to name a few. Despite its popularity, recent research on the MMR approach to moderator variable detection has identified several factors that reduce statistical power below acceptable levels and, consequently, lead researchers to erroneously dismiss theoretical models that include moderated relationships. The present article (1) briefly describes MMR, (2) reviews factors that affect the statistical power of hypothesis tests conducted using this technique, (3) proposes solutions to low power situations, and (4) discusses areas and problems related to MMR that are in need of further investigation.

If we want to know how well we are doing in the biological, psychological, and social sciences, an index that will serve us well is how far we have advanced in our understanding of the moderator variables of our field
—Hall & Rosenthal, 1991, p. 447.

Numerous theories in management have reached a sufficient level of sophistication and development that researchers are interested in detecting not only the main effects of independent variables, but also their interactive (i.e., moderating) effects. The existence of a moderating effect implies that the relationship between two variables (e.g., X and Y) varies as a function of the value of a third variable (e.g., Z), labeled a *moderator* (Zedeck, 1971).

In recent years, interest in moderator variables in numerous management subdisciplines has increased substantially. For example, Bruton, Oviatt and White (1994) investigated acquisitions and tested whether the impact of business relatedness on acquisition performance was moderated by the degree of firm

Direct all correspondence to: Herman Aguinis, University of Colorado at Denver, Department of Psychology, Campus Box 173, P.O. Box 173364, Denver, CO 80217-3364. Internet: <haguinis@castle.cudenver.edu>

distress; Duarte, Goodson and Klich (1994) hypothesized that the relationship between performance and supervisory ratings was moderated by dyadic quality and duration; Nesler, Aguinis, Quigley and Tedeschi (1993) examined whether the effects of objective power on power perceptions was moderated by information regarding source credibility; and Williams and Alliger (1994) inquired whether the effects of work-family role juggling on self-reports of negative affect and calmness were moderated by the setting of the activities (i.e., work vs. home). Moderating effects play critical roles in theories in several other specialties of management and the social and behavioral sciences in general (Bedeian & Mossholder, 1994; Benson, Kemery, Sauser & Tankesley, 1985; Chaplin, 1991; Compas, Ey & Grant, 1993; Cordes & Dougherty, 1993; Lyness, 1993; Sackett & Wilk, 1994; Schmitt, Hattrup & Landis, 1993; Snell & Dean, 1994; Stiff, 1986; Whisman, 1993). These and numerous other theoretical developments support the position that moderator variables are "at the very heart of the scientific enterprise" (Hall & Rosenthal, 1991, p. 447).

A number of statistical procedures have been used to test for the presence of hypothesized moderating effects, one of these being moderated multiple regression (MMR) (Cohen & Cohen, 1983; Saunders, 1956; Zedeck, 1971). Several independent evaluations conducted over the past four decades indicate that MMR is an appropriate method for detecting the effects of moderator variables (Cleary, 1968; Cohen & Cohen, 1983; Friedrich, 1982; Saunders, 1956; Stone, 1988; Stone & Hollenbeck, 1984, 1989; Zedeck, 1971). Consequently, MMR is a frequently used technique for detecting these effects, as illustrated by Cortina (1993), who reported that MMR was used in at least 123 attempts to detect moderating effects in the 1991 and 1992 volumes of the *Journal of Applied Psychology*.

In spite of the increased use of MMR in management research, concerns have been raised regarding difficulties associated with its use (e.g., Aguinis, 1993; Alexander & DeShon, 1994; Bobko & Russell, 1994; Cronbach, 1987; Evans, 1985; McClelland & Judd, 1993). Numerous researchers (e.g., Evans, 1985; Morris, Sherman & Mansfield, 1986) argue that tests of hypotheses pertaining to the effects of moderators often have very low statistical power. In the context of MMR, power is the probability of rejecting a false null hypothesis of no moderating effect. If power is low, Type II statistical error rates are high and, thus, researchers may *erroneously* dismiss theoretical models that include moderating effects. In other words, in low power conditions, conclusions of null moderating effects may often be incorrect. Consequently, Cronbach (1987) noted the need for studies of statistical power issues associated with the use of MMR. Moreover, Dunlap and Kemery (1988) argued that research was needed to identify the conditions under which researchers are likely to be misled by the results of MMR analyses. Perhaps as a consequence of these calls for research and the all too frequent complaint that sound, theory-driven moderator variable hypotheses are not empirically supported (e.g., Jaccard, Helbig, Wan, Gutman & Kritz-Silverstein, 1990; Pablo, 1994; Zedeck, 1971), the last five years have witnessed a noticeable increase in the number of articles pertaining to the application of MMR, and low power conditions which may lead to invalid conclusions regarding moderating effects.[1]

The present article is motivated by the (1) increasing theoretical importance of moderating effects in numerous management subdisciplines, (2) widespread use of MMR to test hypotheses regarding moderating effects, and (3) growing concerns about statistical power problems in MMR-based tests of moderating effects. The remainder of this article is organized into the following sections: (1) brief description of MMR, (2) recent findings pertaining to artifacts that influence the power of MMR and thus threaten the validity of MMR-based conclusions, (3) solutions proposed regarding the impact of these artifacts on the power of MMR, and (4) issues related to MMR that deserve further investigation.

Moderated Multiple Regression (MMR)

MMR consists of comparing two least-squares regression equations (Cohen & Cohen, 1983). Given a criterion or dependent variable Y, a predictor X and a second predictor Z hypothesized to be a moderator, Equation 1 shows the sample-based ordinary least squares (OLS) regression that tests the additive model of the main effects for predicting Y from X and Z, such that:

$$Y = a + b_1 X + b_2 Z + e \qquad (1)$$

where

> $a =$ the least-squares estimate of the intercept
> $b_1 =$ the least-squares estimate of the population regression coefficient for X
> $b_2 =$ the least-squares estimate of the population regression coefficient for Z, and
> $e =$ a residual term.

This model assumes that the population data present the following five characteristics: (1) The expected value of the residual term is zero ($E(e) = 0$); (2) residuals are not correlated; (3) residuals exhibit constant variance (homoscedasticity) across values of each predictor; (4) covariance between the residual term and the predictors is zero; and (5) there is less than complete multicollinearity (Jaccard, Turrisi & Wan, 1990, pp. 15-16).

The second equation is formed by creating a new variable, the product between the predictors (i.e., X^*Z), and including it as a third term in the regression, yielding the following equation:

$$Y = a + b_1 X + b_2 Z + b_3 X^* Z + e \qquad (2)$$

where b_3 is the sample-based least squares estimate of the population regression coefficient for the product term.

To test for the statistical significance of the moderating effect, the coefficients of determination (i.e., squared multiple correlation coefficients, R^2) are compared for Equation 1 (R_1^2) and Equation 2 (R_2^2). An F-statistic

(distributed with $k_2 - k_1$ and $N - k_2 - 1$ degrees of freedom) is computed using the following formula:

$$F = [(R_2^2 - R_1^2) / (k_2 - k_1)] / [(1 - R_2^2) / (N - k_2 - 1)] \qquad (3)$$

where k_2 is the number of predictors in Equation 2, k_1 is the number of predictors in Equation 1, and N is the total sample size. Alternatively, a t-statistic can be computed to test the null hypothesis of $\beta_3 = 0$. (The p values associated with the t and F tests are identical, Cohen & Cohen, 1983).

MMR can be easily conducted on statistical software packages such as BMDP, SAS, and SPSS using the regression procedure (hierarchical multiple regression). First a new variable is created (X^*Z), and then a hierarchical regression analysis is conducted forcing variables X and Z into the equation predicting Y (see Equation 1), followed by a second step at which variable X^*Z is entered (see Equation 2). All three packages provide statistics regarding R^2s for the first and the second model, as well as standardized and unstandardized regression coefficients. Also, all three packages compute an F-statistic (see Equation 3) based on the difference between the two R^2s. The significance of this F-statistic indicates the presence of an X^*Z interaction.

Artifacts Influencing the Power of MMR

Factors identified as detrimental to statistical power in MMR hypothesis tests are related to (1) variable distributions (predictor variable range restriction, error variance heterogeneity), (2) operationalizations of criterion and predictor variable (measurement error, inappropriate metrics, artificial dichotomization or polychotomization), (3) sample size (total sample size, sample size across moderator-based subgroups), and (4) predictor intercorrelation.[2] Next, each of these factors and their impact on the power of MMR are described.

Variable Distributions

Predictor variable range restriction. Range restriction is an ubiquitous phenomenon in management research, especially in research conducted in field settings (cf. Cook & Campbell, 1979; Linn, 1968; McClelland & Judd, 1993). Range restriction is an instance of nonrandom or biased sampling in that not all subjects in a population have an equal probability of being selected to be members of a sample (Alexander, Barrett, Alliger & Carson, 1986).

There are many examples of range restriction problems in management research. In human resources management research, for example, personnel selection procedures are a major cause of range restriction. Decisions regarding which individuals to select for an opening are frequently based on their standing on a predictor variable, X (e.g., test of job aptitude): Only those who obtain a score that exceeds a specific cutoff point x are selected (cf. Guion, 1991; Guion & Cranny, 1982), leading to range restriction on X. As a result, in tests of relationships between the predictor, X, and a criterion, Y (e.g., measure of job

performance), data will only be available for individuals whose scores on X exceed a defined cutoff score x (i.e., $[(X,Y)|X>x]$).

Results of two recent Monte Carlo studies (Aguinis & Stone-Romero, 1994; McClelland & Judd, 1993) showed that by manipulating range restriction on a predictor variable, X, the variance of $X*Z$ scores in a range-restricted sample was lower than the variance of $X*Z$ scores in the unrestricted population, and this affected the ability of MMR to detect a moderating effect. For example, results reported by Aguinis and Stone-Romero showed that for a total sample size of 300 and no range restriction, the power to detect a medium-size moderating effect (i.e., $f^2=.075$, cf. Aiken & West, 1991, p. 157) was .81. However, when the scores were sampled from only the top 80% of the population distribution, power decreased to .51, well below the recommended .80 level (Cohen, 1988). Thus, even in the presence of a relatively mild degree of range restriction, power loss poses a serious threat to the validity of MMR-based conclusions.

Error variance heterogeneity. In tests of *dichotomous* and other categorical moderator variable hypotheses (e.g., gender: $Z = 1$, females; $Z=2$, males), homogeneity of within-moderator-based subgroup error variances ($\sigma^2_{e_i}$) is systematically violated (Aguinis & Pierce, 1995; Alexander & DeShon, 1994; Dretzke, Levin & Serlin, 1982; Hsu, 1994). The error variance for each of the two moderator-based subgroups is:

$$\sigma^2_{e_i} = \sigma_{Y(i)}(1-\rho^2_{XY(i)}) \tag{4}$$

where $\sigma^2_{Y(i)}$ and $\rho_{XY(i)}$ are the Y variance and the X-Y correlation in each subgroup, respectively. In the presence of a moderating effect in the population, the X-Y correlations for the two moderator-based subgroups differ and, thus, the error terms necessarily differ.

Alexander and DeShon (1994) conducted a Monte Carlo study and found that under unequal subgroup sample size conditions, when the subgroup with the largest n (e.g., males) shows the largest residual variance (i.e., smallest X-Y relationship), power is reduced noticeably (see also DeShon & Alexander, 1994b). Consequently, they suggested that future research should explore alternatives to MMR's F test (e.g., a chi-square test) when testing hypotheses regarding categorical moderators.

Operationalizations of Predictor and Criterion Variables

Measurement error. The impact of unreliability of scores on the statistical power of MMR to detect moderating effects has been investigated using both conceptual and empirical approaches (e.g., Busemeyer & Jones, 1983; Dunlap & Kemery, 1988; Evans, 1985). Because constructs in most management specialties are rarely measured with perfect or near perfect reliability, the observed regression coefficients in MMR are usually attenuated. Stated differently, in situations of less than perfect reliabilities in the predictor scores, the reliability of the product term is adversely affected, and the sample-based regression coefficient associated with the product term $X*Z$ (b_3) underestimates

the population coefficient (i.e., β_3) (cf. Equation 2). Moreover, when the reliability of the *criterion* scores is also less than perfect, relationships between Y and the predictors (e.g., $X*Z$) are attenuated even more (Aguinis & Stone-Romero, 1994). Busemeyer and Jones (1983) (cf. Bohrnstedt & Marwell, 1978) provided the following expression which estimates the product term reliability based on predictor scores reliabilities for the case when both X and Z are standardized:

$$\rho_{xz,xz} = (\rho_{xz}^2 + \rho_{xx}\rho_{zz}) \ / \ (\rho_{xz}^2 + 1) \qquad (5)$$

As it is apparent from Equation 5, when $\rho_{xz}^2 = 0$ (i.e., the predictors X and Z are orthogonal), the reliability of the product term is reduced to the product of the reliabilities of the predictors.

Reports providing empirical evidence (Evans, 1985) showed that the estimated effect size for the product term was reduced when the reliabilities of the predictors were small. Also, Dunlap and Kemery (1988) examined the effects of the reliabilities of X and Z and their correlation on statistical power, and replicated previous findings by Evans. For example, when the reliabilities for X and Z were .50, for a high correlation between the predictors (e.g., .80), statistical power was .706. However, when the correlation between the predictors was lower (e.g., .20), power decreased to .561. Although Dunlap and Kemery's simulation did not vary sample size (i.e., $N = 30$ for all conditions), research by Paunonen and Jackson (1988) used a similar Monte Carlo design and corroborated Dunlap and Kemery's results for a larger sample size ($N = 100$).

In contrast to the research described above, a more recent Monte Carlo simulation (Aguinis & Stone-Romero, 1994) manipulated several values of both *predictor* and *criterion* reliability. The results proved to be discouraging for MMR users: Even for reliabilities considered appropriate in management and social science research in general (i.e., .80, Nunnally, 1978), the power to detect moderating effects was typically much smaller than the recommended level of .80 (Cohen, 1988). For example, in the absence of measurement error (i.e., perfect criterion and predictor variable reliabilities), the power to detect a large population effect with a sample size of 300 was above .85. However, when the reliabilities took on values of .80, the probability that the moderator would be detected dropped to less than .45.

Inappropriate metrics. Two issues have appeared in the MMR literature regarding the metrics utilized in operationalizing the constructs measured. The first pertains to the measurement of predictors (i.e., ratio vs. interval level scales), and the second pertains to the measurement of the criterion (i.e., scale coarseness).

(a) Ratio versus interval level of measurement. Several authors have argued that MMR analyses can only be conducted when predictor variables are measured on ratio scales (e.g., Southwood, 1978). This recommendation is based on the assumption that in the presence of lower-order level of measurement (e.g., interval), the zero point on the scale is arbitrary, and simple

additive transformations on predictor variable scores may change the statistical test of the product term. However, as demonstrated by the work of Arnold and Evans (1979), Friedrich (1982), and Jaccard et al. (1990), this recommendation is unwarranted. Moreover, Jaccard et al. (1990, p. 29) convincingly demonstrated that "it is entirely appropriate to evaluate interaction effects for interval level data."

   (b) Scale coarseness.   A second issue recently investigated by Russell and his colleagues (Bobko & Russell, 1994; Russell & Bobko, 1992; Russell, Pinto & Bobko, 1991) refers to criterion variable "scale coarseness." This phenomenon refers to the operationalization of a criterion variable that does not include sufficient scale points. This insufficient number of scale points results in possible information loss and, therefore, prevents a moderating effect from being detected.

   For instance, if the predictor X and hypothesized moderator Z are measured on 7-point Likert-type scales, the product term $X*Z$ has a possible range of 7 X 7 = 49 distinct responses. However, if Y is measured on a "coarse" 7-point scale (which is typically the case) rather than on a 49-point scale (which is typically never the case), information regarding the relationship between Y and $X*Z$ is lost, the population moderating effect is underestimated, and power drops inevitably.

   In an experiment supporting the aforementioned argument (Russell & Bobko, 1992), subjects were assigned to one of two conditions: They responded to a dependent variable (1) consisting of a 5-point Likert type scale, or (2) consisting of a graphic line segment on which subjects had to place a mark indicating their response. Results confirmed that the estimated size of the moderating effect was larger when respondents utilized the continuous scale.

   Artificial moderator dichotomization and polychotomization.   Despite the availability of MMR, numerous researchers still opt for the dichotomization of a continuous variable (e.g., a median split resulting in "high" and "low" subgroups) and then conduct an analysis of variance (ANOVA) using the artificially created subgroups. Cohen (1983) demonstrated that this practice is inappropriate in probing the main effects because ANOVA subsequent to artificial subgrouping has substantially lower power than multiple regression. Recently, Stone-Romero and Anderson (1994) demonstrated that artificial dichotomization of a continuous predictor also leads to substantial power loss in tests for moderator variables.[3] They examined two tests for detecting moderators: (1) comparing X-Y correlations across artificially created moderator-based groups (e.g., high, medium, low), and (2) MMR based on the original untransformed moderator scores. Results indicated that MMR yielded higher power rates for virtually all conditions of sample size, reliability of predictor scores, and number of k subgroups created based on the moderator. Thus, they recommended against using MMR or any other technique based on artificially dichotomized or polychotomized variables.

Sample Size

   Total sample size.   Sample size is positively related to the statistical power of any inferential test (Cohen, 1988). The size of the sample on which the MMR

analysis is performed is perhaps one of the most important single factors affecting power. In recent years, several Monte Carlo simulations have explored the effects of sample size on the power of MMR to detect moderators (e.g., Alexander & DeShon, 1994; Stone-Romero & Anderson, 1994). Stone-Romero and Anderson, for example, found that what they defined as a small effect size was typically undetected when sample size was as large as 120, and unless a sample size of at least 120 was used, even medium and large moderating effects were, in general, also undetected.

Unequal sample size across moderator-based subgroups. When the moderator Z is a *categorical* variable such as gender or ethnicity, it might be the case that there are unequal sample sizes across the levels of Z, such that there are less minority (e.g., $Z=1$, n_1) than majority (e.g., $Z=2$, n_2) subgroup members. As a consequence of this situation, which is typical of studies in such fields as human resources management (e.g., Hattrup & Schmitt, 1990; Hunter, Schmidt & Rauschenberger, 1984), the power to detect ethnicity or gender as a moderator variable is reduced.

In general terms, Hsu (1993) showed that the *effective* total sample size (n') for two independent sample tests of means, correlations, and proportions is the harmonic mean of the two subgroup sample sizes:

$$n' = [2(n_1 n_2)] / [n_1 + n_2] \tag{6}$$

Consequently, in unequal subgroup sample sizes situations (i.e., $n_1 \neq n_2$), when the size of one of the subgroups is fixed at n_1, the statistical power of an inferential test cannot exceed the power of a test involving two subgroups, each of size $2(n_1)$, regardless of the size of the second subgroup.

In the specific context of MMR, the reduction of statistical power is due, in addition, to the nature of the product term $X*Z$, which is a composite of a continuous and a dichotomous (e.g., male-female, majority-minority) variable. The reason for this reduction is that the power to detect the moderating effect depends upon the strength of the sample-based semi-partial correlation between the criterion variable and the product term (i.e., $r_{Y(X*Z.XZ)}$). However, this semi-partial correlation is based on $X*Z$ scores, and these scores are a composite including a dichotomous variable. Thus, a ceiling is placed on the possible magnitude of the sample-based estimate of the population semi-partial correlation depending on the proportion of cases in subgroups (Nunnally, 1978). The maximum value for the correlation occurs when the subgroup proportions are equal (i.e., $p=.50$), but as p departs from .50, the ceiling on the sample-based semi-partial correlation declines, and there is a concomitant decrease in the statistical power to detect the moderator.

A recent empirical examination of this issue using Monte Carlo simulations (Stone-Romero, Alliger & Aguinis, 1994) corroborated the theoretical prediction. Total sample size and number of cases in one subgroup relative to total sample size were manipulated ($p_1 = n_1/N = .10$, .30, or .50). The results showed that there was a considerable decrease in power when the size of subgroup 1 was .10 relative to total sample size regardless of total sample size

(30, 60, 180, 300) and size of the moderating effect in the population (small, medium, large). The effect of unequal subgroup proportions on statistical power was significant above and beyond the effect of total sample size. A proportion of .30, closer to the optimum value of .50, also reduced the statistical power of MMR, but to a lesser extent.

Predictor Intercorrelation

In MMR analyses, predictor scores (X and Z) are used to compute the product term ($X*Z$) which carries information about the interaction. Thus, X and $X*Z$, and Z and $X*Z$ tend to be highly correlated (i.e., multicollinear). Some researchers (Morris et al., 1986; Smith & Sasaki, 1979) have argued that the presence of multicollinearity in MMR leads to an ill-conditioned solution in which the regression coefficients are unstable, error terms are larger, and power is decreased. In high multicollinearity situations, small observed score changes due to measurement error may be magnified and result in large changes in **B** (i.e., the vector of unstandardized regression coefficients) and, consequently, there is a larger capitalization on chance. Thus, because multicollinearity is virtually guaranteed in MMR, and is known to lead to unstable coefficients (including b_3), it has been posited that the power of MMR is not sufficient to detect moderating effects (Morris et al., 1986).

Given this supposed power problem, two strategies have been proposed to mitigate multicollinearity: (a) "centering" predictor variables. The most common centering approach is to subtract the mean from each score (i.e., $X' = X - \bar{X}$; $Z' = Z - \bar{Z}$; $X'Z' = [X - \bar{X}][Z - \bar{Z}]$) (cf. Tate, 1984). Tate (1984, p. 253) illustrates how centering reduces collinearity by noting that the slope of $X*Z$ on X for a central value (mean) of Z is $\bar{Z}$, whereas the slope of $(X - \bar{X})(Z - \bar{Z})$ on $(X - \bar{X})$ at $Z = \bar{Z}$ is zero. (b) Morris et al. (1986) introduced the use of principal-components regression (PCR), which was advocated as not being as affected by multicollinearity and, consequently, as being a more powerful method than MMR for tests of moderator hypotheses.

Fortunately, recent developments suggest that concerns about the detrimental impact of multicollinearity on power are unwarranted. Cronbach (1987) stated that the effects of multicollinearity on MMR analyses are: (1) increased rounding error, (2) increased regression coefficient sampling errors, and (3) difficulty in regression coefficient interpretation, especially for lower-order terms (see also Aiken & West, 1991). However, Cronbach asserted that multicollinearity is *not* detrimental to the power of MMR, as Morris et al. contended. The reasons for the apparent loss of power are that (1) the number of predictors reduces the degrees of freedom for the numerator of the F ratio, and (2) in the presence of multicollinearity, additional predictors contribute little to the sum of squares for regression. Thus, in one example provided by Morris et al., it seemed that decreasing the number of predictors from 13 to 10 reduced collinearity and increased power. However, Cronbach showed that in the typical MMR analysis, there are two measured predictors, in addition to one derived from them. In this more typical situation, collinearity does not adversely affect power. It should be noted, however, that high multicollinearity may cause

computational problems and, thus, it is recommended that predictors be centered before computing the product term (Cronbach, 1987; Jaccard et al., 1990).

In addition to Cronbach's argument against the assumed detrimental impact of multicollinearity on power of MMR analyses, Dunlap and Kemery (1987) demonstrated that Morris et al.'s finding of a nonsignificant moderating effect with MMR and significant effects using PCR may have been a result of some artifact of PCR. Moreover, additional empirical evidence provided by a Monte Carlo study conducted by Mason and Perreault (1991) found that fears of the effects of multicollinearity are, in practice, not typically justified.

Solving Power Problems

The presence of the aforementioned artifacts may cause power rates to drop to values of .50 or lower. That is, unbeknownst to the researcher using MMR, support for a *correctly* specified theoretical model including a moderated relationship may be decided by the flip of a coin. In addition to hindering the advancement of management theory, low statistical power can also lead to incorrect conclusions regarding a host of issues for which MMR is used as a decision-making tool such as recruiting, hiring, and promoting employees (DeShon & Alexander, 1994b; Hattrup & Schmitt, 1990). What can researchers do to remedy power-related problems? First, and as is often preached in the research methods literature, the importance of fully considering research design (e.g., potential range restriction problems) and measurement issues (e.g., operationalization of variables, valid and reliable measurement of constructs) prior to the time of data collection cannot be overemphasized. For example, the impact of scale coarseness can be eliminated by using a continuous criterion scale: Responses to paper-and-pencil Likert-type questions can be recorded on a graphic line segment and then be measured manually (Russell & Bobko, 1992). However, this procedure may be regarded as impractical by most researchers because it is time consuming and prone to errors. Fortunately, a computer program that overcomes these shortcomings is available (see Aguinis, Bommer & Pierce, in press). This program administers questionnaires on IBM and IBM-compatible personal computers by prompting respondents to indicate their answer by clicking on a graphic line segment displayed on the screen. Then, responses are stored directly into an ASCII file thereby eliminating the need to perform any additional steps before conducting the subsequent MMR analysis.

A second design-based suggestion for solving power problems was recently advanced by McClelland and Judd (1993; see also Aiken & West, 1993a, 1993b) and consists of sampling extreme scores from the population so as to enhance variable variances and, concomitantly, the power of MMR. However, the results obtained using oversampling techniques may not generalize to the original population distribution, but to a distribution of extreme scores only.

Despite the availability of the aforementioned design-based strategies to improve the power of MMR, many circumstances may prevent researchers from

having control over these variables (e.g., research conducted in field settings with limited accessibility to subjects, unavailability of highly reliable measures). Thus, researchers may have to use the sample and measures in hand, even in the presence of factors suspected to threaten the power of MMR. In these situations, there is a pressing need to estimate the power of the MMR test given specified sample characteristics (e.g., sample size, measurement error, predictor range restriction), especially if a hypothesized moderator is not detected. Fortunately, there are ways to perform such an estimation.

First, tables are readily available which allow investigators to estimate the sample sizes associated with the consensually acceptable power rate of .80 (Cohen, 1988) for $\alpha = .05$ for models including a continuous predictor and a continuous moderator (Aiken & West, 1991: 159, Table 8.2; Jaccard et al., 1990, p. 37, Table 3.1). These tables can also be used *prior* to data collection as a power analysis tool (cf. Cohen, 1988) (i.e., to estimate sample size needed to reach a .80 power level). In order to use these tables, MMR users need to estimate the values for the R^2s derived from Equations 1 and 2 above. These values can be estimated based on the sample in hand. Second, a computer program is available (see Aguinis, Pierce & Stone-Romero, 1994) that estimates power rates for $\alpha = .05$ in continuous predictor-dichotomous moderator models based on specific values for (1) total sample size, (2) sample sizes across the two categories of the hypothesized dichotomous moderator, and (3) correlation coefficients between predictor and criterion scores for each of the two moderator-based subgroups.

Despite the usefulness of the tables and computer program described above, they have at least two shortcomings. First, they fail to take into account several variables known to affect the power of MMR to detect moderating effects (e.g., predictor range restriction, unreliability of criterion scores). Second, the artifacts reviewed in this article not only have independent effects on power, but, perhaps more importantly, interactive (i.e., nonadditive) effects as well (Aguinis & Stone-Romero, 1994). Because interactive effects may have an even more detrimental effect on power than simple additive effects, the simultaneous presence of artifacts threatens the validity of MMR-based conclusions even to a greater extent (note that the Aguinis, Pierce & Stone-Romero program does consider interactive effects, but only for the three parameters mentioned above, and for the detection of *dichotomous* moderators only). Because of these limitations, future research using MMR is in great need of further developments regarding power calculations.

Once a researcher determines the existence of a low power situation, there are several possible courses of action. Unfortunately, none of them may be fully satisfactory. First, sample size can be increased, but practical considerations may not allow researchers to utilize this strategy. Second, statistical corrections could be implemented, for example, to estimate relationships between variables that suffer from range restriction or measurement error (Ghiselli, Campbell & Zedeck, 1981). However, this strategy raises questions about the appropriateness of testing moderator variable hypotheses using corrected estimates of squared semi-partial correlation coefficients (Cohen & Cohen, 1983). For

example, the standard error for the corrected estimate may not be known and, even if it is known, it may increase after performing the correction. Therefore, overall, there may be little gain over the test of the uncorrected estimate. A third strategy consists of compensating for low power by increasing the Type I error rate above the traditional levels, for example to $\alpha = .10$. However, researchers may be reluctant to raise the pre-set significance level above .05. A fourth alternative advanced by Darrow and Kahl (1982) consists of forcing the $X*Z$ term into the equation predicting Y before entering the X and Z terms. However, this strategy has been extensively criticized and should *not* be utilized because it violates basic principles of multiple regression (Stone, 1988; Wise, Peters & O'Connor, 1984).

Suggestions for Future Research

Table 1 summarizes (1) the artifacts affecting the power of MMR, (2) proposed solutions to alleviate low-power situations, and (3) shortcomings of these solutions. As indicated in this table, and perhaps because of researchers' frustration with the inability to detect moderating effects using MMR, there is a periodic appearance of new methods to detect moderators, such as those suggested by Darrow and Kahl (1982) and Morris et al. (1986). However, these purportedly superior techniques have been shown to be inadequate, or simply incorrect (Cronbach, 1987; Wise et al., 1984). Thus, and despite the power problems described above, McClelland and Judd (1993, p. 377) recently confirmed that there is "no credible published refutation of the appropriateness of testing the reliability of the partial regression coefficient for the product [$X*Z$] as a test of moderator effects." However, a logical extension of McClelland and Judd's statement is that more research is needed regarding the conditions under which MMR-based conclusions may be invalid, and about ways to revert these conditions. The fact that none of the simulation studies reviewed in this article completely account for low power rates suggests that there are still unexamined artifacts which have a detrimental effect on MMR. Thus, future research can pursue at least three avenues.

One issue which warrants attention is a closer examination of the criterion variable. As Bobko and Russell (1994) correctly noted, research on MMR has typically focused on properties of the predictor variables. Simulation work is needed to evaluate the simultaneous impact of various characteristics of the dependent variable (e.g., measurement error, scale coarseness, nonnormality) and predictor variables (e.g., range restriction) on the power of MMR.

A second area that deserves to be investigated is whether quadratic terms (e.g., X^2) should be systematically incorporated as covariates in MMR (Lubinski & Humphreys, 1990; Shepperd, 1991). Recent articles in *Psychological Bulletin* have debated this issue. Lubinski and Humphreys (1990; see also Cortina, 1993) argued that many reports on the existence of moderator variables may be incorrect because they are based on research in which the variance due to the quadratic effects of X or Z (represented by the X^2 and Z^2 terms in a regression equation) may be unduly attributed to the X by Z linear interaction (represented

Table 1. Summary of Artifacts Affecting the Power of MMR, Proposed Solutions, and Shortcomings of These Solutions

Artifact	Proposed Solution	Shortcomings
Predictor range restriction (Aguinis & Stone-Romero, 1994; McClelland & Judd, 1993).	Statistical corrections could be implemented to better estimate population parameters.	It may not be appropriate to use MMR using corrected estimates of squared semi-partial correlation coefficients (cf. Cohen & Cohen, 1983). Also, the standard error may increase after performing the correction so, in the end, there may be little gain over the test of the uncorrected estimate.
Error variance heterogeneity caused by dichotomous moderators (Aguinis & Pierce, 1995; Alexander & DeShon, 1994).	Use tests that do not require homogeneity of error variance across subgroups such as the chi-square test for equality of rs (Alexander & DeShon, 1994).	Chi-square test is, conceptually, not a test of equality of slopes across groups; it is a test of the equality of correlation coefficients across groups.[a]
Measurement error in predictor (e.g., Dunlap & Kemery, 1988; Evans, 1985) and criterion (Aguinis & Stone-Romero, 1994) variables.	Use measures of constructs that are valid and reliable. Use structural equation modeling (SEM) in lieu of MMR (Jaccard & Wan, 1995).	Measures with demonstrated validity and high reliability may not be available. The implementation of SEM may seem cumbersome to most researchers; user-friendly software packages are needed.
Level of measurement lower than ratio (Southwood, 1978).	"[I]t is entirely appropriate to evaluate interaction effects for interval level data" (Jaccard, Turrisi & Wan, 1990, p. 29).	
Scale coarseness (e.g., Bobko & Russell, 1994; Russell & Bobko, 1992).	Responses to paper-and-pencil Likert-type questions can be recorded on a graphic line segment and then be measured manually (Russell & Bobko, 1992). Also, use computer program by Aguinis, Bommer, and Pierce (in press).	
Small total sample size (e.g., Stone-Romero & Anderson, 1994) and unequal sample sizes across moderator-based subgroups (Stone-Romero, Alliger & Aguinis, 1994).	Increase total sample size. Increase sample size in smaller subgroup.	Practical constraints may not allow for this solution.

Table 1. (Continued)

Artifact	Proposed Solution	Shortcomings
High predictor intercorrelation (i.e., multicollinearity) (e.g., Morris, Sherman & Mansfield, 1986; Smith & Sasaki, 1979).	Multicollinearity does *not* have a detrimental impact on the power of MMR (Cronbach, 1987).	
Low power (in general).	Oversampling techniques (Aiken & West, 1993a, 1993b; McClelland & Judd, 1993).	Results obtained may not generalize to the original population distribution, but to a distribution of extreme scores only.
	Estimate statistical power *a priori or post-hoc* using tables provided by Aiken and West (1991, p. 159, Table 8.2) and Jaccard, Turrisi, and Wan (1990, p. 37, Table 3.1).	These tables do not take into account several variables known to affect the power of MMR to detect moderating effects (e.g., predictor range restriction, unreliability of criterion scores).
	Estimate statistical power *a priori or post-hoc* using computer program by Aguinis, Pierce, and Stone-Romero (1994).	This program *only allows* for the estimation of power of MMR to detect *dichotomous* moderators.
	Increase alpha level above the traditional .05 value.	Researchers (and journal editors) may be reluctant to implement this practice.
	Use "backward regression" (Darrow & Kahl, 1982) in lieu of MMR.	This strategy should *not* be used; it violates basic principles of multiple regression (Stone, 1988; Wise, Peters & O'Connor, 1984).
	Use Principal Components Regression (PCR) (Morris, Sherman & Mansfield, 1986) in lieu of MMR.	Dunlap and Kemery (1987) demonstrated that Morris et al.'s finding of a nonsignificant moderating effect with MMR and significant effects using PCR may have been a result of some artifact of PCR. PCR should *not* be preferred over MMR (Cronbach, 1987).

Note: a. Note that the tests for homogeneity of slopes and homogeneity of correlation coefficients are equivalent under the ideal (and perhaps rare) condition of equal variances in both X and Y (i.e., $\sigma_{x_1}^2 = \ldots = \sigma_{x_k}^2$ and $\sigma_{Y_1}^2 = \ldots = \sigma_{Y_k}^2$) across the moderator-based subgroups (i.e., $Z_1 \ldots Z_k$). This equivalence is based on the fact that $\beta_i = \rho_i(\sigma_{Y_i}^2 / \sigma_{x_i}^2)$ (see Alexander & DeShon, 1994).

by $X*Z$). On the other hand, Shepperd (1991, p. 316) asserted that the recommendation to systematically investigate quadratic effects when using MMR "may be unwise and, at the very least, should be viewed cautiously" because this procedure may lead to spurious findings (i.e., falsely rejecting a null hypothesis) regarding *quadratic* effects. Empirical evidence to solve this controversy is seriously needed.

A third avenue for research is the closer scrutiny of structural equation modeling (SEM), which is also based on the general linear model, as a moderator detection strategy. There is preliminary work (Bollen, 1989, pp. 405-406; Jaccard & Wan, 1995; Kenny & Judd, 1984) indicating that SEM may be used in lieu of MMR, with the advantage that measurement error is incorporated into the model tested and, thus, power may be enhanced. However, the procedure may seem cumbersome to most researchers and, thus, user-friendly software packages which can easily handle moderator analyses in SEM are needed.

Conclusions

Although imperfect, MMR seems to be the preferred statistical method to detect moderating effects, especially in models with continuous predictor variables. MMR users should be aware that several artifacts may affect their conclusions regarding moderator variable hypotheses, leading to the *incorrect* inference that there is no moderating effect. Potential solutions to power problems include design considerations and a power analysis prior to data collection. In tests conducted at suspected low power rates, null findings should be interpreted cautiously. Finally, given the increasing importance of moderated relationships in numerous management subdisciplines, it is expected that in the next years there will be further research on MMR and other statistical procedures that can be used for detecting moderating effects. Hopefully, this will allow management researchers to further extend theoretical models to incorporate complex and rich moderated relationships.

Notes

1. A few researchers have argued that MMR may also lead to a second type of erroneous conclusion due to Type I error rates artificially inflated to levels above the pre-set nominal level (typically $\alpha=.05$) (e.g., DeShon & Alexander, 1994a; Lubinski & Humphreys, 1990). In these situations, MMR users may *erroneously* conclude that a moderating effect has been found and, thus, discover a "false" moderator. However, perhaps because management researchers may be more concerned about not discarding a correct model than accepting an incorrect one, most research efforts have been aimed at identifying artifacts leading to low power situations; consequently, these are the factors reviewed here.
2. In addition to these artifacts, Cohen (1988) has identified two factors that are positively related to the power of *any* inferential test (including MMR): (a) magnitude of the effect in the population, and (b) Type I error rate (i.e., alpha level) chosen by the researcher.
3. Note, however, that Maxwell and Delaney (1993) demonstrated that the simultaneous dichotomization of *more than one* predictor not only decreases the power of MMR but, in some situations, it may also artificially increase Type I error rates.

References

Aguinis, H. (1993). The detection of moderator variables using moderated multiple regression: The effects of range restriction, unreliability, sample size, and the proportion of subjects in subgroups (Doctoral dissertation, University at Albany, State University of New York, 1993). *Dissertation Abstracts International (B)*, 54/04: 2253.

Aguinis, H., Bommer, W.H. & Pierce, C.A. (in press). Improving the estimation of moderating effects by using computer-administered questionnaires. *Educational and Psychological Measurement*.

Aguinis, H. & Pierce, C.A. (1995). *Combatting heterogeneity of residual variance in moderator variable detection.* Paper presented at the annual meeting of the Rocky Mountain Psychological Association, Boulder, CO.

Aguinis, H., Pierce, C.A. & Stone-Romero, E.F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement, 54*: 690-692.

Aguinis, H. & Stone-Romero, E.F. (1994). *Methodological artifacts in moderated multiple regression: Effects on power.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Nashville, TN.

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

————. (1993a). *Detecting interactions in multiple regression: Measurement error, power, and design considerations.* Paper presented at the annual meeting of the American Psychological Association, Toronto.

————. (1993b). Detecting interactions in multiple regression: Measurement error, power, and design considerations. *The Score, 16*(1): 7, 14-15.

Alexander, R.A., Barrett, G.V., Alliger, G.M. & Carson, K.P. (1986). Toward a general model of nonrandom sampling and the impact of population correlations: Generalizations of Berkson's Fallacy and restriction of range. *British Journal of Mathematical and Statistical Psychology, 39*: 90-105.

Alexander, R.A. & DeShon, R.P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin, 115*: 308-314.

Arnold, H.J. & Evans, M.G. (1979). Testing multiplicative models does *not* require ratio scales. *Organizational Behavior and Human Performance, 24*: 41-59.

Bedeian, A.G. & Mossholder, K.W. (1994). Simple question, not so simple answer: Interpreting interaction terms in moderated multiple regression. *Journal of Management, 20*: 159-165.

Benson, P.G., Kemery, E.R., Sauser, W.I. & Tankesley, K.E. (1985). Need for clarity as a moderator of the role ambiguity-job satisfaction relationship. *Journal of Management, 11*: 125-130.

Bobko, P. & Russell, C.J. (1994). On theory, statistics, and the search for interactions in the organizational sciences. *Journal of Management, 20*: 193-200.

Bohrnstedt, G.W. & Marwell, G. (1978). The reliability of products of two random variables. Pp. 254-273 in K.F. Schuessler (Ed.), *Sociological methodology.* San Francisco: Jossey-Bass.

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

Bruton, G.D., Oviatt, B.M. & White, M.A. (1994). Performance of acquisitions of distressed firms. *Academy of Management Journal, 37*: 972-989.

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

Chaplin, W.F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59*: 143-178.

Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*: 115-124.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*: 249-253.

————. (1988). *Statistical power analysis for the behavioral sciences,* 2nd ed. Hillsdale, NJ: Erlbaum.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences,* 2nd ed. Hillsdale, NJ: Erlbaum.

Compas, B.E., Ey, S. & Grant, K.E. (1993). Taxonomy, assessment, and diagnosis of depression during adolescence. *Psychological Bulletin, 114*: 323-344.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Cordes, C.L. & Dougherty, T.W. (1993). A review and integration of research on job burnout. *Academy of Management Review, 18*: 621-656.

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

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

Darrow, A.L. & Kahl, D.R. (1982). A comparison of moderated regression techniques considering strength of effect. *Journal of Management, 8*: 35-47.

DeShon, R.P. & Alexander, R.A. (1994a). A generalization of James's second-order approximation to the test for regression slope equality. *Educational and Psychological Measurement, 54*: 328-335.

_____. (1994b). *Power of the test for regression slope differences in differential prediction research*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Nashville, TN.

Dretzke, B.J., Levin, J.R. & Serlin, R.C. (1982). Testing for regression homogeneity under variance heterogeneity. *Psychological Bulletin, 91*: 376-383.

Duarte, N.T., Goodson, J.R. & Klich, N.R. (1994). Effects of dyadic quality and duration on performance appraisal. *Academy of Management Journal, 37*: 499-521.

Dunlap, W.P. & Kemery, E.R. (1987). Failure to detect moderating effects: Is multicollinearity the problem? *Psychological Bulletin, 102*: 418-420.

_____. (1988). Effects of predictor intercorrelations and reliabilities on moderated multiple regression. *Organizational Behavior and Human Decision Processes, 41*: 248-258.

Evans, M.G. (1985). A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis. *Organizational Behavior and Human Decision Processes, 36*: 305-323.

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

Ghiselli, E.E., Campbell, J.P. & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, CA: Freeman.

Guion, R.M. (1991). Personnel assessment, selection, and placement. Pp. 327-397 in M.D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology*. Palo Alto, CA: Consulting Psychologists Press.

Guion, R.M. & Cranny, C.J. (1982). A note on concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology, 67*: 239-244

Hall, J.A. & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs, 58*: 437-448.

Hattrup, K. & Schmitt, N. (1990). Prediction of trades apprentices' performance on job sample criteria. *Personnel Psychology, 43*: 453-466.

Hunter, J.E., Schmidt, F.L. & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. Pp. 41-99 in C.R. Reynolds & R.T. Brown (Eds.), *Perspective on bias in mental testing*. New York: Plenum Press.

Hsu, L.M. (1993). Using Cohen's tables to determine the maximum power attainable in two-sample tests when one sample is limited in size. *Journal of Applied Psychology, 78*: 303-305.

_____. (1994). More on transformations and moderated regression analysis: Advantages of additivity and homoscedasticity transformations. *Journal of Applied Behavioral Science, 30*: 217-226.

Jaccard, J.J., Helbig, D.W., Wan, C.K., Gutman, M.A. & Kritz-Silverstein, D.C. (1990). Individual differences in attitude-behavior consistency: The prediction of contraceptive behavior. *Journal of Applied Social Psychology, 20*: 575-617.

Jaccard, J.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.

Kenny, D.A. & Judd, C.M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 90*: 201-210.

Linn, R.L. (1968). Range restriction problems in the use of self-selected groups for test validation. *Psychological Bulletin, 69*: 69-73.

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.

Lyness, S.A. (1993). Predictors of differences between Type A and B individuals in heart rate and blood pressure reactivity. *Psychological Bulletin, 114*: 266-295.

Mason, C.H. & Perreault, W.D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research, 28*: 268-280.

Maxwell, S.E. & Delaney, H.D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*: 181-190.

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

Morris, J.H., Sherman, J.D. & Mansfield, E.R. (1986). Failures to detect moderating effects with ordinary least squares-moderated multiple regression: Some reasons and a remedy. *Psychological Bulletin, 99*: 282-288.

Nesler, M.S., Aguinis, H., Quigley, B.M. & Tedeschi, J.T. (1993). The effect of credibility on perceived power. *Journal of Applied Social Psychology, 23*: 1407-1425.

Nunnally, J.C. (1978). *Psychometric theory*, 2nd ed. New York: McGraw-Hill.

Pablo, A.L. (1994). Determinants of acquisition integration level: A decision-making perspective. *Academy of Management Journal, 37*: 803-836.

Paunonen, S.V. & Jackson, D.N. (1988). Type I error rates for moderated multiple regression analysis. *Journal of Applied Psychology, 73*: 569-573.

Russell, C.J. & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. *Journal of Applied Psychology, 77*: 336-342.

Russell, C.J., Pinto, J. & Bobko, P. (1991). Appropriate moderated regression and inappropriate research strategy: A demonstration of the need to give your respondents space. *Applied Psychological Measurement, 15*: 257-266.

Sackett, P.R. & Wilk, S.L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*: 929-954.

Saunders, D.R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16*: 209-222.

Schmitt, N., Hattrup, K. & Landis, R.S. (1993). Item bias indices based on total test score and job performance estimates of ability. *Personnel Psychology, 46*: 593-611.

Shepperd, J.A. (1991). Cautions in assessing spurious "moderator effects." *Psychological Bulletin, 110*: 315-317.

Smith, K.W. & Sasaki, M.S. (1979). Decreasing multicollinearity: A method for models with multiplicative functions. *Sociological Methods and Research, 8*: 35-56.

Snell, S.A. & Dean, J.W. Jr. (1994). Strategic compensation for integrated manufacturing: The moderating effects of jobs and organizational inertia. *Academy of Management Journal, 37*: 1109-1140.

Southwood, K.E. (1978). Substantive theory and statistical interaction: Five models. *American Journal of Sociology, 83*: 1154-1203.

Stiff, J.B. (1986). Cognitive processing of persuasive message cues: A meta-analytic review of the effects of supporting information on attitudes. *Communication Monographs, 53*: 75-89.

Stone, E.F. (1988). Moderator variables in research: A review and analysis of conceptual and methodological issues. Pp. 191-229 in G.R. Ferris & K.M. Rowland (Eds.), *Research in personnel and human resources management*, Vol. 6. Greenwich, CT: JAI Press.

Stone, E.F. & Hollenbeck, J.R. (1984). Some issues associated with the use of moderated regression. *Organizational Behavior and Human Performance, 34*: 195-213.

————. (1989). Clarifying some controversial issues surrounding statistical procedures for detecting moderator variables: Empirical evidence and related matters. *Journal of Applied Psychology, 74*: 3-10.

Stone-Romero, E.F., Alliger, G.M. & Aguinis, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects of dichotomous variables. *Journal of Management, 20*: 167-178.

Stone-Romero, E.F. & Anderson, L.E. (1994). Techniques for detecting moderating effects: Relative statistical power of multiple regression and the comparison of subgroup-based correlation coefficients. *Journal of Applied Psychology, 79*: 354-359.

Tate, R.L. (1984). Limitations of centering for interactive models. *Sociological Methods and Research, 13*: 251-271.

Whisman, M.A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114*: 248-265.

Williams, K.J. & Alliger, G.M. (1994). Role stressors, mood spillover, and perceptions of work-family conflict in employed parents. *Academy of Management Journal, 37*: 837-868.

Wise, S.L., Peters, L.H. & O'Connor, E.J. (1984). Identifying moderator variables using multiple regression: A reply to Darrow & Kahl. *Journal of Management, 10*: 227-236.

Zedeck, S. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin, 76*: 295-310.