A Generalized Solution for Approximating the Power to Detect Effects of Categorical Moderator Variables Using Multiple Regression

HERMAN AGUINIS
University of Colorado at Denver

ROBERT J. BOIK
CHARLES A. PIERCE
Montana State University

Investigators in numerous organization studies disciplines are concerned about the low statistical power of moderated multiple regression (MMR) to detect effects of categorical moderator variables. The authors provide a theoretical approximation to the power of MMR. The theoretical result confirms, synthesizes, and extends previous Monte Carlo research on factors that affect the power of MMR tests of categorical moderator variables and the low power of MMR in typical research situations. The authors develop and describe a computer program, which is available on the Internet, that allows researchers to approximate the power of MMR to detect the effects of categorical moderator variables given user-input information (e.g., sample size, reliability of measurement). The approximation also allows investigators to determine the effects of violating certain assumptions required for MMR. Given the typically low power of MMR, researchers are encouraged to use the computer program to approximate power while planning their research design and methodology.

Researchers in numerous organization studies disciplines are interested in estimating interactive effects involving a categorical and a continuous variable. For example, differential prediction is operationalized as an interaction between ethnicity (e.g., minor-

Authors' Note: This research was supported in part by a grant from the Graduate School of Business Administration (University of Colorado at Denver) to Herman Aguinis. A previous version of this article was presented as part of a symposium (J. M. Cortina, chair) at the meeting of the Society for Industrial and Organizational Psychology, Dallas, TX, April 1998. Portions of the research reported herein were conducted while Herman Aguinis was on sabbatical leave from the University of Colorado at Denver and holding visiting appointments at China Agricultural University–The International College of Beijing (People's Republic of China), City University of Hong Kong (People's Republic of China), Nanyang Technological University (Singapore), University of Science Malaysia (Penang, Malaysia), and University of Santiago de Compostela (Spain). Correspondence and reprint requests should be addressed to Herman Aguinis, Graduate School of Business Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217-3364.

Organizational Research Methods, Vol. 4 No. 4, October 2001 291-323

ity, nonminority) and test scores (e.g., general cognitive ability) on a measure of performance (e.g., supervisory ratings) (Cleary, 1968). If differential prediction is found, ethnicity is labeled a "moderator" of the relationship between test scores and performance. Gender is another illustration of a categorical variable whose moderating effect is of interest in several organization studies disciplines. Numerous additional examples were reviewed recently by Aguinis, Beaty, Boik, and Pierce (2000).

Moderated Multiple Regression (MMR)

The moderating effect of a categorical variable on the relationship between a continuous predictor and a continuous criterion is typically estimated using MMR. Assume that X is a continuous predictor variable (e.g., preemployment test scores), Y a continuous criterion variable (e.g., supervisory ratings of performance), and Z a categorical predictor variable hypothesized to be a moderator (e.g., gender, dummy coded 1 = men and 2 = women). Equation 1 shows the linear regression model for predicting Y from X, Z, and the interaction between X and Z (i.e., moderating effect of Z) represented by the $X \bullet Z$ product term (Aiken & West, 1991; Cohen & Cohen, 1983; Saunders, 1956):

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \bullet Z + \varepsilon,$$

where β_0 is the intercept, β_1 is the regression coefficient for X, β_2 is the regression coefficient for Z, β_3 is the regression coefficient for the product term that carries information about the interaction between X and Z, and ε is a normally distributed random error term. Rejecting the null hypothesis H_0 that the product term's regression coefficient is zero indicates the presence of a moderating or interaction effect. Stated differently, rejecting this null hypothesis indicates that the regression of Y on X is unequal across levels of Z (e.g., male and female subgroups). Note that although this illustration addresses a binary moderator variable (i.e., two levels), the MMR model allows the categorical moderator to take on any number of levels (e.g., a moderator with three levels could be ethnicity coded with African American, Latino/Latina, and White categories).

Statistical Power Problems With MMR

MMR is widely used to test hypotheses regarding the effects of categorical moderator variables in organization studies (e.g., Bobko & Russell, 1994), as well other fields including education (Aguinis, Nesler, Quigley, Lee, & Tedeschi, 1996), marketing (e g., Mason & Perreault, 1991), and sociology (e.g., Smith, & Sasaki, 1979), among others.

Despite its pervasive use, researchers have lamented the low statistical power of MMR for nearly three decades (Aguinis, 1995; Aguinis, Bommer, & Pierce, 1996; McClelland & Judd, 1993; Zedeck, 1971). When MMR analyses are conducted at low levels of statistical power, researchers who fail to find support for their hypotheses regarding moderating effects do not know whether (a) their hypotheses are incorrect, or (b) their hypotheses are correct, but they failed to detect the moderating effect. This situation creates great uncertainty in theory development involving hypothesized moderating effects. Uncertainty regarding the presence of moderating effects is particularly serious because moderator variables are considered to be "at the very heart of the scientific enterprise" (Hall & Rosenthal, 1991, p. 447) and to serve as indicators of a field's scientific advancement and maturity (Hall & Rosenthal, 1991).[1]

The following factors have been identified as culprits for the typical low power of MMR: (a) reduced variance in the predictor variables (McClelland & Judd, 1993, also referred to as "range restriction" by Aguinis & Stone-Romero, 1997), (b) error variance heterogeneity across moderator-based subgroups (Alexander & DeShon, 1994), (c) measurement error (Busemeyer & Jones, 1983), (d) small total sample size (e.g., Alexander & DeShon, 1994), and (e) unequal sample size across the moderator-based subgroups (Stone-Romero, Alliger, & Aguinis, 1994) (see Aguinis, 1995, and Aguinis & Pierce, 1998a, for reviews). In practice, however, researchers often have little control over the size of their samples or the reliability levels of their measures. Thus, it is often the case that researchers conduct their MMR analyses under low power conditions. As a result, researchers may erroneously conclude that there is no moderating effect.

Given that most of the factors known to affect the power of MMR are not under the control of researchers and that researchers may not be aware of the low power of MMR in the sample at hand, Aguinis and colleagues (Aguinis & Pierce, 1998b; Aguinis, Pierce, & Stone-Romero, 1994) developed computer programs to estimate the power of MMR. These programs allow researchers to estimate the power of MMR tests for specific situations (e.g., large vs. small moderating effect magnitude). Despite the fact that these programs are available and useful to researchers in the quest for moderating effects, they suffer from five limitations. These limitations exist because the programs are based on algorithms derived from empirical (i.e., Monte Carlo) studies (Aguinis & Pierce, 1998b, is based on Aguinis & Stone-Romero, 1997; and Aguinis et al., 1994, is based on Stone-Romero et al., 1994).

The first limitation is that these programs do not include all the factors known to affect the power of MMR. More specifically, they assume that the measures of Y, X, and Z are free from measurement error. This is a tenable assumption for the categorical moderator Z but untenable for the predictor X and the criterion Y.

The second limitation of these programs is that the Monte Carlo studies on which they are based included only a limited range of values for factors affecting the power of MMR. For instance, Aguinis and Stone-Romero (1997) used values for total sample sizes of 60 and 300. Thus, the program by Aguinis and Pierce (1998b) based on the simulation results by Aguinis and Stone-Romero might not provide accurate power estimates for situations having sample size values far from 60 or 300. In addition, the Aguinis and Stone-Romero simulation did not include a negative relationship between X and Y for any of the moderator-based subgroups. Although these relationships are typically positive in the context of educational and preemployment testing, this is not the case in other research areas. For example, a health psychologist might want to test the moderating effect of gender on the negative relationship between the predictor "optimism" and the criterion "time to recover from surgery." As will be illustrated later in this article, using computer programs based on Monte Carlo simulation results in which certain value ranges were not included (e.g., negative correlations) can lead to inaccurate power estimates.

The third limitation of the currently available programs to compute the power of MMR is that restriction on X is assumed to take on only the simplest form of truncation. That is, the programs allow users to specify whether X scores are truncated at a specific cutoff point. This truncation is known to lower the power of MMR when *X* is normally distributed (Aguinis & Stone-Romero, 1997). However, simple truncation is only one of many situations that result in sample variances being smaller than population variances. For instance, in many situations, the probability that an individual will be selected in a sample does not depend only on his or her standing on variable *X* but also on his or her standing regarding other measured and unmeasured variables (Aguinis & Whitehead, 1997). More precisely, in educational and preemployment testing situations, many of the top scorers turn down an offer, so there might not be a precise score on *X* above which individuals have a higher probability of being included in the sample (Murphy, 1986). In short, differences between the sample and population variances for *X* might be due to a more complex sampling restriction mechanism, and the available programs to compute power do not allow for this perhaps more frequent research scenario.

The fourth limitation is that investigators can only compute power in situations in which the categorical moderator has two levels. This is a useful feature for researchers interested in testing the effects of binary moderator variables such as gender. However, there are many research situations in which the categorical moderator takes on more than two values (e.g., ethnicity coded using African American, Latino/Latina, and White categories). Although two levels are very frequently observed, researchers might also be interested in investigating the effects of moderators with more than two categories. However, at present there is no tool available to compute the power of MMR in these research situations.

Finally, the fifth limitation of existing programs is that they include situations with one continuous predictor *X* and one categorical moderator. Thus, existing programs do not allow researchers to compute the power of MMR in situations including two or more categorical moderator variables. One such common situation is an MMR model including the categorical moderators ethnicity and gender (cf. Aguinis et al., 2000). In such situations, a researcher might wish to compute the power to detect the *X* by ethnicity interaction, the *X* by gender interaction, and the ethnicity by gender interaction. In addition, if an interaction is expected, then a researcher might wish to compute the power to detect interaction contrasts among the *Y* on *X* slopes (see Boik, 1979, 1993, for a discussion of main and interaction effect contrasts). Computing power for these types of effects is not possible with the programs available at present.

Present Study

Given the limitations of previous efforts to produce highly accurate estimates of the power of MMR to detect the effects of categorical moderator variables, the goal of the present article is to develop a theory-based solution for approximating power. Overcoming limitations of previous empirically based research, we offer a theoretical result that allows researchers to approximate the power of MMR. First, we describe the theoretical approximation. Second, we present results of a Monte Carlo simulation showing its accuracy. Third, we describe a user-friendly computer program (MMRPOWER) available on the World Wide Web that we developed to implement the theoretical power approximation given user-input values. Thus, MMRPOWER allows users to compute power given their precise expected or actual situation (e.g., total sample size, sample sizes in each of the moderator-based subgroups, measurement error for *X* and *Y* in each of the moderator-based subgroups). Finally, we use MMRPOWER to generate values for a number of typical situations in organization studies to examine the relative effects of the various factors that influence power.

In the sections that follow, we use the term *power* as a synonym of *power function*, which is the probability of rejecting the null hypothesis H_0 given specified values for all of the parameters (Casella & Berger, 2002). When H_0 is true and the homogeneity of error variance assumption is satisfied, then power equals the nominal preset rejection rate α (also referred to as Type I error rate or test size). However, when H_0 is true and the homogeneity assumption is not satisfied, then power can be greater or smaller than nominal α. Thus, as we discuss later in more detail, MMR should not be used when the homogeneity of error variance assumption is violated (Aguinis, Petersen, & Pierce, 1999; Aguinis & Pierce, 1998a).

An Analytic Approximation to the Power of MMR

Assumptions of the Model

Appendix A describes the assumptions used in developing the analytic approximation. Briefly, it is assumed that either *X* and/or *Y* could be measured with error. Regardless of the reliability of *X* and *Y*, it is assumed that a normal linear regression model holds for *Y* conditional on *X*, where *Y* and *X* are the observable scores rather than the true scores. This differs from the conventional errors-in-variables model in which Y_{true} conditional on X_{true} is assumed to follow a regression model (Brown & Fuller, 1990; Carroll & Ruppert, 1995). The MMR model is more appropriate when interest is in the relationship among the observable *X* and *Y* scores.

The analytic approximation allows for sampling restrictions that lead to a difference between the expected sample variance of *X* and the population variance of *X*. We use the term *variance multiplying factor* to refer to the expected sample variance of *X* divided by population variance of *X*. This ratio need not be the same in each moderator-based subgroup. Variance multiplying factors that differ from 1 can arise because of truncation (i.e., scores are included in the sample only if they are above or below a specific cutoff point) or other sampling restrictions. If *X* is normally distributed, then MMRPOWER will compute the variance multiplying factor. If *X* is not normally distributed, or if a sampling restriction other than truncation holds, then the user needs to input a value for the variance multiplying factor.

Factors Affecting Power

Appendix B provides a technical presentation of the null hypothesis, the MMR model, and the *F* statistic used in assessing the presence of a moderating effect in MMR. Appendix B defines the various components that are used in deriving the analytic approximation to power in Appendix C. Appendix C presents (a) the distribution of the *F* statistic conditional on *X* and (b) an approximation to the unconditional distribution of the *F* statistic. Note that Gatsonis and Sampson (1989) discussed the distinction between conditional and unconditional power in a simpler model than that considered here.

Theorem 2 in Appendix C gives an analytic expression for the power of the MMR F test. The expression is technical, but this does not limit the practical usefulness of the theorem. The required computations are readily performed using MMRPOWER (see the Computer Program section below).

An examination of the analytic power approximation given in Theorem 2 reveals that the power of the MMR F test depends on the following quantities: (a) preset nominal test size, α; (b) number of moderator-based subpopulations, k; (c) sample sizes across moderator-based subgroups, n_j for $j = 1, \ldots, k$; (d) difference in slopes of Y on X across moderator-based subpopulations, $\beta_j - \beta_k$ for $j = 1, \ldots, k - 1$; (e) reliabilities (e.g., Cronbach's alpha, split-half, test-retest) for Y in the k moderator-based subpopulations, $\alpha_{y,j}$ for $j = 1, \ldots, k$; (f) reliabilities for X in the k moderator-based subpopulations, $\alpha_{y,j}$ for $j = 1, \ldots, k$; (g) correlations between X and Y in each of the moderator-based subpopulations, ρ_j for $j = 1, \ldots, k$; (h) marginal variance of Y in the k subpopulations, $\sigma^2_{y,j}$, for $j = 1, \ldots, k$; (i) variance of X in the k subpopulations, $\sigma^2_{x,j}$ for $j = 1, \ldots, k$; and (j) ratio of expected sample variance of X/population variance of X. Note, however, that some of these factors are not independent. For example, slopes are functions of reliabilities, correlations, marginal variances of Y, and variances of X. That is, for slopes to change, at least one of these factors also needs to change.

Theorem 2 provides a theory-based synthesis of previous research regarding the variables that affect the power of MMR. For instance, consider the following variables: (a) sampling restriction on X (Aguinis & Stone-Romero, 1997), (b) measurement error (e.g., Busemeyer & Jones, 1983), and (c) unequal sample sizes across moderator-based subgroups (Stone-Romero et al., 1994). Sampling restriction is incorporated by allowing the expected sample variance of X to differ from the population variance of X. Stated differently, the analytic solution goes beyond simple truncation (e.g., an individual is included in the sample if his or her X score is above a specific cutoff point) and addresses the more general issue that differences between sample and population variances affect power (McClelland & Judd, 1993). Measurement error is incorporated by allowing reliabilities for X and Y to differ across moderator-based subgroups. Sample sizes across moderator-based subgroups, which are related to the variance of the moderator variable, also are explicitly included in the model.[2]

Theorem 2 also reveals that power decreases as the reliability of X and/or Y decreases. Furthermore, to a first-order approximation and if reliabilities are homogeneous across moderator-based subgroups, power depends on the reliabilities α_y and α_x only through their product $\alpha_y\alpha_x$. Accordingly, power is affected by measurement error in X and Y in a symmetric fashion. For example, holding constant all other factors affecting power, the power of MMR will be approximately the same for the case (a) $\alpha_y = .80$ and $\alpha_x = .80$ (i.e., $\alpha_y\ \alpha_x = .64$) as for the case (b) $\alpha_y = .90$ and $\alpha_x = .71$ (i.e., $\alpha_y\ \alpha_x = .639$).

Also, Theorem 2 clarifies the effects of X and Y variance heterogeneity on power. More precisely, holding constant all other factors shown in the theorem and given a situation with two moderator-based subgroups, $\sigma_{x,1}$, $\sigma_{x,2}$, $\sigma_{y,1}$, and $\sigma_{y,2}$ affect power only through the ratios $\sigma_{x,1}/\sigma_{x,2}$ and $\sigma_{y,1}/\sigma_{y,2}$. This effect is illustrated later in the article in the section Relative Impact of Factors Affecting Power.

Finally, Theorem 2 also provides theoretical evidence that complements previous empirical results regarding the interactive effects of various factors that affect the power of MMR (Aguinis & Stone-Romero, 1997; Stone-Romero et al., 1994). The presence of interactive effects on power suggests that even if the value regarding one factor (e.g., total sample size) is favorable in terms of power (i.e., large), the existence of at least one other factor with an unfavorable value (e.g., poor reliability for X) may reduce power substantially. Thus, an unfavorable condition regarding any of the factors known to affect the power of MMR imposes a ceiling for the power of the F test. These interactive effects explain the typical concerns regarding the low power of MMR and the all-too-frequent failures to find so-called elusive moderating effects (Zedeck, 1971).

Accuracy of the Power Approximation

We conducted a Monte Carlo simulation to evaluate the power approximation shown in Theorem 2. The simulation was not intended to examine thoroughly a large number of conditions regarding the factors known to affect power (e.g., sample/population variances, reliabilities, sample size). Rather, our goal was to assess the accuracy of Theorem 2. Thus, we examined a set of diverse conditions considered to be typical in organization studies rather than a full factorial including all possible combinations of independent variable values included in the design. We implemented the simulation using a MATLAB 5 (http://www.mathworks.com/products/matlab) program.

Independent Variables

The simulation manipulated the following variables: (a) number of moderator-based subpopulations, (b) total sample size, (c) sample size across the moderator-based subgroups, (d) true score correlations between X and Y for the moderator-based subpopulations, (e) sampling restriction on X, (f) variances of X and Y, (g) reliabilities of X and Y, and (h) deviation from normality for X. Table 1 shows the 26 combinations of independent variable values, or cases, included in the simulation. Each of the 26 cases was sampled 10,000 times.

Commentary is needed regarding the manipulated variables. First, we chose to manipulate correlations as opposed to slopes. Although Appendix C shows that the moderating effect is defined as differences in slopes across the values of the moderator Z and the null hypothesis regarding the presence of the moderating effect is based on differences in slopes, the simulation varied correlations so as to make the results more easily interpretable. In other words, the standardized correlation metric might be more familiar and easier to interpret by most researchers than the unstandardized regression coefficient metric.

Second, regarding the X distribution, we sampled X scores from either a normal or a beta (1.5, 3.0) distribution. This beta distribution has a skewness coefficient of .51 and a kurtosis coefficient of –.46.

Third, regarding sampling restriction on X, we implemented the following four types of restriction on the normal and beta distributions: (a) left truncation (i.e., scores are sampled if they are above a specific cutoff point), (b) right truncation (i.e., scores are sampled only if they are below a specific cutoff point), (c) sparse left (i.e., low scores are sparsely sampled, and high scores are more densely sampled), and (d) sparse right (i.e., high scores are sparsely sampled, and low scores are more densely sampled). These four types of sampling restrictions are defined mathematically in Appendix C. Each form of sampling restriction was crossed with four values of a truncation-like parameter (i.e., T = .00, .25, .50, and .75), which is the proportion of

Table 1

Independent Variable Values Used in the Simulation Assessing the Accuracy of the Analytic Power Approximation

Case Number	k	n_j	ρ_j	T	$\sigma_{x,j}$	$\sigma_{y,j}$	$\alpha_{x,j}$	$\alpha_{y,j}$
1	2	50, 50	.10, .30	.75, .75	.40, .40	.40, .40	.70, .70	.70, .70
2	2	30, 70	.10, .50	.75, .50	.40, 1.2	.40, 2.0	.70, .90	.70, .90
3	2	10, 90	.10, .70	.75, .25	.40, 2.0	.40, 1.2	.90, .70	.90, .70
4	2	125, 125	.30, .10	.50, .75	1.2, .40	2.0, .40	.90, .90	.90, .90
5	2	75, 175	.30, .50	.50, .50	1.2, 1.2	1.2, 2.0	.70, .70	.70, .70
6	2	25, 225	.30, .70	.50, .25	1.2, 2.0	2.0, 1.2	.70, .90	.70, .90
7	2	200, 200	.50, .10	.25, .75	2.0, .40	1.2, .40	.90, .70	.90, .70
8	2	120, 280	.50, .30	.25, .50	2.0, 1.2	2.0, .40	.90, .90	.90, .90
9	2	40, 360	.50, .70	.25, .25	2.0, 2.0	2.0, 1.2	.70, .90	.70, .90
10	2	125, 125	.10, .50	0, 0	1.2, 1.2	2.0, 2.0	.90, .90	.90, .90
11	2	125, 125	.30, .30	0, 0	1.2, 1.2	2.0, 2.0	.90, .90	.90, .90
12	2	50, 50	.10, .30	.75, .75	.40, .40	.40, .40	.70, .70	.70, .70
13	2	75, 175	.30, .50	.50, .50	1.2, 1.2	1.2, 2.0	.70, .70	.70, .70
14	3	25, 25, 50	.10, .30, .50	.75, .75, .75	.40, .40, 1.2	.40, 1.2, 2.0	.70, .70, .70	.70, .70, .70
15	3	30, 30, 40	.10, .50, .70	.75, .75, .50	.40, 1.2, 2.0	.40, .40, 1.2	.70, .90, .70	.70, .90, .70
16	3	35, 35, 30	.30, .10, .50	.75, .50, .25	.40, 2.0, 2.0	1.2, 2.0, .40	.70, .90, .90	.70, .90, .90
17	3	25, 75, 150	.30, .30, .70	.50, .75, .25	.40, 1.2, 1.2	2.0, 1.2, .40	.90, .70, .70	.90, .70, .70
18	3	50, 75, 125	.50, .10, .70	.50, .50, .25	1.2, .40, 2.0	.40, 1.2, 1.2	.90, .90, .90	.90, .90, .90
19	3	75, 75, 100	.50, .10, .30	.50, .25, .25	2.0, .40, 1.2	1.2, 1.2, .40	.90, .90, .70	.90, .90, .70
20	3	50, 75, 275	.70, .30, .50	.25, .75, .50	1.2, 2.0, .40	2.0, 2.0, .40	.70, .70, .90	.70, .70, .90
21	3	75, 75, 250	.70, .50, .10	.25, .50, .25	2.0, 2.0, .40	1.2, .40, 2.0	.70, .90, .90	.70, .90, .90
22	3	90, 90, 220	.10, .70, .50	.25, .50, .75	1.2, 1.2, .40	2.0, 1.2, .40	.90, .70, .70	.90, .70, .70
23	3	75, 75, 100	.10, .30, .50	0, 0, 0	1.2, 1.2, 1.2	2.0, 2.0, 2.0	.90, .90, .90	.90, .90, .90
24	3	75, 75, 100	.30, .30, .30	0, 0, 0	1.2, 1.2, 1.2	2.0, 2.0, 2.0	.90, .90, .90	.90, .90, .90
25	3	25, 25, 50	.10, .30, .50	.75, .75, .75	.40, .40, 1.2	.40, 1.2, 2.0	.70, .70, .70	.70, .70, .70
26	3	75, 75, 100	.50, .10, .30	.50, .25, .25	2.0, .40, 1.2	1.2, 1.2, .40	.90, .90, .70	.90, .90, .70

Note. k = number of moderator-based subpopulations; n_j = sample size in each moderator-based subgroup (i.e., total sample size = Σn_j); ρ_j = true score correlations between X and Y for each moderator-based subpopulation; T = truncation proportion on X for each moderator-based subgroup (i.e., proportion of scores that cannot be included in the sample); $\sigma_{x,j}$ = true score standard deviation for X; and $\sigma_{y,j}$ = true score standard deviation for Y for each moderator-based subpopulation; $\alpha_{y,j}$ = reliability for Y; $\alpha_{x,j}$ = reliability for X for each moderator-based subpopulation.

the population that cannot be sampled due to restriction. Thus, the manipulation of the sampling mechanism for X led to various degrees of skewness, kurtosis, and X variance multiplying factors (i.e., expected sample variance of X/population variance of X). Tables showing the values for skewness, kurtosis, and variance multiplying factors associated with each combination of sampling mechanism and T for the normal and the beta (i.e., nonnormal) distributions are available from the authors on request.

Independent variable values. Commentary is needed regarding the choice for the various independent variable values used in the simulation and shown in Table 1. Once again, the goal of this simulation was to assess the accuracy of the analytically derived solution to approximate power. Thus, we were not attempting to include every situation encountered by researchers or to generate extensive power tables. Investigators can generate their own custom power values using the computer program that implements the analytic solution (see the Computer Program section below).

We chose the independent variable values according to the following rationale. First, regarding the number of moderator-based subpopulations, we chose the values of $k = 2$ and 3. This was based on a literature review by Aguinis et al. (2000) concluding that virtually all of the 616 MMR tests of categorical moderator variables reported in major organization studies journals over the past 30 years included moderators with two or three levels.[3]

Second, regarding total sample size, we chose values ranging from 100 to 400. This choice was based on Jaccard and Wan's (1995) review of American Psychological Association (APA) journals indicating that the median sample size is 175.

Third, regarding sample size across the moderator-based subgroups, we divided the total sample size into two or three subgroups based on proportions ranging from .10 to .50. The rationale was that researchers might find situations in which the proportion of scores in one subgroup is as low as .10 of the total sample. On the other hand, there might be situations in which the sample size is fairly equal across moderator-based subgroups (i.e., a proportion of .50). Thus, a range of .10 to .50 covers most typical research situations.

Fourth, regarding the magnitude of the moderating effect, we chose various combinations of correlations ranging from .10 to .70 across moderator-based subpopulations. The rationale was that these are the values typically observed, although .70 is less usual, in organization studies and other social science fields (e.g., political science, psychology). For example, Cohen (1988) defined effects of .10, .30, and .50 as small, medium, and large, respectively. We also included negative correlations within this range (i.e., Cases 12, 13, 25, and 26 in Table 1). Note, however, that the corresponding population correlations based on observable scores may differ from population correlations based on true scores because correlations between observable scores are affected by measurement error.

Fifth, regarding the truncation proportion on X (i.e., T), we chose values of .00, .25, .50, and .75. The rationale was that these values cover a range from $T = 0$, in which all population scores can be included in the samples, to $T = .75$, in which only 25% of the population scores can be included in the sample. Implementing truncation on the normal and beta distributions led to variance multiplying factors (i.e., expected sample variance of X/population variance of X) ranging from 1.00 to .17. Thus, the simulation included situations ranging from no differences between sample and population variances to situations in which there is very severe variance restriction.

Sixth, the values for the standard deviation of *X* and *Y* were chosen after conducting a cursory review of several APA journals. Our review suggested that published studies using Likert-type scales with five and seven anchors yield standard deviations in the .40 to 2.0 range. Thus, this guided our choice for the standard deviation values.

Finally, regarding reliabilities for *X* and *Y*, we chose the values .70 and .90. Similar to Jaccard and Wan (1995), the rationale was that .70 is considered to be the lower bound of acceptable reliability levels (Nunnally & Bernstein, 1994), and .90 is a desirable level that serves as a de facto upper-bound level in many areas of organization studies and other social sciences.

Dependent Variable

The dependent variable, power, was the proportion of times out of each set of 10,000 trials that the null hypothesis of no moderating effect was rejected. We computed this proportion for each of the 26 cases shown in Table 1 crossed with each of the four sampling mechanisms (i.e., left and right truncation and sparse left and right) and the two types of underlying *X* distributions (i.e., normal and beta) described above and defined mathematically in Appendix C. We also computed an approximate (i.e., analytically derived) power value using Theorem 2 as implemented by the computer program MMRPOWER (see Computer Program section below).

Results and Discussion

Tables 2 and 3 show (a) the proportion of times the null hypothesis was empirically rejected, and (b) the power approximation yielded by Theorem 2 for each of the 26 cases shown in Table 1 crossed with the four sampling mechanisms. Table 2 shows results pertaining to an underlying normal *X* distribution, and Table 3 shows results for an underlying beta (i.e., nonnormal) *X* distribution. Tables 2 and 3 indicate that the empirical and theoretical proportions are virtually identical for every condition. The difference between the analytically derived power approximation and the empirically derived rejection rates was in no case greater than $|.019|$. The mean absolute deviation between the simulation and theory-based power values is .0036 for Table 2 and .0037 for Table 3. Given the statistical power metric (i.e., ranging from 0 to 1.00), the differences between the empirical and analytic values are negligible.

Comparison with previous empirically based algorithms to estimate power. We next compared analytically derived power estimates with those generated by the Aguinis and Pierce (1998b) program. Recall that, in contrast to the present analytic solution, the Aguinis and Pierce program suffers from certain limitations (e.g., does not allow for the consideration of measurement error, only allows for power estimation for binary moderator variables). Because of these limitations, the power estimates generated using the Aguinis and Pierce program assume that (a) reliabilities are 1.0 for *X* and *Y* across the two moderator-based subgroups, (b) the variance and truncation for *X* are identical across the two moderator-based subgroups, (c) the variance for *Y* is identical across the two moderator-based subgroups, and (d) *X* and *Y* scores follow a bivariate normal distribution. Because the Aguinis and Pierce program does not allow for the specification of as many variables as the present analytic approximation, in the comparison we could only vary (a) total sample size, (b) sample sizes across moderator-

Table 2
Comparison of Empirical and Analytic (i.e., using Theorem 2)
Power Values (underlying normal distribution for *X*)

Case Number	*Left Truncation*	*Right Truncation*	*Theoretical Approximation*	*Sparse Left*	*Sparse Right*	*Theoretical Approximation*
1	.061	.060	.063	.077	.071	.076
2	.000	.001	.000	.002	.002	.001
3	.000	.000	.000	.000	.000	.000
4	.002	.002	.001	.008	.007	.007
5	.169	.166	.161	.313	.311	.311
6	.307	.309	.317	.313	.305	.317
7	.021	.021	.019	.070	.066	.065
8	.898	.896	.898	.984	.983	.985
9	.295	.295	.288	.278	.295	.291
10	.840	.840	.842	.846	.841	.842
11	.048	.048	.050	.046	.048	.050
12	.058	.063	.063	.080	.073	.076
13	.165	.162	.161	.308	.319	.311
14	.007	.005	.005	.007	.007	.005
15	.010	.008	.007	.021	.019	.017
16	.070	.069	.066	.082	.081	.077
17	.423	.420	.411	.498	.500	.485
18	.121	.122	.113	.220	.214	.204
19	.303	.300	.301	.473	.468	.467
20	.602	.597	.594	.782	.773	.772
21	.101	.100	.102	.181	.189	.182
22	.312	.311	.308	.448	.445	.446
23	.586	.586	.583	.585	.591	.583
24	.052	.050	.050	.053	.050	.050
25	.006	.006	.005	.005	.006	.005
26	.299	.301	.301	.472	.464	.467

Note. Cases are defined in Table 1. Left truncation and right truncation = empirical power values (i.e., proportion of times the null hypothesis of no moderating effect was empirically rejected) implementing left truncation and right truncation sampling mechanisms defined in Appendix C; theoretical approximation = power approximate using Theorem 2 as implemented by the program MMRPOWER; sparse left and sparse right = empirical power values implementing sparse left and sparse right sampling restriction mechanisms defined in Appendix C.

based subgroups, (c) correlations between *X* and *Y* for each moderator-based subpopulation, and (d) truncation proportion. Table 4 shows results of this comparison.

Table 4 shows that, as was shown in Tables 2 and 3, the analytic approximation is virtually identical to the simulation-based results. In addition, Table 4 shows that, under restrictive assumptions such as lack of measurement error across subpopulations, the Aguinis and Pierce (1998b) program power estimate also was close to the simulation-based value in several conditions. For instance, for Cases 1 through 3, the difference between the simulation and the Aguinis and Pierce power values ranged from –.049 to .021. Note, however, that these cases include values that fall within the range of the simulation study on which the Aguinis and Pierce program was based (i.e., Aguinis & Stone-Romero, 1997). Thus, it was expected that the Aguinis and Pierce power estimate would be fairly accurate for these situations.

Table 3
Comparison of Empirical and Analytic (i.e., using Theorem 2) Power Values (underlying beta [1.5, 3.0] distribution for *X*)

Case Number	*Left Truncation*		*Right Truncation*		*Sparse Left*		*Sparse Right*	
	Simulation	*Theory*	*Simulation*	*Theory*	*Simulation*	*Theory*	*Simulation*	*Theory*
1	.064	.065	.051	.053	.085	.084	.055	.062
2	.000	.000	.000	.000	.004	.002	.000	.000
3	.000	.000	.000	.000	.000	.000	.000	.000
4	.001	.001	.000	.000	.013	.012	.001	.000
5	.225	.217	.080	.077	.400	.391	.228	.228
6	.308	.318	.308	.319	.308	.316	.307	.318
7	.025	.023	.002	.002	.105	.101	.016	.016
8	.958	.958	.695	.690	.991	.993	.969	.969
9	.297	.290	.295	.287	.300	.291	.289	.290
10	.841	.842	.841	.842	.841	.842	.847	.842
11	.045	.050	.050	.050	.050	.050	.049	.050
12	.064	.065	.053	.053	.085	.084	.058	.062
13	.219	.217	.077	.077	.395	.391	.238	.228
14	.007	.005	.005	.004	.008	.006	.005	.005
15	.008	.006	.001	.001	.030	.024	.004	.004
16	.072	.068	.069	.070	.086	.083	.066	.066
17	.456	.444	.381	.372	.535	.520	.466	.453
18	.155	.143	.083	.078	.289	.270	.156	.149
19	.382	.381	.211	.217	.538	.535	.406	.404
20	.675	.667	.384	.378	.826	.822	.661	.657
21	.134	.133	.064	.067	.234	.230	.145	.141
22	.380	.371	.198	.196	.503	.498	.386	.386
23	.590	.583	.590	.583	.590	.583	.590	.583
24	.048	.050	.048	.050	.048	.050	.048	.050
25	.006	.005	.006	.004	.007	.006	.007	.005
26	.392	.381	.220	.217	.540	.535	.401	.404

Note. Cases are defined in Table 1. Left and right truncation = sampling mechanisms based on truncation defined in Appendix C; sparse left and sparse right = sparse left and sparse right sampling restriction mechanisms defined in Appendix C; simulation = empirical power values (i.e., proportion of times the null hypothesis of no moderating effect was empirically rejected); theory = power approximate using Theorem 2 as implemented by the program MMRPOWER.

Alternatively, Cases 4 and 5 show situations including values falling outside of the range of the simulation work on which the Aguinis and Pierce (1998b) program was based (i.e., for Case 4, there is a negative correlation, and for Case 5, the correlations are greater than .80). In these situations, the difference between the simulation and the Aguinis and Pierce power values was greater. In Case 4, the Aguinis and Pierce program yielded a negatively biased value (i.e., –.170), and in Case 5, the program yielded a positively biased value (i.e., .789). These discrepancies illustrate one of the weaknesses described in the introduction section regarding the available computer programs to estimate power: Because they are based on Monte Carlo data, their accuracy is confined only to values similar to those used in the simulation on which the programs are based. The Aguinis and Pierce program is based on the Aguinis and Stone-Romero (1997) simulation, which did not include correlations lower than .20 or

Table 4
Comparison of Power Values Based on the Present Analytic Approximation and the Aguinis and Pierce (A&P) (1998b) Computer Program

				Power		
Case Number	n_j	ρ_j	T	*Analytic Approximation*	*A&P*	*Simulation*
1	150, 150	.20, .80	.60	.941	.961	.940
2	50, 100	.20, .40	.00	.228	.184	.233
3	80, 120	.10, .40	.90	.147	.158	.145
4	75, 100	-.40, .40	.20	.989	.816	.986
5	150, 150	.95, .99	.60	.124	.910	.121

Note. n_j = sample size in each moderator-based subgroup (i.e., total sample size = $n_1 + n_2$); ρ_j = correlations between *X* and *Y* for each moderator-based subpopulation; *T* = truncation proportion on *X* for each moderator-based subgroup (i.e., proportion of scores that cannot be included in the sample); analytic approximation = power approximate using Theorem 2 as implemented by the program MMRPOWER; A&P = power value generated using the Aguinis and Pierce (1998b) computer program; simulation = empirical power values (i.e., proportion of times the null hypothesis of no moderating effect was empirically rejected). The comparison in this table holds *k* constant at 2 (i.e., the moderator variable has two levels), reliabilities of *X* and *Y* constant at 1.00, and variance of *Y* constant at 1.00 (i.e., the Aguinis & Pierce, 1998b, program does not allow for variations in *k*, reliabilities, and *Y* variance).

higher than .80. The foregoing example illustrates the superiority of an analytic solution that can be generalized to any parameter value range.

It is also of interest to approximate power for the cases shown in Table 4 once some of the Aguinis and Pierce (1998b) assumptions are relaxed. For instance, Case 1 yielded a satisfactory power value (.961 for the Aguinis & Pierce program and .940 for the simulation; a difference of just .021). Recall that these power values assume perfect reliability for both *X* and *Y* scores for the two moderator-based subgroups. Relaxing this assumption, we generated a power value via simulation assuming a more realistic scenario in organization studies in which reliabilities for *X* and *Y* are .80 for each of the two moderator-based subgroups. The resulting simulation power value was .734, below the recommended value of .80 (Cohen, 1988). Thus, just by relaxing one of the assumptions (i.e., lack of measurement error), the difference between the Aguinis and Pierce and the simulation power values increased from .021 to .229. In contrast, the power value generated using the analytic approximation was .736, a difference of just .002 from the simulation result. This example shows a second limitation mentioned in the introduction section pertaining to previous research attempting to estimate the power of MMR based on empirical work. Once again, computer programs based on empirically derived algorithms to estimate power (e.g., Aguinis et al., 1994; Aguinis & Pierce, 1998b) cannot go beyond the empirical work on which they are based; for instance, if the original research assumed lack of measurement error, so do the corresponding computer programs. In contrast, the present analytic solution is based on theory and goes beyond previous empirical work. Consequently, it allows power values to be approximated based on more realistic conditions (e.g., measurement error for *X* and *Y* across moderator-based subgroups, nonnormal *X* distributions, range restriction beyond simple truncation on *X*). In short, researchers using the present analytic result have a more generalizable and accurate approximation of power and, consequently, are likely to make more informed conclusions regarding the operation of moderating effects of categorical variables.

Computer Program

As noted above, we developed a computer program that performs all the necessary computations required by the theoretical approximation shown in Theorem 2. This program (MMRPOWER) is available at http://www.math.montana.edu/~rjboik/power.html. We chose to make MMRPOWER available on the Internet so as to reach the largest possible number of users regardless of operating system platform (e.g., Windows 95/98/NT/2000, Macintosh, OS2). The program was written in FORTRAN 77 and consists of a main program that calls several FORTRAN subprograms and functions that were originally published in *Applied Statistics* (a list of the algorithms used is available from the authors on request).

The first screen of the program prompts the user to provide information regarding (a) number of moderator-based subgroups (the maximum number handled by the program is 20), (b) desired significance level or test size (i.e., preset Type I error rate), (c) desired test (i.e., overall test of equality of slopes across moderator-based subgroups or tests of specific contrasts of slopes), and (d) sampling restrictions (i.e., none, sampling from truncated X normal distributions, or sampling from nonnormal X distributions). The program also prompts the user to provide information regarding whether the input format includes correlations based on true scores, correlations based on observable scores, slopes based on true scores, or slopes based on observable scores.

The second screen of the program prompts the user to input the necessary information to compute power. In addition to sample size and reliabilities for each of the moderator-based subgroups, the necessary input varies depending on the choices made on the first screen. That is, the user is prompted for correlations or slopes based on true or observable scores for each moderator-based subgroup. In addition, the user is prompted for the truncation proportion for X (i.e., T) if truncated normal distributions were noted on the first screen, no information if no sampling restrictions were noted on the first screen, and variance multiplying factors (i.e., expected sample variance/population variance) if nonnormal X distributions were noted on the first screen. Finally, if tests of specific contrasts were requested on the first screen, the second screen prompts the user for contrast coefficients (see Boik, 1979, 1993, for a discussion of main and interaction effect contrasts).

Needed Input

All the information required by the program is typically available to researchers (except for truncation and variance multiplying factor information). If the true score options are chosen, sample-based statistics must be used to estimate parameters. The program will yield an accurate power value as long as the estimates are accurate.

Information regarding truncation and the variance multiplying factor (when the sampling mechanism is other than truncation) may not be available. If this information is not available, we suggest that researchers use an estimate based on relevant literature. If a literature-based estimate is not available, researchers can input a best-case scenario (i.e., use a truncation proportion of 0.00 and assume no truncation or a variance multiplying factor of 1.00 and assume no variance difference between the sample and the population) and a worse-case scenario (i.e., use a severe truncation proportion such as 0.75 or a severe variance multiplying factor of 0.25). Given the absence of information on truncation and the variance multiplying factor, researchers will know that the power of their MMR test lies somewhere between the best-case and worse-case situations.

Relative Impact of Factors Affecting Power

As noted in the introduction section, several empirical studies have examined the impact of each of the many factors known to affect the power of MMR. Typically, these Monte Carlo simulations have investigated only the concurrent impact of two or three factors. For instance, Stone-Romero et al. (1994) only manipulated three design-related factors: (a) sample size in each of two moderator variable-based subgroups, (b) total sample size, and (c) magnitude of moderating effect. Likewise, Stone-Romero and Anderson (1994) also varied only three factors: (a) total sample size, (b) unreliability of predictor variable scores, and (c) magnitude of moderating effect. Another contribution of the present analytic approximation is that, in contrast to previous empirical work, power values can be easily and concurrently generated for a diverse set of variables and values. Consequently, we can now gain a better understanding of, when other variables are held constant, what is the relative impact of improving one factor (e.g., increasing total sample size) as compared to improving conditions regarding another factor (e.g., minimizing restriction on *X*).

We present a total of 34 illustrative cases in Tables 5 and 6 to show the relative impact of the various factors affecting the power of MMR. In Table 5, we illustrate the effects of total sample size, sample size across moderator-based subgroups, differences in correlations across moderator-based subgroups, truncation (for the *X* normal distribution case) and variance multiplying factor (for the arbitrary *X* distribution case), and reliabilities on *X* and *Y*.

Regarding Table 5, to make the various comparisons easier to understand, we present a situation in which there are two moderator-based subgroups (e.g., gender). Also, truncation, variance multiplying factor, and *X* and *Y* variances ($\sigma_x = \sigma_y = 1.0$) are not varied across the subpopulations (effects of *X* and *Y* variance heterogeneity are shown in Table 6). In addition, because the effects of reliability of *X* and *Y* are symmetrical when reliabilities are homogeneous across moderator-based subgroups, we varied the product $\alpha_y\alpha_x$ and not the reliability for each variable. Finally, we also held $\alpha_y\alpha_x$ constant across the two subgroups.

Case 1 in Table 5 is what could be labeled an “optimal” situation for detecting a moderating effect. Total sample size is 400 (much larger than the median of 175 reported by Jaccard & Wan, 1995), sample size is equal across the subgroups, there is a .50 difference between the $X - Y$ correlations across the two subgroups, there is no restriction on *X*, *X* and *Y* variances are equal across the two subgroups, and the product of the reliability terms is .81 (i.e., $\alpha_y = .90$ and $\alpha_x = .90$). Not surprisingly, given these optimal conditions, the power of the MMR test is .998.

Case 6 is what could be labeled an “average” or more typical situation for detecting a moderating effect. $N = 175$, the sample sizes ratio across the subgroups is .67 (i.e., they are dissimilar but not drastically different), there is a .30 difference between $X - Y$ correlations across the two subgroups, there is some restriction (i.e., scores can be sampled from 75% of the range of population scores), variances of *X* and *Y* are identi-

Table 5
Effects of Sample Size, Effect Size, Truncation or Variance Multiplying Factor, and Reliability on *X* and *Y* on Power of Moderated Multiple Regression for Illustrative Research Situations

Case Number	n_j	ρ_j	T	δ	$x_y x_y$	*Power*
Underlying normal *X* distribution						
1	200, 200	.10, .60	.00		.81	.998
2	105, 70	.10, .60	.00		.81	.867
3	200, 200	.10, .40	.00		.81	.791
4	200, 200	.10, .60	.25		.81	.942
5	200, 200	.10, .60	.00		.64	.988
6	105, 70	.10, .40	.25		.64	.203
7	200, 200	.10, .40	.25		.64	.431
8	105, 70	.10, .60	.25		.64	.494
9	105, 70	.10, .40	.00		.64	.341
10	105, 70	.10, .40	.25		.81	.248
Underlying arbitrary *X* distribution						
11	200, 200	.10, .60		1.00	.81	.998
12	105, 70	.10, .60		1.00	.81	.867
13	200, 200	.10, .40		1.00	.81	.791
14	200, 200	.10, .60		0.75	.81	.987
15	200, 200	.10, .60		1.00	.64	.988
16	105, 70	.10, .40		0.75	.64	.268
17	200, 200	.10, .40		0.75	.64	.562
18	105, 70	.10, .60		0.75	.64	.638
19	105, 70	.10, .40		1.00	.64	.341
20	105, 70	.10, .40		0.75	.81	.329

Note. n_j = sample size in each moderator-based subgroup (i.e., total sample size = $n_1 + n_2$); ρ_j = correlations between *X* and *Y* for each moderator-based subpopulation; *T* = truncation proportion on *X* for each moderator-based subgroup (i.e., proportion of scores that cannot be included in the sample) for normal distribution; δ = variance multiplying factor (i.e., expected sample variance of *X*/population variance of *X*) for arbitrary distribution; α_y = reliability for *Y* and α_x = reliability for *X*; power = power approximate using Theorem 2 as implemented by the program MMRPOWER. Power values were approximated using a preset nominal Type I error = .05. It is assumed that σ_y (true score standard deviation for *Y*) = σ_x (true score standard deviation for *X*) = 1.00 for each moderator-based subpopulation.

cal across subpopulations, and the product of the reliability terms is .64 (i.e., $\alpha_y = \alpha_x$ = .80 across subgroups). The resulting power for what can be considered a more typical research situation is .203. Given this low value, it is not surprising that moderating effects have been labeled "elusive" (Zedeck, 1971). If Case 6 indeed represents a situation faced by the majority of organization science researchers using MMR to test hypotheses regarding moderating effects of categorical variables, the chances of rejecting a null hypothesis would be greater if a coin toss were used instead of MMR.

Cases 2 through 5 in Table 5 are follow-ups to the optimal Case 1. We changed the value for each of the factors from optimal to average without altering the optimal conditions regarding all other factors. For instance, Case 2 shows that decreasing the total sample size and not having identical sizes across the subgroups decreases power from .998 to .867, holding all other factors at their optimal value. Case 3 shows that decreasing the difference in correlations from .10 in Subgroup 1 and .60 in Subgroup 2 to .10

Table 6
Effects of *X* and *Y* Variance Heterogeneity on Power of Moderated Multiple Regression for Illustrative Research Situations

Case Number	n_j	ρ_j	$x_y x_y$	$\sigma_{y,1}/\sigma_{y,2}$	$\sigma_{x,1}/\sigma_{x,2}$	σ^2_{ej}	β_j	*Power*
1	200, 200	.10, .60	.81	2/2 = 1	2/2 = 1	4.41, 3.15	.09, .54	.998
2	200, 200	.10, .60	.81	2/2 = 1	2/4 = .5	4.41, 3.15	.09, .27	.678
3	200, 200	.10, .60	.81	2/4 = .5	2/2 = 1	4.41, 12.59	.09, 1.08	1.000
4	200, 200	.10, .60	.81	2/4 = .5	2/4 = .5	4.41, 12.59	.09, .54	.994
5	200, 200	.10, .60	.81	2/2 = 1	2/1 = 2	4.41, 3.15	.09, 1.08	1.000
6	200, 200	.10, .60	.81	2/1 = 2	2/2 = 1	4.41, .79	.09, .27	.645
7	200, 200	.10, .60	.81	2/1 = 2	2/1 = 2	4.41, .79	.09, .54	.988
8	105, 70	.10, .40	.64	2/2 = 1	2/2 = 1	4.97, 4.49	.08, .32	.341
9	105, 70	.10, .40	.64	2/2 = 1	2/4 = .5	4.97, 4.49	.08, .16	.109
10	105, 70	.10, .40	.64	2/4 = .5	2/2 = 1	4.97, 17.95	.08, .64	.681
11	105, 70	.10, .40	.64	2/4 = .5	2/4 = .5	4.97, 17.95	.08, .32	.287
12	105, 70	.10, .40	.64	2/2 = 1	2/1 = 2	4.97, 4.49	.08, .64	.582
13	105, 70	.10, .40	.64	2/1 = 2	2/2 = 1	4.97, 1.12	.08, .16	.065
14	105, 70	.10, .40	.64	2/1 = 2	2/1 = 2	4.97, 1.12	.08, .32	.108

Note. n_j = sample size in each moderator-based subgroup (i.e., total sample size = $n_1 + n_2$); ρ_j = correlations between *X* and *Y* for each moderator-based subpopulation; α_y = reliability for *Y* and α_x = reliability for *X*; $\sigma_{y,1}$ and $\sigma_{y,2}$ = true score standard deviation for *Y* for moderator-based Subpopulations 1 and 2, respectively; $\sigma_{x,1}$ and $\sigma_{x,2}$ = true score standard deviation for *X* for moderator-based Subpopulations 1 and 2, respectively; σ^2_{ej} = error variance for each moderator-based subpopulation computed using Equation C1; β_j = slope of *Y* on *X* for each moderator-based subpopulation computed using Equation A6; power = power approximate using Theorem 2 as implemented by the program MMRPOWER. Power values were approximated using a preset nominal Type I error = .05. It is assumed that $T = 0$ (T = proportion of *X* scores that cannot be included in the sample). Also, it is assumed that *X* has an underlying normal distribution (results based on an arbitrary *X* distribution were identical because *T* was set at 0, that is, no restriction on *X*).

and .40, respectively, decreases power from .998 to .791. Case 4 shows the effect of adding some restriction (i.e., scores can only be sampled from 75% of the population score range), and Case 5 shows the effect of adding more measurement error. Relatively speaking, and given the values shown in Cases 2 through 5 in Table 5, the following is a rank ordering of factors that affect the power of MMR in order of importance: (a) moderating effect magnitude, (b) total sample size, (c) sampling restriction on *X*, and (d) measurement error. Of course, this rank ordering may change if noticeably different values are chosen as optimal and average for each of the factors.

Table 5 also shows various situations (Cases 7-10) in which there is an average condition regarding all factors, and there is improvement in one factor at a time, from average to optimal, without altering the other factors. As expected, the pattern of power improvement is similar to the pattern observed for Cases 2 through 5 regarding power decrements. Given that conditions are average for all other factors, increasing effect size and sample size, reducing sampling restriction on *X*, and improving reliability increase power, in this respective order.

The bottom half of Table 5 shows a similar pattern of results when normality is not assumed for the *X* distribution in assessing the effects of differences between expected sample variance of *X* and population variance of *X*. Effect size and sample size are the two most important factors affecting power. However, the effects of reliability and restriction are similar in magnitude. That is, improving one or the other factor yields similar gains in power.

An additional conclusion can be drawn from Table 5. Aguinis and Stone-Romero (1997) concluded that the power function is nonlinear. Table 5 (and additional tables and graphs available from the authors on request) shows further that the pattern of nonlinearity depends on the factor that is varied as well as the magnitude of the change in the values of the factors manipulated.

Table 6 shows the effects on power of heterogeneity of X and Y variance across moderator-based subpopulations. As in Table 5, to make the various comparisons easier to understand, we present a situation in which there are two moderator-based subgroups (e.g., gender). Also, we assume no restriction on X (i.e., $T = 0$ for the normal case and variance multiplying factor = 1.00 for the arbitrary distribution). In addition, because the effects of reliability of X and Y are symmetrical when X and Y reliabilities are homogeneous across moderator-based subgroups, we varied the product $\alpha_y\alpha_x$ and not the reliability for each variable. Finally, we also held $\alpha_y\alpha_x$ constant across the two subgroups.

Case 1 in Table 6 shows what could be labeled an optimal case for detecting a moderating effect. Sample size is large (i.e., 400) and equal across moderator-based subgroups, there is a .50 difference in the $X - Y$ correlations between the subpopulations, reliabilities for X and Y are .90 for each of the subgroups, and variances for X and Y are identical across the subpopulations. More important, because it is the variance ratios that have an impact on power (holding all other variables constant) rather than the absolute variance values, the ratios are $\sigma_{y,1}/\sigma_{y,2} = \sigma_{x,1}/\sigma_{x,2} = 1.00$. Finally, the error variances are homogeneous, and their ratio is 1.40 (DeShon & Alexander, 1996, concluded that error variances are homogeneous if the ratio of the largest to the smallest variance is less than 1.50; see Aguinis & Pierce, 1998a, for a review of effects of error variance heterogeneity on the power of MMR). Not surprisingly, given these optimal, and perhaps rare, conditions, the power of MMR is .998.

In Cases 2 through 4 in Table 6, we systematically varied the $\sigma_{y,1}/\sigma_{y,2}$ and $\sigma_{x,1}/\sigma_{x,2}$ ratios to include values of 0.5, 1.0, and 2.0 and held all other variables identical to the optimal Case 1. Thus, in Case 2 the ratios are $\sigma_{y,1}/\sigma_{y,2} = 1.0$ and $\sigma_{x,1}/\sigma_{x,2} = 0.5$; in Case 3 the ratios are $\sigma_{y,1}/\sigma_{y,2} = 0.5$ and $\sigma_{x,1}/\sigma_{x,2} = 1.0$; and in Case 4 the ratios are $\sigma_{y,1}/\sigma_{y,2} = 0.5$ and $\sigma_{x,1}/\sigma_{x,2} = 0.5$. Note that as X and Y variances change, so do error variances and slopes. In Cases 5 through 7, we varied the X and Y variance ratios in an identical fashion as compared to Cases 2 through 4 but replaced each 0.5 ratio with a 2.0 ratio. Taken together, Cases 1 through 7 show that even when other conditions are optimal for the detection of a moderating effect with respect to reliability, sample size, and differences in correlations, heterogeneity of variance across subgroups can be detrimental to power. Some patterns of variance heterogeneity lead to a decrease in differences in slopes and, consequently, a decrease in power. For instance, Case 6 shows that despite the fact that X variances are identical across subpopulations, a 2:1 ratio regarding Y variance leads to a power of .645 as compared to a power of .998 when the Y variances are identical. This effect is particularly noteworthy given that a 2:1 ratio of variances might not be perceived by most researchers as posing a serious threat to MMR-based conclusions. Nevertheless, the decrease in power is substantial.

Case 8 in Table 6 shows what could be labeled an average situation for detecting a moderating effect. Total sample size is 175, sample sizes are moderately different across subgroups, $X - Y$ correlations are .10 and .40 across the two subpopulations, and reliabilities for *X* and *Y* are assumed to be .80 for each subgroup. Note, however, that in Case 8 the *X* and *Y* variances take on the optimal value such that they are identical across the subpopulations. Moreover, the ratio of error variances is also optimal. The ratio is 1.11, which is less than the 1.50 rule of thumb recommended to decide whether error variances are heterogeneous (DeShon & Alexander, 1996). This average situation yielded a power value of .341 (note that power is higher than the .203 value for the average situation in Case 6 in Table 5 because in Table 6 it is assumed that there is no restriction on *X*). Once again, as illustrated in Table 5, this shows that numerous, if not most, hypothesis tests using MMR are prone to Type II errors (i.e., incorrectly failing to reject a false null hypothesis).

Next, in Cases 9 though 14 in Table 6, we replicated the pattern of *X* and *Y* variance ratios used for Cases 2 through 7 to investigate the impact of the variance ratios on the power of this average situation. For instance, Case 10 shows that increasing *Y* variance heterogeneity can increase power, even when there is error variance heterogeneity. Taken together, power values shown in Cases 9 through 14 reinforce a conclusion reached by DeShon and Alexander's (1996) empirical study: Error variance can lead to increases or decreases in power because Type I error rates are not fixed at their preset nominal level (.05 in this case). Most notably, Cases 9 through 14 illustrate the impact of variance heterogeneity on slope differences and, in turn, on power. For instance, Cases 10 and 13 provide a clear example. Both cases are identical. The only difference is that for Case 10, $\sigma_{y,1}/\sigma_{y,2} = .5$, whereas for Case 13, $\sigma_{y,1}/\sigma_{y,2} = 2.0$. This difference in *Y* variances led to power = .681 for Case 10 and power = .065 for Case 13. In short, what might be considered small differences in *Y* variances across subpopulations can have dramatic effects on power.

There is an important point that should be emphasized regarding the effects of error variance heterogeneity in MMR. Results based on MMR cannot be trusted in the presence of error variance heterogeneity because Monte Carlo studies as well as the analytic approximation have demonstrated that MMR's *F* test does not control Type I error at the nominal rate when error variances are heterogeneous (Aguinis et al., 1999; Aguinis & Pierce, 1998a). Depending on population and sample characteristics, violating the homogeneity of error variance assumption (a) increases or decreases Type I error rates (e.g., a researcher mistakenly concludes that a moderating effect exists), and (b) increases or decreases Type II error rates (i.e., a researcher mistakenly concludes that a moderating effect does not exist). For instance, regarding Type I error, Dretzke, Levin, and Serlin (1982) showed that error rates are artificially inflated when sample sizes are unequal across subgroups, and this is most noticeable when the smaller subgroup sample size is paired with the larger error variance. Regarding Type II error, Alexander and DeShon (1994) found that error rates increase (i.e., statistical power is lowered) when the subgroup with the larger sample size is associated with the larger error variance (see Aguinis & Pierce, 1998a, for a review). Because of the Type I and Type II error rate problems due to heterogeneity of error variance, Aguinis et al. (1999) developed the computer program ALTMMR (available at http://members.aol.com/imsap/altmmr.html) to allow MMR users to (a) assess whether error variances are homogeneous, and (b) compute alternatives to MMR's *F* test if they are heterogeneous. Thus, we suggest that, after data have been collected, researchers use ALTMMR to check whether the homogeneity of error variance assumption is satisfied before proceeding to conduct an MMR analysis.

The present analytic solution to computing the power of MMR does not solve the problem that the *F* statistic does not have an *F* distribution when the homogeneity of error variance assumption is violated. MMR's results about whether H_0 should be rejected cannot be trusted when error variances are heterogeneous, and this is a problem associated with MMR and not with the present analytic solution. In fact, results show that power values generated using the analytic solution are virtually identical to those generated empirically, even when error variances are heterogeneous. The algorithm and program yield accurate power rates, but when the assumption is violated, these power rates correspond to a test whose size differs from the nominal α. In short, when error variances are heterogeneous, MMR results cannot be trusted, and MMR should not be used (Aguinis et al., 1999).

An additional contribution of the present analytic solution is that MMRPOWER can also be used to approximate the actual α, as well as learn about the deviation of the actual α from the nominal α given a violation of the homogeneity of error variance assumption. To do this, users would first select identical slopes for the moderator-based subgroups and then input various combinations of values for the factors known to cause a violation of the assumption (e.g., sample sizes, standard deviations).

MMRPOWER issues a caution statement in situations when H_0 is true and the actual α differs from the preset nominal α, and MMRPOWER's output includes each group's error standard deviation (i.e., "W-Group Error SD"). The same caution statement is issued for any contrasts if the contrast H_0 is true and error variances are heterogeneous. In such situations, MMRPOWER outputs the following message:

> The null hypothesis is true, but the actual significance level of the test is α' [a numeric value is shown on the screen]. To obtain a new test with actual significance equal to 0.05000 set the nominal alpha to α" [a numeric value is shown on the screen]. The adjusted critical F value is F [a numeric value is shown on the screen].

The ease with which power values are generated using the present analytic solution as implemented by the program MMRPOWER also has practical research value. More precisely, generating power values for various hypothetical scenarios can aid researchers in making decisions about study design and allocating their research resources. For instance, assume a typical situation in which research resources are limited. Also, assume that a researcher is planning a study including a self-report survey and has a choice between two sets of measures for *X* and *Y*. Based on the previous use of these measures, the first set would lead to $\alpha_y\alpha_x = .81$ (i.e., $\alpha_y = \alpha_x = .90$), and these are paper-and-pencil instruments that take about 40 minutes to be completed. The second set of measures would lead to $\alpha_y\alpha_x = .60$ (i.e., $\alpha_y = .80$ and $\alpha_x = .75$), and these instruments take approximately 20 minutes to be completed. Assume that past research shows that, because of the difference in time and effort involved in completing the measures, using the second set of measures improves response rate by about 15%. Should this researcher attempt to increase *N* by using the second set of measures or, alternatively, should this researcher use the first set of measures so as to have less measurement error in *X* and *Y* scores? Which of the two study scenarios will lead to greater power in estimating whether *Z* moderates the relationship between *X* and *Y*? What would be the resulting power value if $\alpha_y\alpha_x$ decreases from .81 to .60 (i.e., by using the second set of measures) but total sample size increases from 200 to 230 (i.e., a 15% increase)? With

MMRPOWER, these questions can be answered via the analytic approximation to compute power for each of the anticipated situations.

Conclusion

MMR is a pervasively used statistical technique to estimate and test effects of categorical moderator variables in organization studies and the social sciences in general. Despite its popularity, researchers often express the concern that the power of MMR is inadequate to test hypotheses regarding the operation of moderating effects. Theorem 2 in Appendix C offers an analytic synthesis of factors affecting the power of MMR. Results of our Monte Carlo simulation indicate that Theorem 2 is accurate at approximating the power of MMR under a very diverse set of conditions including, among other factors, heterogeneity of variance of *X*, heterogeneity of variance of *Y*, and heterogeneity of reliabilities across moderator-based subgroups. We also developed the computer program MMRPOWER that implements the algorithm shown in Theorem 2. MMRPOWER can be used to study further the effects of violating certain assumptions on the power of MMR. Using MMRPOWER with values typically encountered in research situations in which MMR is used yielded power values substantially below Cohen's (1988) recommended .80 value. Thus, it is likely that conclusions of many past studies that used MMR and did not find support for a hypothesized categorical moderator variable were actually the product of a Type II error (i.e., inability to correctly reject a false null hypothesis). In closing, we encourage researchers to approximate power while planning their research design (e.g., sample size) and methodology (e.g., reliability of measurement). By doing so, researchers will make more informed conclusions about hypotheses entailing moderating effects of categorical variables.

APPENDIX A
Assumptions

We consider two cases. In Case 1, the continuous criterion variable *Y* and the continuous predictor *X* are assumed to follow a bivariate normal distribution within each categorical moderator-based subpopulation. In Case 2, the conditional distribution of *Y* given $X = x$ is assumed to be normal within each moderator-based subpopulation, but the marginal distributions of *X* are arbitrary.

True Scores and Error

For Cases 1 and 2, it is assumed that the observable random variables *X* and *Y* can be decomposed as the sum of two statistically independent components, namely, true scores and error. That is,

$$\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} Y_{\text{true}} \\ X_{\text{true}} \end{pmatrix} + \begin{pmatrix} \varepsilon_y \\ \varepsilon_x \end{pmatrix}, \tag{A1}$$

where X_{true} and Y_{true} are true scores and ε_y and ε_x are random errors. Denote the *k* subpopulations by $P_1, P_2, \ldots P_k$. Then, the expectation and covariance matrix corresponding to the true scores in subpopulation *j* can be written as

$$\mathrm{E}\begin{pmatrix} Y_{\text{true}} \\ X_{\text{true}} \end{pmatrix} \Big| P_j \Big) = \boldsymbol{\mu}_j = \begin{pmatrix} \mu_{y,j} \\ \mu_{x,j} \end{pmatrix} \text{And}$$

$$\mathrm{Var}\begin{pmatrix} Y_{\text{true}} \\ X_{\text{true}} \end{pmatrix} \Big| P_j \Big) = \boldsymbol{\Sigma}_{\text{true},j} = \begin{pmatrix} \sigma^2_{y,j} & \rho_j \sigma_{y,j}\sigma_{x,j} \\ \rho_j \sigma_{y,j}\sigma_{x,j} & \sigma^2_{x,j} \end{pmatrix}, \tag{A2}$$

where the notation $|P_j$ means that the result is specific to subpopulation j; ρ_j is the correlation between X_{true} and Y_{true} in subpopulation j; and $\sigma_{x,j}$ and $\sigma_{y,j}$ are the X_{true} and Y_{true} population standard deviations.

The random errors ε_y and ε_x are assumed to be distributed independently of one another. Specifically,

$$\varepsilon_y \mid P_j \sim N(0, \sigma^2_{\varepsilon_{y,j}}) \text{ and } \varepsilon_x \mid P_j \sim (0, \sigma^2_{\varepsilon_{x,j}}). \tag{A3}$$

Note that ε_y is assumed to be normally distributed, whereas only the first two moments, but not the distribution of ε_X, is specified in Equation A3. In Case 1, it is assumed that ε_X is normally distributed, but in Case 2, the distribution of ε_X is arbitrary.

It follows from Equations A1, A2, and A3 that the vector of observable scores has the following expectation and covariance matrix:

$$\mathrm{E}\begin{pmatrix} Y \\ X \end{pmatrix} \Big| P_j \Big) = \mu_j \text{ and } \mathrm{Var}\begin{pmatrix} Y \\ X \end{pmatrix} \Big| P_j \Big) = \Sigma_j = \begin{pmatrix} \sigma^2_{y,j} + \sigma^2_{\varepsilon_{y,j}} & \rho_j \sigma_{y,j}\sigma_{x,j} \\ \rho_j \sigma_{y,j}\sigma_{x,j} & \sigma^2_{x,j} + \sigma^2_{\varepsilon_{x,j}} \end{pmatrix}. \tag{A4}$$

Accordingly, the reliabilities of X and Y are

$$\alpha_{x,j} = \frac{\sigma^2_{x,j}}{\sigma^2_{x,j} + \sigma^2_{\varepsilon_{x,j}}} \text{ and } \alpha_{y,j} = \frac{\sigma^2_{y,j}}{\sigma^2_{y,j} + \sigma^2_{\varepsilon_{y,j}}},$$

respectively.

Case 1: Bivariate Normal Distribution for (*Y*, *X*)

In Case 1, it is assumed that the vector of true scores as well as the vector of random errors follow bivariate normal distributions. That is,

$$\begin{pmatrix} Y_{\text{true}} \\ X_{\text{true}} \end{pmatrix} \Big| P_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{\text{true},j}) \text{ and } \begin{pmatrix} \varepsilon_y \\ \varepsilon_x \end{pmatrix} \Big| P_j \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\varepsilon_{y,j}} & 0 \\ 0 & \sigma^2_{\varepsilon_{x,j}} \end{pmatrix}\right], \tag{A5}$$

where μ_j and $\Sigma_{\text{true},j}$ are given in Equation A2. It follows from Equations A1 and A5 that the vector of observable scores also follows a bivariate normal distribution:

$$\begin{pmatrix} Y \\ X \end{pmatrix} \Big| P_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j),$$

where μ_j is given in Equation A2, and Σ_j is given in Equation A4.

Using standard conditioning arguments, it can be shown that if the pair (Y, X) is randomly drawn from population j, then the distribution of Y conditional on $X = x$ is the following:

$$Y \mid (X = x, P_j) \sim N\left[\beta_{0j} + \beta_{1j}x, \frac{\sigma^2_{y,j}}{\alpha_{y,j}}(1 - \rho^2_j \alpha_{x,j} \alpha_{y,j})\right], \tag{A6}$$

where

$$\beta_{0j} = \mu_{y,j} - \mu_{x,j}\beta_{1j} \text{ and } \beta_{1j} = \rho_j \alpha_{x,j} \frac{\sigma_{y,j}}{\sigma_{x,j}}.$$

Case 2: Arbitrary Distribution for *X*

In Case 2, the distributions of X_{true}, ε_x, and X are arbitrary, provided that the moments in Equations A2 and A3 exist. No specific marginal distributions are assumed for either Y_{true} or Y, but it is assumed that Y is conditionally normal given $X = x$. Specifically, it is assumed that the regression model in Equation A6 holds.

APPENDIX B
Testing Hypotheses Regarding Moderating Effects

In testing moderating effect hypotheses a sample of n_j pairs is randomly selected from the jth population, for $j = 1, \ldots, k$. These selections may be made completely at random, or they may be made according to a restricted random process. In any case, it is assumed that conditional on $X = x$, the probability of selecting the pair (Y, X) is independent of Y. If this assumption is satisfied, then the conditional distribution of Y given $X = x$ for sampled pairs (Y, X) is identical to the conditional distribution of Y given $X = x$ in the entire population (i.e., Eq. A6). One simple example of this type of restricted sampling is truncation on X. That is, selection could be random subject to the restriction that each (Y, X) pair in sample j satisfies $X \geq x^*_j$ or $X \leq x^*_j$, where x^*_j is a fixed lower or upper cutoff.

Null Hypothesis

To test the moderating effect hypotheses shown in Equation 1, one tests the hypothesis H_0: $\beta_{11} = \beta_{12} = \ldots = \beta_{1k}$. This hypothesis is equivalent to

$$H_0: \rho_1 \alpha_{x,1} \frac{\sigma_{y,1}}{\sigma_{x,1}} = \rho_2 \alpha_{x,2} \frac{\sigma_{y,2}}{\sigma_{x,2}} = \ldots = \rho_k \alpha_{x,k} \frac{\sigma_{y,k}}{\sigma_{x,k}}.$$

However, as reviewed by Aguinis and Pierce (1998a), a false null hypothesis could be attributable to heterogeneity among the correlations, the X variances, the Y variances, and/or the X reliabilities.

Denote the $k \times 1$ vector of slope parameters by $\boldsymbol{\beta}_1$. That is,

$$\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1k} \end{pmatrix}.$$

A vector of moderating effects can be obtained by computing contrasts among the entries of β_1. Specifically, let **C** be a $k \times (k-1)$ matrix of contrast coefficients with entries c_{ij} for $i = 1, \ldots, k$ and $j = 1, \ldots, k-1$ defined as follows:

$$c_{ij} = \begin{cases} 1 \text{ if } i = j; \\ -1 \text{ if } i = k; \text{ and} \\ 0 \text{ otherwise.} \end{cases} \tag{B1}$$

For example, if $k = 4$, then

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix}.$$

Using this matrix notation, the vector of moderating effects is $\mathbf{C}'\boldsymbol{\beta}_1$. For convenience, the vector slope differences, $\mathbf{C}'\boldsymbol{\beta}_1$, will be denoted by $\boldsymbol{\psi}$. That is,

$$\mathbf{C}'\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_{11} - \beta_{1k} \\ \beta_{12} - \beta_{1k} \\ \vdots \\ \beta_{1,k-1} - \beta_{1k} \end{pmatrix} = \boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{k-1} \end{pmatrix}.$$

In short, the null hypothesis can be written either as H_0: $\mathbf{C}'\beta_1 = 0$ or as H_0: $\psi = 0$. (Note that one-degree-of-freedom contrasts also could be tested as opposed to an omnibus test given that a theory-based hypothesis exists regarding differences between/among specific groups; West, Aiken, & Krull, 1996.)

Conventional MMR Model

For convenience, the total sample size, $\Sigma_{j=1}^{k} n_j$ is denoted by N. In addition, the partial sum $\Sigma_{g=1}^{j} n_g$ is denoted by N_j, for $j = 1, \ldots, k$; and N_0 is defined as zero. That is, $N_0 = 0$; $N_1 = n_1$; $N_2 = n_1 + n_2$, and so forth up to $N_k = N$. The pooled sample of pairs is (Y_i, X_i) for $i = 1, \ldots, N$, and the sample of size n_j from population j is (Y_i, X_i) for $i = 1 + N_{j-1}, 2 + N_{j-1}, \ldots, N_j$.

The moderator variable Z can take on the values $1, 2, \ldots, k$. The value of Z indicates the subpopulation from which a (Y, X) pair was drawn. That is, if the pair (Y_i, X_i) was drawn from subpopulation j, then $z_i = j$. For each z_i, a set of $k - 1$ binary indicator variables, w_{ij} for $j = 1, \ldots, k - 1$, can be defined as follows:

$$w_{ij} = \begin{cases} 1 \text{ if } z_i = j; \\ 0 \text{ otherwise.} \end{cases} \tag{B2}$$

In the special case of a binary moderator variable (i.e., $k = 2$), the subscript j on w_{ij} can be omitted, and the binary indicator variable, w_i, simplifies to

$$w_i = \begin{cases} 1 \text{ if } z_i = 1; \\ 0 \text{ if } z_i = 2. \end{cases}$$

The conventional MMR model for (Y_i, X_i), $i = 1, \ldots, N$ can be written as follows:

$$Y_i = \beta_0 + \beta_1 x_i = \sum_{j=1}^{k-1} w_{ij}\tau_j + \sum_{j=1}^{k-1} w_{ij} x_i \psi_j + \varepsilon_i, \tag{B3}$$

where w_{ij} is defined in Equation B2, and the parameters of the MMR model are functions of the parameters of the conditional distributions in Equation A6. Specifically,

$$\beta_0 = \beta_{0k};\ \beta_1 + \beta_{1k};\ \tau_j = \beta_{0j} - \beta_{0k} \text{ for } j = 1, \ldots, k-1; \text{ and}$$

$$\psi_j = \beta_{1j} - \beta_{1k} \text{ for } j = 1, \ldots, k-1.$$

Also, ε_i for $i = 1, \ldots, N$ are independently distributed random errors. The random error terms are independently distributed as

$$\varepsilon_i | P_j \sim N\left[0, \frac{\sigma^2_{y,j}}{\alpha_{y,j}}(1 - \rho^2_j \alpha_{x,j}\alpha_{y,j})\right] \text{ for } i = 1 + N_{j-1}, 2 + N_{j-1}, \ldots, N_j.$$

F Test

The conventional MMR F test for moderating variable effects is to reject H_0 if $F_x \geq F^{1-\alpha}_{k-1,N-2k}$, where F_x is the computed value of the test statistic, and $F^{1-\alpha}_{k-1,N-2k}$ is the critical value for a size α test. The subscript x on F_x is a reminder that the test is conducted conditional on $X = x$. The test statistic can be computed as follows:

$$F_x = \frac{\hat{\boldsymbol{\psi}}'(\mathbf{C}'\mathbf{D}_x\mathbf{C})^{-1}\hat{\boldsymbol{\psi}}}{(k-1)MSE}, \tag{B4}$$

where $\mathbf{C}$ is defined in Equation B1; $\hat{\boldsymbol{\psi}}$ is the ordinary least squares estimator of $\boldsymbol{\psi}$ in the MMR model (Eq. B3);

$$\mathbf{D}_x = \text{Diag}(SSX_j^{-1}; j = 1, \ldots, k);$$

$$SSX_j = \sum_{i=1+N_{j-1}}^{N_j} (x_i - \bar{x}_j)^2;\ \bar{x}_j = \frac{1}{n_j} \sum_{i=1+N_{j-1}}^{N_j} x_i;$$

$$MSE = \frac{SSE}{N - 2k} = \frac{\Sigma_{j=1}^{k} SSE_j}{N - 2k};$$

Diag(a_j; $j = 1, \ldots, k$) is a diagonal matrix given by

$$\mathrm{Diag}(a_j;\, j = 1, \ldots, k) = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_k \end{pmatrix};$$

SSE is the sum of squared errors from the least squares fit of the MMR model (Eq. B3), and SSE_j for $j = 1, \ldots, k$ is the sum of squared errors from the least squares fit of the parameters in the *j*th conditional model described in Equation A6.

APPENDIX C
Distribution of the Test Statistic Conditional on *X* and Approximation to the Unconditional Distribution of the Test Statistic

Distribution of the Test Statistic Conditional on *X*

In general, the test statistic F_x will not be distributed as an *F* random variable. Under the model shown in Equation A6, the condition that must be satisfied for F_x to follow an *F* distribution is homogeneity of the conditional variances of *Y* given *X* (Aguinis & Pierce, 1998a). This homogeneity condition can be written as follows:

$$\frac{\sigma^2_{y,j}}{\alpha_{y,j}}(1 - \rho^2_j \alpha_{x,j}\alpha_{y,j}) = \sigma^2 \text{ for } j = 1, \ldots, k, \tag{C1}$$

where σ^2 is a positive constant. If the homogeneity condition is not satisfied, then the F_x is not distributed as an *F* random variable. Nonetheless, the exact distribution of F_x, conditional on *X*, can be obtained by using well-known linear models theory (e.g., Stapleton, 1995). The result is summarized in Theorem 1.

Theorem 1: Conditional Distribution of F_x

Define $\mathbf{V}_x$ as

$$\mathbf{V}_x = \mathrm{Diag}\left[\frac{\sigma^2_{y,j}(1-\rho^2_j\alpha_{x,j}\alpha_{y,j})}{\alpha_{y,j}SSX_j};\, j = 1,\ldots,k\right]$$

and denote the eigenvalues and eigen-vectors of $(\mathbf{C}'\mathbf{D}_x\mathbf{C})^{-1}\,\mathbf{C}'\mathbf{V}_x\mathbf{C}$ by $\omega_{x,j}$ and $\mathbf{u}_{x,j}$, respectively. That is,

$$(\mathbf{C}'\mathbf{D}_x\mathbf{C})^{-1}\,\mathbf{C}'\mathbf{V}_x\mathbf{C}\mathbf{u}_{x,j} = \mathbf{u}_{x,j}\omega_{x,j}$$

for $j = 1, \ldots, k - 1$. The distribution of F_x, conditional on *X*, is

$$F_x \sim \left(\frac{N-2k}{k-1}\right)\left[\frac{\sum_{j=1}^{k-1}\omega_{x,j}G_{x,j}}{\sum_{j=1}^{k}\frac{\sigma_{y,j}^2(1-\rho_j^2\alpha_{x,j}\alpha_{y,j})}{\alpha_{y,j}}H_j}\right],$$

where $G_{x,j}$ for $j = 1, \ldots, k-1$ and H_j for $j = 1, \ldots, k$ are independently distributed chi-squared random variables. Specifically, $H_j \sim \chi^2(n_j - 2)$ for $j = 1, \ldots, k$ and $G_{x,j} \sim \chi^2(1, \lambda_{x,j})$ for $j = 1, \ldots, k-1$, where $\lambda_{x,j}$ is a noncentrality parameter and is given by

$$\lambda_{x,j} = \frac{(\mathbf{u}'_{x,j}\mathbf{C}'\boldsymbol{\beta}_l)^2}{2\mathbf{u}'_{x,j}\mathbf{C}'\mathbf{V}_x\mathbf{C}\mathbf{u}_{x,j}}.$$

In the special case of a binary moderator variable, the conditional distribution of F_x simplifies substantially. In this special case, $k - 1 = 1$;

$$\omega_{x,1} = \frac{\left[\frac{\sigma_{y,1}^2(1-\rho_1^2\alpha_{x,1}\alpha_{y,1})}{\alpha_{y,1}SSX_1}+\frac{\sigma_{y,2}^2(1-\rho_2^2\alpha_{x,2}\alpha_{y,2})}{\alpha_{y,2}SSX_2}\right]}{\left(\frac{1}{SSX_1}+\frac{1}{SSX_2}\right)};\text{ and}$$

$$\lambda_{x,1} = \frac{(\beta_{11}-\beta_{12})^2}{2\left(\frac{\sigma_{y,1}^2(1-\rho_1^2\alpha_{x,1}\alpha_{y,1})}{\alpha_{y,1}SSX_1}+\frac{\sigma_{y,2}^2(1-\rho_2^2\alpha_{x,2}\alpha_{y,2})}{\alpha_{y,2}SSX_2}\right)}.$$

Conditional on the X variable, the power of the moderating effect tests is given by

$$\text{Power}_x = \Pr(F_x \geq F^{1-\alpha}_{k-1,N-2k}). \tag{C2}$$

Approximation to the Unconditional Distribution of the Test Statistic

To obtain the unconditional power of the F test, the conditional power in Equation C2 must be averaged over all possible realizations of the X variable. That is, the unconditional power of the F test is

$$\text{Power} = \text{E}(\text{Power}_x), \tag{C3}$$

where Power_x is given in Equation C2, and the expectation is taken with respect to the distribution of X. Exact analytic expressions for the unconditional power are not known. Accordingly, the unconditional power must be approximated.

The X variable plays a role in the conditional power solely through SSX_j for $j = 1,\ldots, k$. The strategy for the proposed approximation is straightforward. First, for each j, replace $1/SSX_j$ by its expectation. Second, use Theorem 1 and act as though $1/SSX_j = \mathrm{E}(1/SSX_j)$ for $j = 1, \ldots, k$. This strategy is equivalent to expanding the unconditional power in Equation C3 in a Taylor series around $1/SSX_j = \mathrm{E}(1/SSX_j)$ for $j = 1, \ldots, k$ and truncating the series after the linear term. With an error of order n_j^{-2}, the expectation of $1/SSX_j$ is

$$\mathrm{E}\left(\frac{1}{SSX_j}\right) = \frac{n_j + 1}{(n_j - 1)\mathrm{E}(SSX_j)}.$$

The relationship between $\mathrm{E}(SSX_j)$ and $\sigma^2_{x,j}$ depends on the reliability of X in subpopulation j and on the manner in which (Y, X) pairs were selected from the jth subpopulation. Let Q be a random variable that takes on the value $Q = 1$ if the pair (Y, X) can be selected from the population and takes on the value $Q = 0$ if the pair (Y, X) cannot be selected from the population. If the (Y, X) pairs are a simple random sample from the jth population, then $Q = 1$ for all pairs, and

$$E(S^2_{x_j}) = \frac{\sigma^2_{x,j}}{\alpha_{x,j}}, \text{where } S^2_{x_j} = \frac{SSX_j}{n_j - 1}$$

is the sample variance of X in the jth sample. If the probability of selecting the pair (Y, X) depends on X, then

$$\mathrm{E}(S^2_{x_j}) = \frac{\sigma^2_{x,j}\delta_j}{\alpha_{x,j}} \text{ for } j = 1, \ldots, k; \text{ where } \delta_j = \frac{\mathrm{Var}(X|P_j, Q = 1)}{\mathrm{Var}(X|P_j)}. \tag{C4}$$

The quantity δ_j is a multiplying factor that depends on the distribution of X and on the manner in which selection probabilities depend on X.

If sampling is restricted by left truncation on X, then

$$Q = \begin{cases} 1 \text{ if } X \geq x_j^*, \text{and} \\ 0 \text{ otherwise.} \end{cases}$$

If the distribution of X is known, then, for this simple form of restricted sampling, the multiplying factor δ_j can be computed explicitly. For example, if X is normally distributed, then (using moment generating functions) the multiplying factor can be shown to be

$$\delta_j = 1 + \frac{\varphi(h_j)}{1 - T_j}\left(h_j - \frac{\varphi(h_j)}{1 - T_j}\right);$$

where φ is the probability density function of the standard normal distribution, $h_j = \sqrt{\alpha_{x,j}}$ ($x_j^* - \mu_{x,j})/\sigma_{x,j}$; and T_j is the truncation proportion for subpopulation j. That is, the sample consists of pairs (Y, X) that are randomly selected from all pairs in population j for which X lies in the upper or lower $100[1 - T_j]\%$ of the distribution.

In the simple case of normally distributed X and left truncation, the multiplying factor δ_j always is less than or equal to 1.00. In other cases, δ_j can be less than, equal to, or greater than 1.00. For example, suppose that the density function of X in subpopulation j is $f_x(x)$. Consider the selection mechanism

$$\text{Prob}(Q = 1|X = x) = F_x(x)^{\frac{T}{1-T}}, \text{ where } F_x(x) = \int_{-\infty}^{x} f_x(u)du$$

is the cumulative distribution function of X, and T is a constant in (0, 1). With this selection mechanism, the probability of selection gradually increases as X increases. This selection mechanism is a continuous version of left truncation, and it will be called sparse left sampling. It can be shown that $E[\text{Prob}(Q = 1|X)] = 1 - T$. That is, the percentage of the distribution that cannot be sampled is $100T\%$. The value of T, therefore, is analogous to the truncation proportion discussed above. The conditional density of X given $Q = 1$ is

$$f_{x|Q}(x|Q = 1) = \frac{1}{1-T} f_x(x)F_x(x)^{\frac{T}{1-T}}.$$

To compute δ_j, the density function f_x(x)must be known. As an example, suppose that X follows an exponential distribution in subpopulation j. That is, the density function is

$$f_X(x|\lambda) = \lambda e^{-\lambda x}, \text{ for } 0 < x < \infty.$$

It can be shown that δ_j increases from $\delta_j = 1$ to $\delta_j = 1.645$ as the "truncation proportion" (i.e., T) increases from 0 to 1. That is, the sample variances tend to be inflated, and the degree of inflation increases as T increases.

In sparse right sampling, the probability of selection is high for low X scores, and it decreases as X increases. The probability of selection in sparse right sampling is

$$\text{Prob}(Q = 1|X = x) = [1 - F_x(x)]^{\frac{T}{1-T}}.$$

If an exponential distribution (in subpopulation j) is subjected to sparse right sampling, then δ_j decreases from $\delta_j = 1.00$ to $\delta_j = 0.00$ as T increases from 0.00 to 1.00. If the selection mechanism is changed to the left truncation mechanism,

$$\text{Prob}(Q = 1|X = x) = \begin{cases} 1 \text{ if } x > x^*, \text{and} \\ 0 \; x < x^* \end{cases},$$

then for the exponential distribution, $\delta_j = 1$ regardless of the cutoff value x^*. In summary, if X follows an exponential distribution, then $\delta_j = 1$, $\delta_j < 1$, or $\delta_j > 1$ depending on the sampling mechanism.

Next, Theorem 2 summarizes the power approximation obtained by making the required substitutions.

Theorem 2: Unconditional Power Approximation

The power of the MMR F test is

$$\text{Power} \approx \Pr\left[\left(\frac{k-1}{N-2k}\right)F_{k-1,N-2k}^{1-\alpha}\sum_{j=1}^{k}\frac{\sigma_{y,j}^2(1-\rho_j^2\alpha_{x,j}\alpha_{y,j})}{\alpha_{y,j}}H_j - \sum_{j=1}^{k-1}\omega_j G_j \le 0\right],$$

where ω_j is the jth eigenvalue of $(\mathbf{C'DC})^{-1}\,\mathbf{C'VC}$;

$$\mathbf{D} = \text{Diag}\left[\frac{\alpha_{x,j}(n_j+1)}{(n_j-1)^2\delta_j\sigma^2_{x,j}}; j = 1,\ldots,k\right];$$

$$\mathbf{V} = \text{Diag}\left[\frac{\sigma^2_{y,j}\alpha_{x,j}(1-\rho^2_j\alpha_{x,j}\alpha_{y,j})(n_j+1)}{\alpha_{y,j}(n_j-1)^2\delta_j\sigma^2_{x,j}}; j = 1,\ldots,k\right];$$

and G_j for $j = 1, \ldots, k - 1$ and H_j for $j = 1, \ldots, k$ are independently distributed chi-squared random variables. Specifically, $H_j \sim \chi^2(n_j - 2)$ for $j = 1, \ldots, k$ and $G_j \sim \chi^2(1, \lambda_j)$ for $j = 1, \ldots, k - 1$, where λ_j is a noncentrality parameter;

$$\lambda_j = \frac{(\mathbf{u}'_j\,\mathbf{C}'\boldsymbol{\beta}_1)^2}{2\mathbf{u}'_j\,\mathbf{C}'\mathbf{V}\mathbf{C}\mathbf{u}_j};$$

and $\mathbf{u}_j$ is the jth eigen-vector of $(\mathbf{C}'\mathbf{D}\mathbf{C})^{-1}\,\mathbf{C}'\mathbf{V}\mathbf{C}$.

The accuracy of the approximation in Theorem 2 increases as each n_j increases. Also, for any fixed sample size, the approximation is most accurate if the standardized kurtosis coefficients of the X variable are near zero (i.e., normal kurtosis) in each subpopulation.

Notes

1. We acknowledge that there is also a contrary view. For instance, Luce (1995) stated that

> evidence of interactions is usually a signal of trouble. . . . All too often, in my opinion, the interactions are treated as a finding and not as evidence of a lack of understanding of the combining rule for measures of the independent variables. (p. 21)

However, as noted by Aguinis (in press), Aguinis and Pierce (1998c), Aguinis and Whitehead (1997), and others (e.g., Hall & Rosenthal, 1991), it is the *theory-based* interaction effects that are “at the very heart of the scientific enterprise” (Hall & Rosenthal, 1991, p. 447). Alternatively, unexpected and/or unanticipated interaction effects can be problematic. They might lead to a meaningful discovery or simply indicate that the conception of the research question and/or design is incorrect.

2. Increasing sample size across subgroups is related but not equivalent to reduction of variance in the categorical predictor Z. An increase in sample size can result in a decrease or an increase in the variance of Z. For example, if Z is a categorical moderator that takes on values of 1 and 2, the sample variance of Z is

$$S^2_z = \frac{\Sigma(Z_i - \bar{Z})^2}{N-1} = \frac{Np(1-p)}{N-1},$$

where $N = n_1$ (i.e., sample size in Subgroup 1) + n_2 (i.e., sample size in Subgroup 2), and $p = n_1/N$. As an illustration, if n_1 = 10 and n_2 = 40, then the sample variance is 8/49 = .1636. If n_1 is increased to 40, then the sample variance is 20/79 = .253 (i.e., an increase). If n_1 remains at 10 and n_2 is increased to 50, then the sample variance is 25/177 = .1412 (i.e., a decrease).

3. Note, however, that the computer program MMRPOWER that implements the analytic approximation allows for the inclusion of up to 20 levels or moderator-based subgroups (i.e., $k \leq$

20). Tables showing results for conditions $4 \leq k \leq 20$ are available from the authors. The pattern of results was similar to those reported herein for conditions $2 \leq k \leq 3$.

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, *21*, 1141-1158.

Aguinis, H. (in press). Estimation of sampling variance of correlations in meta-analysis. *Personnel Psychology*.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2000, April). Statistical power of differential prediction analysis: A 30-year review. In F. L. Oswald (Chair), *Differential prediction in personnel selection: Past, present, and future*. Symposium conducted at the meeting for the Society of Industrial and Organizational Psychology, New Orleans, LA.

Aguinis, H., Bommer, W. H., & Pierce, C. A. (1996). Improving the estimation of moderating effects by using computer-administered questionnaires. *Educational and Psychological Measurement*, *56*, 1043-1047.

Aguinis, H., Nesler, M. S., Quigley, B. M., Lee, S., & Tedeschi, J. T. (1996). Power bases of faculty supervisors and educational outcomes for graduate students. *Journal of Higher Education*, *67*, 267-297.

Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods*, *2*, 315-339.

Aguinis, H., & Pierce, C. A. (1998a). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods*, *1*, 296-314

Aguinis, H., & Pierce, C. A. (1998b). Statistical power computations for detecting dichotomous moderator variables with moderated multiple regression. *Educational and Psychological Measurement*, *58*, 668-676.

Aguinis, H., & Pierce, C. A. (1998c). Testing moderator variable hypotheses meta-analytically. *Journal of Management*, *24*, 577-592.

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. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, *82*, 192-206.

Aguinis, H., & Whitehead, R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology*, *82*, 528-538.

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

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.

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

Boik, R. J. (1979). Interactions, partial interactions and interaction contrasts in the analysis of variance. *Psychological Bulletin*, *86*, 1084-1089.

Boik, R. J. (1993). The analysis of two-factor interactions in fixed effects linear models. *Journal of Educational Statistics*, *18*, 1-40.

Brown, P. J., & Fuller, W. A. (Eds.). (1990). *Statistical analysis of measurement error models and applications*. Providence, RI: American Mathematical Society.

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.

Carroll, R. J., & Ruppert, D. (1995). *Measurement error in nonlinear models*. New York: Chapman & Hall.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Belmont, CA: Duxbury.

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

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

DeShon, R. P., & Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods*, *1*, 261-277.

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

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, *106*, 516-624.

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

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.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, *46*, 1-26.

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

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

Murphy, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. *Psychological Bulletin*, *99*, 133-138.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

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

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

Stapleton, J. H. (1995). *Linear statistical models*. New York: John Wiley.

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 for dichotomous variables. *Journal of Management*, *20*, 167-178.

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

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality*, *64*, 1-48.

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

Herman Aguinis (http://www.cudenver.edu/~haguinis) is an associate professor of management at the University of Colorado at Denver. He received a Ph.D. in industrial/organizational psychology from the University at Albany, State University of New York. His current research interests include personnel selection, social power and influence, estimation of interaction effects, meta-analysis, and research methods.

Robert J. Boik (http://www.math.montana.edu/~rjboik) is a professor of statistics at Montana State University. He received a Ph.D. in experimental psychology from Baylor University and a Ph.D. in statistics from Temple University. His current research interests include linear models, multivariate statistics, and Bayesian methods.

Charles A. Pierce (http://www.montana.edu/wwwpy) is an associate professor of industrial/organizational psychology at Montana State University. He received a Ph.D. in social psychology from the University at Albany, State University of New York. His current research interests include workplace romance, sexual harassment, estimation of interaction effects, meta-analysis, and research methods.