Journal of Applied Psychology
1997, Vol. 82, No. 1, 192–206

0021-9010/97/$3.00

Methodological Artifacts in Moderated Multiple Regression and Their Effects on Statistical Power

Herman Aguinis
University of Colorado at Denver

Eugene F. Stone-Romero
University at Albany, State University of New York

Monte Carlo simulations were conducted to examine the degree to which the statistical power of moderated multiple regression (MMR) to detect the effects of a dichotomous moderator variable was affected by the main and interactive effects of (a) predictor variable range restriction, (b) total sample size, (c) sample sizes for 2 moderator variable-based subgroups, (d) predictor variable intercorrelation, and (e) magnitude of the moderating effect. Results showed that the main and interactive influences of these variables may have profound effects on power. Thus, future attempts to detect moderating effects with MMR should consider the power implications of both the main and interactive effects of the variables assessed in the present study. Otherwise, even moderating effects of substantial magnitude may go undetected.

Numerous theories in applied psychology posit the operation of moderator variables, that is, variables that interact with others in explaining variance in a dependent variable. Moreover, in recent years, interest in moderator variables has increased substantially in applied psychology (e.g., MacCallum & Mar, 1995; Nesler, Aguinis, Quigley, & Tedeschi, 1993; Sagie & Koslowsky, 1993; Schmitt, Hattrup, & Landis, 1993) as well as in several related disciplines, including administrative science, education, organizational behavior, sociology, and political science (e.g., Aguinis, Bommer, & Pierce, 1996; Aguinis, Pierce, & Stone-Romero, 1994; Cordes & Dougherty, 1993; Pierce, Gardner, Dunham, & Cummings, 1993; Walsh & Kosnik, 1993; Xie & Johns, 1995). In industrial and organizational psychology, moderating effects of variables such as ethnicity and gender on the relationship between preemployment test scores and measures of performance suggest that the test does not predict performance equally well for the subgroups under consideration (e.g., minorities and nonminorities). Consequently, if a moderator such as ethnicity is found, there is differential prediction, and the preemployment test is considered to be biased for certain subgroups (Bartlett, Bobko, Mosier, & Hannan, 1978; Cleary, 1968; Linn, 1994; Society for Industrial and Organizational Psychology; SIOP, 1987).

In a seminal article, Zedeck (1971) noted that Z is a moderator of the relationship between variables X and Y when the nature (e.g., magnitude) of this relationship varies across levels of Z. He also described a number of statistical procedures that can be used for estimating moderating effects, including moderated multiple regression (cf. Saunders, 1956). In contrast with the time in which Zedeck's article was published, moderated multiple regression (MMR) is now routinely used to estimate and interpret the effects of both dichotomous and continuous moderators (e.g., Bartlett et al., 1978; Cronbach & Snow, 1977; Smith & Sechrest, 1991; Stone, 1988; Stone & Hollenbeck, 1989). Note that MMR is preferred over other strategies, such as the comparison of subgroup-based correlation coefficients for two or more subgroups (Stone-Romero & Anderson, 1994). One reason for this preference is that results of an MMR analysis provide researchers with important information that is not provided by tests of the equality of correlation coefficients. More specifically, MMR provides information about slope differences for various subgroups. This information is

Herman Aguinis, Department of Psychology, University of Colorado at Denver; Eugene F. Stone-Romero, Department of Psychology, University at Albany, State University of New York.

Portions of this article are based on Herman Aguinis's doctoral dissertation, which was completed at the University at Albany, State University of New York, under the supervision of Eugene F. Stone-Romero. A previous version of this article was presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, Nashville, Tennessee, April 1994. We are grateful to George M. Alliger, James J. Jaccard, Lawrence E. Jones, Gary H. McClelland, Charles A. Pierce, and Stephen G. West for helpful comments, suggestions, and advice. However, the opinions expressed in this article are the sole responsibility of the authors.

Correspondence concerning this article should be addressed to Herman Aguinis, Department of Psychology, University of Colorado at Denver, Campus Box 173, P.O. Box 173364, Denver, Colorado 80217-3364. Electronic mail may be sent via Internet to haguinis@castle.cudenver.edu. World Wide Web address is http://www.cudenver.edu/~haguinis.

critical in assessing differential prediction (SIOP, 1987). For instance, a recent review by Sackett and Wilk (1994) concluded that MMR is the most pervasively used statistical technique for assessing differential prediction in organizational and educational settings.

A common strategy for testing hypotheses regarding moderator variables with MMR relies upon the statistical test of the unstandardized regression coefficient carrying information about the moderating (i.e., interaction) effect (Aiken & West, 1991; Cohen & Cohen, 1983; Darlington, 1990; Jaccard, Turrisi, & Wan, 1990). Given a criterion or dependent variable Y, a predictor variable X, and a second predictor variable Z hypothesized to moderate the $X-Y$ relationship, a product term is formed between the predictor and the moderator (i.e., $X \cdot Z$). Then, a regression model is formed including predictor variables X, Z, and the $X \cdot Z$ product term, which carries information regarding their interaction. The statistical significance of the unstandardized regression coefficient of the product term (i.e., $b_{X \cdot Z}$) indicates the presence of the interaction.

Note that a moderator variable Z can be continuous (e.g., age) or dichotomous (e.g., gender). In the case of a dichotomous moderator variable, dummy codes (e.g., 0 = men and 1 = women) may be used in the regression analysis.

Statistical Power Problems in Previous Studies of Moderating Effects

Cronbach (1987) observed that there is a "frequent failure of [hypothesized] interactions [i.e., moderating effects] to be [statistically] significant" (p. 417). Consistent with this argument, an examination of recent attempts to detect the effects of moderator variables using MMR reveals that they are often unsuccessful. Among the recent studies that may have failed to detect moderating effects because of insufficient power are those by Cortina, Doherty, Schmitt, Kaufman, and Smith (1992), Hattrup and Schmitt (1990), Schmitt et al. (1993), and Wohlgemuth and Betz (1991).

In view of the statistical power problems with MMR that militate against the detection of moderating effects, hypothesized moderator variables have been characterized as elusive (e.g., Zedeck, 1971). Not surprisingly, therefore, in recent years there have been several calls for additional research on the power of various statistical techniques, including MMR, to detect moderating effects (e.g., Cronbach, 1987; Smith & Sechrest, 1991). Stated differently, there is the need to determine the conditions under which the use of MMR may lead researchers to conclude, erroneously, that there are no moderating effects (i.e., commit a Type II statistical error). This is not a trivial issue because the failure of empirical research to reveal moderating effects may lead researchers to conclude, erroneously, that theories that entail interaction effects are invalid. In addition, on a more applied note, the failure to detect a moderating effect may lead practitioners to use personnel selection tests that are biased toward specific (e.g., protected) subgroups (e.g., women; Dunbar & Novick, 1988).

Previous Research on the Power to Detect Moderating Effects

In response to concerns regarding the detection of moderating effects, several Monte Carlo studies have recently been conducted to assess the impact of various common methodological artifacts (e.g., small sample size) on the power of MMR to detect moderating effects (referred to hereinafter as *MMR power*). The results of this research show that MMR power problems stem from such study-related artifacts as (a) total sample size (e.g., Alexander & DeShon, 1994; Stone-Romero & Anderson, 1994), (b) unequal sample sizes across moderator variable-based subgroups (Stone-Romero, Alliger, & Aguinis, 1994), (c) unreliability of predictor variable scores (e.g., Bohrnstedt & Marwell, 1978; Busemeyer & Jones, 1983; Dunlap & Kemery, 1988), (d) predictor variable intercorrelation (Dunlap & Kemery, 1988), and (e) magnitude of the moderating effect in the population (e.g., Stone-Romero & Anderson, 1994; Stone-Romero et al., 1994; see Aguinis, 1995, for a review of factors affecting MMR power and potential courses of action regarding low MMR power situations). In general, however, these simulation studies have only considered the effects of a limited number of variables that are presumed to influence MMR power. For example, Dunlap and Kemery (1988) investigated the effects of just two factors: (a) multicollinearity (i.e., degree of predictor intercorrelation), and (b) predictor score unreliability. More recently, Stone-Romero et al. (1994) manipulated only three design-related factors: (a) proportions of cases in 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.

Purpose of the Present Study

In view of the critical implications of low MMR power for applied psychology theory and practice and of the limited focus of many previous studies; the purpose of this study was to address several prior deficiencies and extend the results of previous simulation research on MMR power.

The study differed from previous research in three regards. First, although factors affecting MMR power have typically been examined in the context of continuous moderator variables (e.g., Aiken & West, 1991; Evans, 1985; McClelland & Judd, 1993), the present study's concern was with dichotomous moderators (e.g., gender or minority status), which are variables typically studied by applied psychologists (e.g., personnel selection researchers). Note that although the investigation of factors affecting MMR power to detect dichotomous moderator variables could possibly be considered analytically, empirically mimicking specific design-related conditions typically present in research using MMR provides an explanation for the numerous recent failures to gather support for sound, theory-based hypotheses regarding dichotomous moderators.

Second, in the study we assessed the effects of predictor variable range restriction on the power of MMR to detect moderating effects. Range restriction has interested industrial and organizational psychologists and other social scientists for nearly a century (Pearson, 1903). One reason for this interest is that range restriction is a pervasive phenomenon in applied psychological research, especially in research conducted in field settings (Cook & Campbell, 1979; Guion, 1991; Linn, 1968). For example, personnel selection procedures are a major cause of range restriction (Thorndike, 1949, pp. 169–176). Decisions regarding which individuals to select for an opening are frequently based on their standing on predictor variable X (e.g., test of job aptitude): Only those who obtain a score that exceeds a specific cutoff point (a) are selected, leading to explicit range restriction on X. As a result, in tests of relationships between the predictor X and criterion Y (e.g., measure of job performance), data will be available only for individuals whose scores on X exceed the defined cutoff score a (Thorndike, 1949).

Range restriction has important implications for the ability of MMR to detect moderating effects (i.e., MMR power). More specifically, in the presence of range restriction on predictor variable X, the variance of $X \cdot Z$ scores in a range-restricted sample, $s^{2*}_{X \cdot Z}$ (where the asterisk represents a statistical estimate derived from a range-restricted sample) will be lower than the variance of $X \cdot Z$ scores in the unrestricted population, $\sigma^2_{X \cdot Z}$, and this will affect the ability of MMR to detect a moderating effect. The reason for the predicted decrease in MMR power is that, in testing for a moderating effect, the null hypothesis of $\beta_{X \cdot Z} = 0$ will be rejected only if the squared semipartial correlation between Y and the product term $X \cdot Z$ (i.e., $r^2_{Y(X \cdot Z.XZ)}$) differs from zero. In this situation, R^2 for the model that includes the interaction term will be greater than R^2 for the model that considers only the main effects of X and Z (Cohen & Cohen, 1983). However, note that because the magnitude of $r^2_{Y(X \cdot Z.XZ)}$ is determined, in part, by the variances of X, Y, and Z (Lord & Novick, 1968; Nunnally & Bernstein, 1994), restriction of range on one or more of these variables will lead to a sample estimate of the squared semipartial correlation coefficient, $r^{2*}_{Y(X \cdot Z.XZ)}$, that underestimates $\rho^2_{Y(X \cdot Z.XZ)}$ (i.e., the squared semipartial correlation coefficient in the population; Guilford & Fruchter, 1973; Kelley, 1923; Otis, 1922; Thorndike, 1947). In summary, the second purpose of the present research was to investigate the extent to which the degree of predictor variable (X) range restriction lowered the power of MMR to detect a moderating effect.

Third, unlike several previous simulation studies, this study considered not only the main effects but also the interactive effects of a number of research-related factors that may influence MMR power. This is a very consequential issue because when MMR is used in industrial and organizational psychology research (e.g., validation research), investigators often encounter situations in which several of the factors known to affect power adversely (e.g., small total sample size, unequal sample sizes across dichotomous moderator-based subgroups, and predictor variable range restriction) are present simultaneously (Schmidt, Hunter, Pearlman, & Hirsh, 1985). Note that although the main effects of some of the aforementioned artifacts (e.g., total sample size) are well-known, the estimation of all their potential interactive effects is intractable mathematically. Accordingly, as noted earlier, the third contribution of the present research was to investigate empirically the interactive effects of five factors on MMR power. Our simulation involved the manipulation of the following variables: (a) predictor variable range restriction, (b) total sample size, (c) sample sizes of the two subgroups associated with a dichotomous moderator variable, (d) predictor variable intercorrelation, and (e) magnitude of the moderating effect.

Method

Overview

The present study involved the manipulation of five variables that often serve to reduce MMR power: (a) predictor variable range restriction, (b) total sample size, (c) sample sizes in two subgroups associated with a dichotomous moderator variable, (d) predictor variable intercorrelation, and (e) magnitude of the moderating effect. Monte Carlo simulations were conducted in which the levels of these variables were manipulated concurrently to assess their main and interactive effects on MMR power for situations in which there is a dichotomous moderator variable. The study's dependent variable was the proportion of times that MMR revealed the presence of a moderating effect.

Manipulated Parameters

Range restriction. The degree of range restriction on a continuous predictor variable (X) was operationally defined using the following ratio: range of scores in the sample divided by range of scores in the population, hereinafter referred to as the range ratio (*RR*). *RR* took on values of 1.00, .80, .40, and .20 so as to represent situations varying from no range restriction (i.e., *RR* = 1.00) to a very severe degree of range restriction in which the range of scores in the sample is only 20% of that in the population (i.e., *RR* = .20).

Total sample size. The total size of the sample on which the MMR analyses were based was set at either 60 or 300. These values cover the range typically found in applied psychology. For instance, Lent, Aurbach, and Levin (1971) found that the median sample size in 1,500 validation studies was 68. More recently, Russell et al. (1994) conducted a meta-analysis that included the 138 validation studies of personnel selection systems published in the *Journal of Applied Psychology* and *Personnel Psychology* between 1964 and 1992 and ascertained that the median sample size was 103 (C. J. Russell, personal communication, February 21, 1996).

Proportion of cases in Subgroup 1. In a study involving a dichotomous moderator variable (Z), the total number of cases in a sample equals the sum of the number of cases in each of the two moderator variable-based subgroups, n_1 for Subgroup 1 (Z = 1) and n_2 for Subgroup 2 (Z = 2). Thus, the proportion of cases in Subgroup 1 (i.e., p_1) is $n_1 \div N$. This proportion was set at values of .1, .3, and .5 so as to represent reasonably the range of proportions that are possible in any study. Moreover, note that these proportions are typical in studies that test for the effects of dichotomous moderator variables in personnel psychology (cf. Hunter, Schmidt, & Hunter, 1979; Trattner & O'Leary, 1980).

Correlation between predictor variable and moderator variable (predictor intercorrelation). The correlation between the predictor variable (X) and the moderator variable (Z) (i.e., ρ_{XZ}) was set at levels of .20, .40, and .80. These values were chosen so as to include what might be regarded as a low, medium, and high degrees of multicollinearity between the predictors.

Magnitude of moderating effect. As in recent simulations (e.g., Stone-Romero & Anderson, 1994; Stone-Romero et al., 1994), the magnitude of the moderating effect in the population (referred to hereinafter as *effect size; ES*) was operationally defined as the absolute difference between the ρ_{YX} levels for the two moderator variable-based subgroups (i.e., $|\rho_{YX(1)} - \rho_{YX(2)}|$). The size of this difference was set at levels of .00 (i.e., no effect), .20 (i.e., small effect), .40 (i.e., medium effect), and .60 (i.e., large effect). Converted to Cohen's (1988, pp. 410–412) effect size metric f^2 (see also Aiken & West, 1991, p. 157), these values correspond to f^2 levels of .000, .010, .075, and .145, respectively. Note that Cohen (1988) uses the labels small, medium, and large for f^2s of .020, .150, and .350. However, because the present simulation included parameter values that were assumed to mimic realistically those encountered by applied psychologists concerned with detecting the effects of dichotomous moderators, we did not include *ES* values larger than .145.

Summary. Table 1 provides a summary of the parameters that were manipulated in the simulation and the levels taken on by each such parameter. The manipulations of the independent variables led to a design having a total of 648 cells or conditions (i.e., 4 (*RR*) × 2 (N) × 3 (p_1) × 3 (ρ_{XZ}) × 3 ($\rho_{YX(1)}$) × 3 ($\rho_{YX(2)}$)).

Simulation Procedure and Dependent Variable

Simulation procedure. The simulations were conducted on 50 Mhz IBM-compatible personal computers with QuickBASIC

Table 1
Values of the Manipulated Parameters

Manipulated parameter	Values assumed by parameter in the simulation			
	Level 1	Level 2	Level 3	Level 4
RR	.2	.4	.8	1.0
N	60	300		
p_1	.1	.3	.5	
ρ_{XZ}	.2	.4	.8	
$\rho_{YX(1)}$	.2	.4	.8	
$\rho_{YX(2)}$	.2	.4	.8	

Note. RR = range ratio on predictor variable X; p_1 = proportion of cases in Subgroup 1 (i.e., $n_1 \div N$); ρ_{XZ} = correlation between predictor variables X and Z; $\rho_{YX(1)}$ = correlation between Y and X for Subgroup 1 (i.e., Z = 1); $\rho_{YX(2)}$ = correlation between Y and X for Subgroup 2 (i.e., Z = 2).

4.5 software programs. Taken together, these programs have been shown to be accurate in (a) generating variables with specified properties (e.g., normally distributed and with specified intercorrelations), and (b) testing for moderating effects through assessing the statistical significance of the unstandardized regression coefficient of the $X \cdot Z$ term (i.e., null hypothesis of $\beta_{X \cdot Z} = 0$). The Appendix includes a description of the programs used, as well as the results of checks on their accuracy.

A total of 1,000 samples was drawn for each of the 648 conditions in the simulation. For each such sample, we used MMR to test for the existence of a moderating effect. The null hypothesis of no moderating effect was tested using a two-tailed test and a nominal Type I error rate (α) of .05.

Dependent variable. For each of the 648 cells of the design, we determined the proportion of times which the MMR analysis revealed the presence of a moderating effect (i.e., by rejecting the null hypothesis of $\beta_{X \cdot Z} = 0$). This proportion is equivalent to statistical power in cases in which there actually is an effect in the population, that is, the proportion of times which a false null hypothesis has been rejected. Note, however, that when there is no moderating effect in the population (i.e., $\rho_{YX(1)} = \rho_{YX(2)}$, $ES = 0$) the proportion of times that the null hypothesis is (falsely) rejected represents Type I error. Thus, for the cells in which $\rho_{YX(1)} = \rho_{YX(2)}$, the rejection rate should be .05 (i.e., the preset nominal Type I error rate).

Results

Rejection rates for the null hypothesis of no moderating effect of Z are reported in Tables 2 and 3 for all the cells in the design. In view of the large number of power estimates that are reported in these tables, we used a general linear modeling (GLM) strategy to simplify the description of the study's results concerning the main and interactive effects of the manipulated variables on MMR power.

Regression of Power on Variables Manipulated in the Simulation

Prior to regressing the empirically obtained MMR rejection rates on values of the manipulated parameters, we

Table 2
Power To Detect Moderating Effect and Type I Error Rates as a Function of Range Restriction on X, *Magnitude of the Moderating Effect in the Population, and Predictor Variable Intercorrelation for* N = *60*

			Range ratio											
			p_1 = .1				p_1 = .3				p_1 = .5			
$\rho_{YX(1)}$	$\rho_{YX(2)}$	ρ_{XZ}	1.00	.80	.40	.20	1.00	.80	.40	.20	1.00	.80	.40	.20
.2	*.2*	*.2*	*.032*	*.052*	*.047*	*.041*	*.047*	*.060*	*.047*	*.049*	*.047*	*.047*	*.063*	*.045*
.2	*.2*	*.4*	*.050*	*.050*	*.055*	*.047*	*.051*	*.056*	*.043*	*.048*	*.055*	*.039*	*.060*	*.049*
.2	*.2*	*.8*	*.058*	*.050*	*.036*	*.052*	*.053*	*.042*	*.047*	*.065*	*.048*	*.039*	*.050*	*.050*
.2	.4	.2	.079	.060	.055	.073	.121	.102	.074	.078	**.110**	.085	.085	.078
.2	.4	.4	.102	.078	.070	.060	.122	.093	.094	.062	**.127**	.098	.081	.068
.2	.4	.8	.088	.079	.063	.065	.112	.093	.090	.089	**.131**	.090	.083	.079
.2	.8	.2	.492	.369	.311	.249	.734	.592	.406	.310	**.749**	.562	.352	.287
.2	.8	.4	.496	.395	.315	.272	.740	.567	.388	.339	**.788**	.592	.411	.311
.2	.8	.8	.492	.380	.304	.252	.717	.591	.445	.342	**.739**	.614	.462	.386
.4	.2	.2	.067	.046	.055	.052	.114	.074	.050	.044	**.130**	.098	.062	.060
.4	.2	.4	.054	.062	.044	.035	.093	.078	.064	.055	**.139**	.119	.075	.058
.4	.2	.8	.056	.052	.042	.039	.097	.094	.064	.064	**.123**	.102	.081	.079
.4	*.4*	*.2*	*.055*	*.054*	*.053*	*.040*	*.056*	*.058*	*.056*	*.044*	*.043*	*.061*	*.040*	*.058*
.4	*.4*	*.4*	*.044*	*.052*	*.046*	*.053*	*.053*	*.051*	*.048*	*.043*	*.065*	*.045*	*.040*	*.056*
.4	*.4*	*.8*	*.048*	*.049*	*.053*	*.053*	*.043*	*.051*	*.056*	*.051*	*.050*	*.061*	*.046*	*.048*
.4	.8	.2	.332	.280	.208	.224	.489	.338	.252	.224	**.477**	.317	.216	.141
.4	.8	.4	.347	.281	.235	.220	.490	.350	.236	.230	**.492**	.362	.238	.164
.4	.8	.8	.344	.278	.255	.198	.476	.361	.272	.245	**.449**	.367	.286	.246
.8	.2	.2	.195	.092	.049	.022	.631	.400	.224	.147	**.764**	.578	.365	.278
.8	.2	.4	.198	.101	.046	.022	.644	.411	.214	.150	**.734**	.604	.357	.300
.8	.2	.8	.173	.088	.036	.021	.586	.426	.244	.144	**.743**	.661	.472	.357
.8	.4	.2	.074	.045	.023	.015	.318	.209	.102	.068	**.465**	.336	.186	.144
.8	.4	.4	.081	.044	.028	.021	.343	.210	.121	.099	**.482**	.340	.176	.151
.8	.4	.8	.079	.038	.026	.016	.322	.217	.111	.102	**.472**	.358	.251	.159
.8	*.8*	*.2*	*.059*	*.047*	*.047*	*.038*	*.046*	*.060*	*.048*	*.036*	*.048*	*.042*	*.053*	*.047*
.8	*.8*	*.4*	*.046*	*.048*	*.047*	*.055*	*.048*	*.055*	*.047*	*.045*	*.047*	*.038*	*.044*	*.057*
.8	*.8*	*.8*	*.043*	*.055*	*.055*	*.042*	*.041*	*.041*	*.050*	*.040*	*.053*	*.060*	*.052*	*.044*

Note. $\rho_{YX(1)} = \rho_{YX}$ for $Z = 1$, $\rho_{YX(2)} = \rho_{YX}$ for $Z = 2$, p_1 = proportion of scores in Subgroup 1 (i.e., $n_1 \div N$). Cells showing Type I error rates are italicized (i.e., $ES = 0$). Cells showing power rates for no range restriction and equal number of scores across subgroups are boldface and may be used for comparison (i.e., baseline) purposes.

conducted the following two types of transformations. First, as recommended by Aiken and West (1991), we centered all the predictors so as to ease the interpretation of the resulting regression coefficients. Second, because the MMR power function is nonlinear (cf. McClelland & Judd, 1993), GLM-based regression coefficients may be affected adversely (i.e., biased) by scaling artifacts. Hence, we linearized the power function by transforming the empirically derived power estimates shown in Tables 2 and 3, which are expressed in proportion metric, using an arcsin square root transformation (Winer, Brown, & Michels, 1991, p. 356, case *ii*). One should note, however, that Tables 2 and 3 show the untransformed rejection rates for all the conditions in the simulation, including the cells for which the moderating effect size was zero in the population.[1]

In adopting a GLM approach to modeling the data generated by the simulation, we regressed the transformed power estimates on the main and interactive effects of the manipulated parameters of (a) range restriction on predictor X (i.e., the range ratio; *RR*), (b) total sample size (N), (c) proportion of cases in Subgroup 1 (i.e., $p_1 = n_1/N$), (d) degree of predictor variable intercorrelation (i.e., Fisher's Z transformation of ρ_{XZ}), and (e) magnitude of the moderating effect in the population (i.e., the absolute value of the difference between the Fisher's Z scores of $\rho_{YX(1)}$ and $\rho_{YX(2)}$; ZD). Note that the dependent variable in these analyses was the arcsin square root of the power of MMR to detect a moderating effect (i.e., a transformation of the proportion of times that the simulation led to the rejection of a false null hypothesis of no moderating effect). Thus, this GLM analysis used data from only the 432 cells in the design for which there actually was a moderating effect (i.e., $\rho_{YX(1)} \neq \rho_{YX(2)}$).

[1] Although not the focus of the present research, a perusal of Tables 2 and 3 indicates that rejection rates for cells corresponding to no moderating effect were close to the a priori specified Type I error rate of .05. This can be observed for virtually all 216 conditions for which $\rho_{YX(1)} - \rho_{YX(2)} = .00$ (i.e., $ES = 0$). Thus, the design artifacts manipulated in the present study do not seem to have artificially increased Type I error rates.

Table 3
Power To Detect Moderating Effect and Type I Error Rates as a Function of Range Restriction on X, *Magnitude of the Moderating Effect in the Population, and Predictor Variable Intercorrelation for* N = *300*

			Range ratio											
			p_1 = .1				p_1 = .3				p_1 = .5			
$\rho_{YX(1)}$	$\rho_{YX(2)}$	ρ_{XZ}	1.00	.80	.40	.20	1.00	.80	.40	.20	1.00	.80	.40	.20
.2	*.2*	*.2*	*.055*	*.048*	*.046*	*.054*	*.063*	*.039*	*.060*	*.058*	*.047*	*.043*	*.052*	*.057*
.2	*.2*	*.4*	*.058*	*.051*	*.045*	*.060*	*.052*	*.053*	*.057*	*.057*	*.055*	*.042*	*.054*	*.047*
.2	*.2*	*.8*	*.040*	*.064*	*.046*	*.051*	*.056*	*.049*	*.053*	*.051*	*.036*	*.056*	*.052*	*.059*
.2	.4	.2	.196	.139	.102	.102	.399	.270	.177	.130	**.448**	.322	.173	.163
.2	.4	.4	.210	.160	.118	.104	.382	.255	.192	.123	**.469**	.302	.209	.152
.2	.4	.8	.196	.162	.098	.124	.394	.262	.171	.150	**.441**	.334	.251	.172
.2	.8	.2	.958	.846	.655	.568	1.000	.986	.915	.829	**1.000**	.998	.963	.874
.2	.8	.4	.958	.865	.701	.560	1.000	.994	.927	.828	**1.000**	1.000	.966	.892
.2	.8	.8	.962	.861	.676	.575	.999	.994	.936	.869	**1.000**	.999	.968	.935
.4	.2	.2	.171	.102	.074	.083	.369	.251	.152	.112	**.457**	.300	.173	.136
.4	.2	.4	.187	.120	.074	.070	.366	.277	.164	.131	**.431**	.322	.210	.147
.4	.2	.8	.184	.129	.069	.051	.358	.247	.164	.141	**.440**	.335	.210	.188
.4	*.4*	*.2*	*.046*	*.049*	*.049*	*.050*	*.047*	*.067*	*.050*	*.042*	*.052*	*.051*	*.045*	*.051*
.4	*.4*	*.4*	*.053*	*.055*	*.064*	*.057*	*.050*	*.046*	*.053*	*.042*	*.047*	*.050*	*.046*	*.061*
.4	*.4*	*.8*	*.036*	*.044*	*.047*	*.050*	*.054*	*.054*	*.047*	*.042*	*.032*	*.059*	*.044*	*.038*
.4	.8	.2	.811	.620	.436	.358	.980	.879	.677	.566	**.995**	.927	.748	.599
.4	.8	.4	.808	.637	.448	.395	.973	.906	.698	.575	**.989**	.929	.782	.651
.4	.8	.8	.808	.622	.456	.392	.981	.900	.737	.612	**.990**	.951	.810	.722
.8	.2	.2	.943	.742	.386	.263	1.000	.998	.910	.771	**1.000**	1.000	.961	.872
.8	.2	.4	.944	.777	.440	.279	1.000	.998	.935	.791	**1.000**	1.000	.972	.918
.8	.2	.8	.942	.767	.415	.262	1.000	.998	.968	.870	**1.000**	1.000	.998	.978
.8	.4	.2	.656	.381	.202	.113	.974	.883	.565	.426	**.993**	.938	.762	.603
.8	.4	.4	.657	.408	.197	.126	.981	.902	.663	.443	**.990**	.958	.807	.666
.8	.4	.8	.651	.389	.185	.113	.978	.905	.679	.522	**.995**	.966	.873	.779
.8	*.8*	*.2*	*.054*	*.044*	*.052*	*.071*	*.054*	*.045*	*.047*	*.056*	*.051*	*.049*	*.040*	*.058*
.8	*.8*	*.4*	*.055*	*.038*	*.048*	*.042*	*.044*	*.053*	*.046*	*.048*	*.051*	*.053*	*.041*	*.051*
.8	*.8*	*.8*	*.043*	*.053*	*.053*	*.050*	*.043*	*.039*	*.059*	*.041*	*.037*	*.045*	*.045*	*.065*

Note. $\rho_{YX(1)} = \rho_{YX}$ for $Z = 1$, $\rho_{YX(2)} = \rho_{YX}$ for $Z = 2$, p_1 = proportion of scores in Subgroup 1 (i.e., $n_1 \div N$). Cells showing Type I error rates are italicized (i.e., $ES = 0$). Cells showing power rates for no range restriction and equal number of scores across subgroups are boldface and may be used for comparison (i.e., baseline) purposes.

The regression analysis took place in four steps. At Step 1 of the analysis, the main effects of the five predictor variables were forced into the equation; at Step 2, the 10 two-way interaction terms were entered; at Step 3, the three-way interaction terms were entered; at Step 4, the four-way interactions were entered; and, finally, at Step 5, the five-way interaction term was entered. However, the block of three-way interaction terms accounted for .3% of the variance in MMR power above and beyond the main and two-way interaction effects, the four-way interaction terms accounted for .2%, and the five-way interaction accounted for 0%. As a result, regression coefficients for terms higher than the second order are neither reported in Table 4 nor discussed. The regression coefficients presented in Table 4 are for the stage at which all main and two-way product terms were in the model.

In describing the results in Table 4, we first consider the interaction effects. The reason for this is that the existence of interaction effects implies that the main effects of variables that enter into the interaction represent the average of the effects of the same variables across relevant levels of the other variables (Darlington, 1990; Jaccard et al., 1990). The interpretation of interactions yields the most detailed and precise information regarding the impact of the manipulated parameters on power for specific conditions. Alternatively, the interpretation of main effects provides more generalizable, but less precise information.

To facilitate the interpretation of the interaction effects reported in Table 4, we make reference to representative results found in Tables 2 and 3. In addition, to ease the interpretation of the results, we graphically display empirically derived power rates in Figures 1 to 5.

Interactive Effects

Table 4 shows that the block of two-way interactions accounted for 16.9% of the variance in MMR power above and beyond the variance accounted for by the main effects. A perusal of Table 4 indicates that there were interaction effects for (a) range ratio by moderating effect magnitude, (b) range ratio by total sample size, (c) total sample size by moderating effect magnitude, (d) moderat-

Table 4
Moderated Multiple Regression Analysis: Regression of Power on Manipulated Parameters

Step and variable	*b*	*SE b*	*B*	*F*	R^2	ΔR^2
1. Main effects					.779**	.779**
RR	0.5338	.0230	.210	538.56**		
N	0.0025	.0000	.371	1,674.63**		
p_1	0.9708	.0445	.197	475.00**		
$Z\rho_{XZ}$	0.0317	.0191	.015	2.76		
ZD	1.6527	.0200	.747	6,809.03**		
2. Two-way interactions					.948**	.169**
$RR \times ZD$	1.212	.0633	.173	366.26**		
$RR \times N$	0.001	.0002	.057	39.41**		
$RR \times Z\rho_{XZ}$	−0.102	.0603	−.015	2.87		
$RR \times p_1$	0.166	.1409	.011	1.39		
$N \times ZD$	0.006	.0002	.321	1.253.58**		
$ZD \times p_1$	2.300	.1227	.170	351.70**		
$ZD \times Z\rho_{XZ}$	0.087	.0526	.015	2.74		
$N \times p_1$	0.002	.0004	.058	41.45**		
$N \times Z\rho_{XZ}$	0.000	.0002	.004	0.67		
$Z\rho_{XZ} \times p_1$	0.178	.1169	.014	2.32		
3. Three-way interactions					.951**	.003**
4. Four-way interactions					.953**	.002**
5. Five-way interaction					.953**	.000

Note. b = unstandardized regression coefficient; B = standardized regression coefficient; RR = range ratio on predictor variable X; p_1 = proportion of scores in Subgroup 1 (i.e., $n_1 \div N$); $Z\rho_{XZ}$ = Fisher's Z transformation of predictor intercorrelation (ρ_{XZ}); and ZD = moderating effect magnitude (absolute difference between Fisher's Z scores for $\rho_{YX(1)}$ and $\rho_{YX(2)}$). The regression coefficients shown are for the stage at which all main and two-way product terms are in the model (intercept = 1.082). Three-way, four-way, and five-way interactions accounted for less than 1% of variance above and beyond main and two-way interaction effects on power. Consequently, they were not included in the model.
** $p < .001$.

ing effect magnitude by subgroup proportion, and (e) subgroup proportion by total sample size. Next, we describe and graphically illustrate each of these interactions.

Range Ratio × Moderating Effect Magnitude Interaction. The range ratio (RR) by moderating effect magnitude (ZD) interaction depicted in Figure 1 shows that the

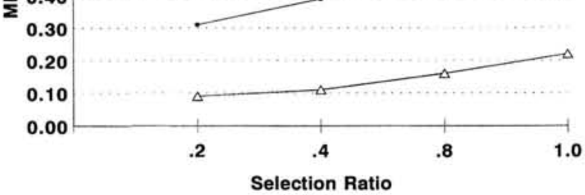

Figure 1. Interactive effects of RR (range ratio on X) and ZD (magnitude of the moderating effect) on moderated multiple regression (MMR) power. ES = effect size.

predictor variable range restriction had a greater impact in decreasing power when the magnitude of the moderating effect in the population was larger. For instance, when the effect size was large, no range restriction (i.e., $RR = 1.0$) yielded an acceptable MMR power of .79, whereas a range ratio of .20 decreased MMR power to .48. However, when the effect size magnitude was small, power was already low when there was no range restriction (i.e., MMR power = .22), and the introduction of range restriction (e.g., RR = .20) lowered the rejection rates, but the relative impact was smaller because the values were already near the chance level (i.e., .05).

Range Ratio × Total Sample Size interaction. The results reported in Tables 2 and 3 and those displayed in Figure 2 provide a basis for understanding the nature of the range ratio (RR) by total sample size interaction. The range ratio has a different impact on power at total sample size levels of 60 and 300: For N = 60, MMR power is low overall (.35 or lower). As a consequence, downward shifts in RR do not lead to marked decreases in power. However, for N = 300, there is a considerable decrease in power as RR shifts from 1.0 to .20, that is, MMR power decreases from .74 to .44.

Total Sample Size × Moderating Effect Magnitude interaction. The results presented in Figure 3 show that, regardless of the level of the moderating effect magnitude

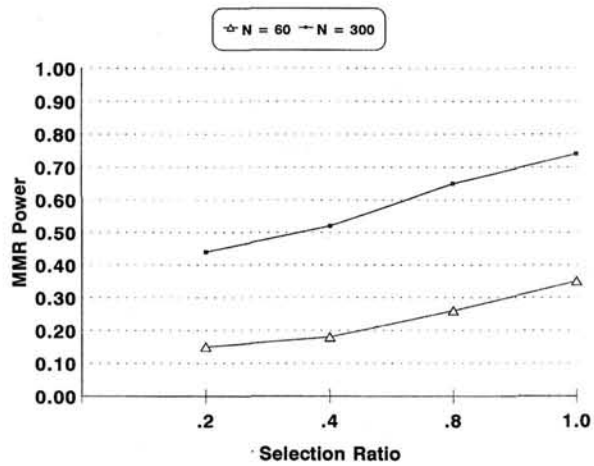

Figure 2. Interactive effects of RR (range ratio on X) and total sample size on moderated multiple regression (MMR) power.

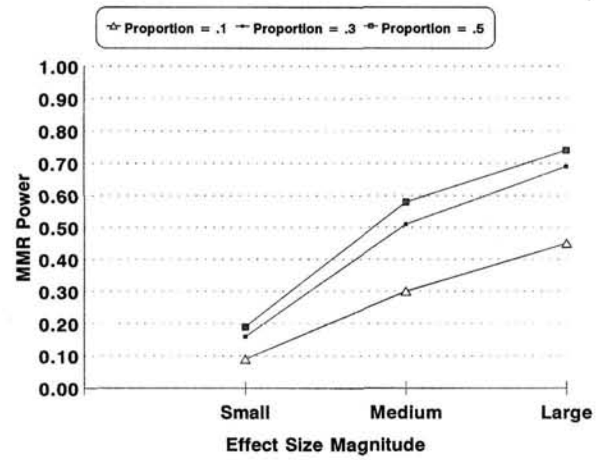

Figure 4. Interactive effects of ZD (magnitude of moderating effect) and p_1 (proportion of cases in Subgroup 1) on moderated multiple regression (MMR) power.

(ZD), statistical power is greater for a total sample size of 300 than it is for $N = 60$. It should be noted, however, that changes in ZD have a linear impact on MMR power when $N = 60$, but this is not the case when $N = 300$. For example, if a sample of $N = 300$ is used, shifting from a small to a medium moderating effect increases MMR power from .21 to .69, whereas using a sample of $N = 60$, the same change in the moderating effect magnitude increases MMR power to a much smaller degree (i.e., from .08 to .24).

Moderating Effect Magnitude × Subgroup Proportion interaction. The results in Figure 4 show that across all values of moderating effect magnitude (ZD), when subgroup proportion (p_1) = .1 power is uniformly lower than when $p_1 = .3$ or .5. Also, a proportion of .3 does not have such a detrimental effect on power as a proportion of .1. For instance, given a large moderating effect, a proportion of .5 yielded MMR power of .69, whereas a proportion of .3 resulted in a comparable MMR power of .74.

Subgroup Proportion × Total Sample Size interaction. Figure 5 displays how unequal subgroup proportions and total sample size interactively influence MMR power. This figure shows that for a total sample size of 60, an increase in subgroup proportion from .1 to .5 led to only a modest increase in power (from .15 to .31). Also, when the total sample size was 300, a number that would typically be considered adequate in applied psychology, a shift in subgroup proportion from .1 to .5 led to an MMR power

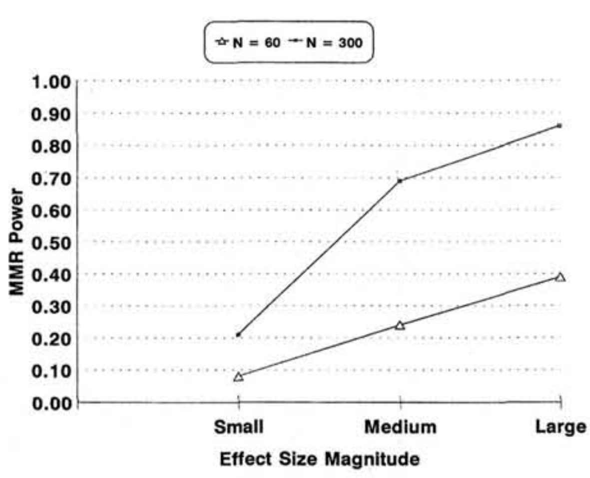

Figure 3. Interactive effects of ZD (magnitude of moderating effect) and total sample size on moderated multiple regression (MMR) power.

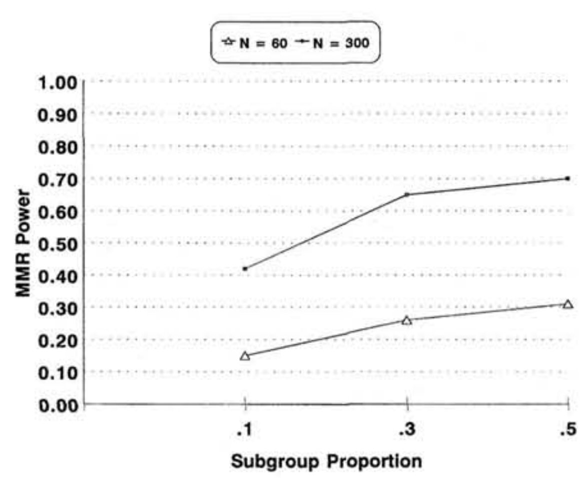

Figure 5. Interactive effects of p_1 (proportion of cases in Subgroup 1) and total sample size on moderated multiple regression (MMR) power.

change from .42 to .70. This illustrates that even in the presence of a somewhat large total sample size, if subgroup sample sizes are unequal, MMR power may be inadequate to detect some moderating effects.

Main Effects

The results reported in Table 4 indicate that four of the five manipulated parameters affected power. Not surprisingly, our results showed that MMR power increased with increases in total sample size and increases in moderating effect size.

More interestingly, MMR power dropped as the range ratio decreased. For example, Table 2 shows that in the presence of a large moderating effect, predictor intercorrelation of .40, equal proportions across groups (i.e., p_1 = .5), and total sample size of 60, the absence of range restriction (i.e., RR = 1.00) yielded an acceptable MMR power of .734. However, for the same conditions (also shown on line 20), a range ratio of .80 decreased power to .604, a range ratio of .40 lowered it further to .357, and an even lower range ratio of .20 led to MMR power of .30.

Finally, replicating results reported by Stone-Romero et al. (1994), MMR power decreased to the extent that the proportion of cases in moderator variable-based subgroups deviated from .5. For instance, a perusal of Table 3 (i.e., N = 300) shows that there are drastic decreases in MMR power when subgroup proportion decreases from .5 to .1. However, there is a more marked decrease when this proportion shifts from .3 to .1 as compared with the .5 to .3 change.

Discussion

The overall purpose of this study was to examine the power of MMR to detect moderating effects under a number of conditions that are common in many empirical studies in applied psychology. We assessed the main and interactive effects of predictor variable range restriction (operationally defined as the range ratio; RR), total sample size, sample sizes in two subgroups associated with a dichotomous moderator variable (i.e., p_1), predictor variable intercorrelation (i.e., ρ_{XZ}), and magnitude of the moderating effect (i.e., ZD) on MMR power. In the paragraphs that follow, we detail the implications of our findings.

Main Effect of Predictor Variable Range Restriction

The implications regarding predictor variable range restriction are particularly interesting because no previous research had empirically examined the effects of this artifact on the power of MMR to detect dichotomous moderating effects. The present results indicate that decreases in the range ratio lead to marked decreases in power. In general, across all levels of the other manipulated variables, as the range ratio decreased, so did MMR power. Thus, even when the other conditions that are critical for the detection of moderating effects are optimal, range restriction leads to substantial decreases in power. For example, for a medium moderating effect, subgroup proportion of .5, total sample size of 300, and no range restriction (i.e., RR = 1.00), MMR power is .99. However, for the same optimal (and rare) levels of moderating effect magnitude, subgroup proportion, and total sample size, decreases in RR to .80, .40, and .20 reduce power to .94, .80, and .67, respectively.

The impact of range restriction is even more noticeable in situations more typically encountered in applied psychological research (e.g., personnel selection). For example, if the moderating effect magnitude is small, subgroup proportion is .1, and total sample size is 300, MMR power equals .73 when there is no range restriction (i.e., RR = 1.00). As RR decreases to .80, .40, and .20, MMR power decreases to .51, .32, and .25, respectively. Thus, the effects of range restriction on MMR power are considerable and, consequently, predictor range restriction is a factor that cannot be overlooked in future efforts to estimate the effects of moderator variables using MMR.

Main Effects of Additional Manipulated Variables: Replication and Extension of Previous Research

Several previous studies investigating continuous predictor variables and continuous moderator variables have used Monte Carlo simulation methods to assess the impact of such variables as sample size and effect size on MMR power (e.g., Dunlap & Kemery, 1988; Evans, 1985; Stone-Romero & Anderson, 1994; Stone-Romero et al., 1994). Unlike most such studies, the present research was concerned with the effects of such variables on MMR power in situations in which there is a continuous predictor variable and a dichotomous moderator. The results of our research are largely consistent with those of studies that investigated continuous moderator variables: Small sample size, small moderating effect magnitude, and departures from sample size equality across moderator-based subgroups have a detrimental impact on MMR power.

In addition, our study's results add to the extant knowledge base concerned with the detection of moderating effects using MMR by providing empirically derived power rates for a large number of conditions typically encountered by MMR users. Trattner and O'Leary (1980) provided tables showing sample sizes needed to achieve a statistical power level of .80 in tests of differences in validity coefficients across subgroups. However, the Trattner and O'Leary tables do not consider the impact of, for example, range restriction. Thus, researchers using these tables in situations of range-restricted samples may overestimate the probability of detecting a population moderating effect. Overcoming this likely problem, Tables 2 and 3 allow for estimating the power of MMR analysis under various conditions of range restriction, total sample size, sample sizes across moderator-based subgroups, predictor intercorrelation, and moderating effect magnitude. Consequently, researchers can use the information presented in these tables to estimate power levels a priori in designing research in which moderated regression will be used to assess moderating effects. Moreover, the information shown in Tables 2 and 3 may be used to assess power levels on a post hoc basis in studies that have been conducted without an a priori power analysis. If the resulting power estimate is low, researchers should interpret null findings with caution and not easily dismiss the existence of a hypothesized moderating effect.

Admittedly, the range of values in Tables 2 and 3 may not cover all possible situations encountered by applied psychologists. However, because they consider the impact of additional artifacts, they provide a definite improvement over the tables available at present (i.e., Trattner & O'Leary, 1980).

Interactive Effects of Manipulated Variables

Our study's results augment existing knowledge concerning the detection of moderating effects in yet a third regard. In several previous simulation studies, such variables as sample size, effect size, and predictor variable intercorrelation were manipulated in isolation or in pairs. By contrast, our study examined the interactive effects of five independent variables on MMR power. A noteworthy finding of our study is that the manipulated variables have interactive (i.e., nonadditive) effects on MMR power. The block of two-way interactions accounted for 16.9% of the variance in MMR power above and beyond the effects of the main effects. Converting this value to Cohen's (1988) metric yields an effect size of $f^2 = 3.25$ (see Aiken & West, 1991, p. 157 or Cohen & Cohen, 1983, p. 161, for the equation used to compute f^2). It should be noted that Cohen's (1988) reviews as well as comprehensive meta-analytic reviews in most areas of the literature in social sciences, education, and business indicate that an effect size of $f^2 = .350$ should be considered large (Aiken & West, 1991; see Chapter 8). The magnitude of the interactive effects shown in Table 4 is over nine times larger than what is conventionally considered to be a large effect. The practical implication of such sizable interactive effects is that even if the conditions in research designed to detect moderating effects are very favorable from the standpoint of one or more of the variables that determine MMR power (e.g., large total sample size), the presence of an unfavorable level of one or more of the other variables that influence power (e.g., low subgroup proportion and range ratio levels) will result in power levels that are far below the .80 standard recommended by Cohen (1988). Next, we detail and illustrate the implications of these results for the detection of moderating effects in applied psychology.

Failures of previous studies to detect moderating effects. It is interesting to consider our results regarding the interactive effects of various artifacts on MMR power within the context of recent research that estimated the effects of dichotomous moderator variables using MMR (e.g., Cortina et al., 1992; Hattrup & Schmitt, 1990; Schmitt et al., 1993; Wohlgemuth & Betz, 1991). For instance, Hattrup and Schmitt (1990) failed to find a moderating effect of gender on the relationship between test scores and job performance in a study in which there were marked differences in the sample sizes of the male and female subgroups, resulting in a very low proportion of cases in the female subgroup. Their MMR analysis led to the conclusion that the criterion-related validity of the test did not differ across the two subgroups. However, results of the present study as well as those of a simulation by Stone-Romero et al. (1994) suggest that, largely because of the differences in the subgroup proportions, the power to detect a gender-based moderating effect in the Hattrup and Schmitt study was approximately .25. Assuming that such research as Cortina et al. (1992), Hattrup and Schmitt (1990), Schmitt et al. (1993), and Wohlgemuth and Betz (1991) is representative of a larger body of research in which attempts to detect moderating effects have failed, the present study's results suggest that the power of many studies that have tested for moderating effects using MMR may have been inadequate. That is, because large differences in subgroup proportions seem to have been largely ignored in previous research, many moderating effects have probably gone undetected.

In addition to having power problems resulting from unequal proportions of cases in the moderator variable-based subgroups, the Hattrup and Schmitt (1990) study also suffered from the problem of "explicit selection on the predictor tests for males" (p. 460). Considering this fact in conjunction with the results of our simulation regarding the effects of range restriction on MMR power suggests that the power of the Hattrup and Schmitt study to detect a potential gender-based moderating effect was likely considerably below .25. More precisely, the present study's results suggest that the probability that the same moderating effect would be detected in the Hattrup and Schmitt study was very close to .05. We emphasize that the null findings reported by Hattrup and Schmitt represent only one of the many examples that could have been used here to illustrate how study-related artifacts such as those considered by our simulation can serve to reduce statistical power to unacceptably low levels. Low power seems to be a pervasive problem in research aimed at the estimation of moderating effects, and we do not intend to devalue Hattrup and Schmitt's otherwise excellent study (see DeShon & Alexander, 1994, for additional examples).

Overall, what the results of our study suggest is that for the combined levels of range restriction, sample size, and so forth that are common in research in industrial and organizational psychology (DeShon & Alexander, 1994), educational psychology (Linn, 1983), and in other fields, it is likely that the power to detect moderating effects using MMR is unacceptably low (i.e., far below the .80 level recommended by Cohen, 1988). Also, the combination of such problems as low sample size, range restriction, and departures of subgroup proportions from .5 can lead to power levels that approach mere chance levels (i.e., .05).

Finally, it appears that the results of our simulation provide a very reasonable explanation of the inability of many previous studies to detect moderating effects, even though the attempts to detect such effects were often guided by sound, theory-based predictions. Hence, given the present findings, it should come as no surprise that moderator variables have been viewed as being elusive (e.g., Zedeck, 1971) and that moderating effects found in any single study have proven difficult or impossible to replicate (Smith & Sechrest, 1991).

Inability of selective strategies to solve power problems. Aguinis (1995) proposed several possible strategies for overcoming low power situations in personnel psychology and human resources management research. For example, one such strategy is the well-known and often difficult to implement increase in total sample size. However, another implication of the results regarding the interactive effects of the manipulated variables is that, in any given study, attempts to enhance the probability of detecting moderating effects that rely on the use of a single strategy (e.g., increasing total sample size) may not be successful. Thus, in research aimed at estimating moderating effects, our results highlight the importance of fully considering a wide range of design issues prior to the point when data are collected. If samples are too small, subgroup proportions are unequal, predictor variables suffer from range restriction, and so forth, existing moderating effects are very likely to go undetected.

Potential Limitations of the Present Study

The purpose of this study was to assess the simultaneous impact of five research design-related artifacts on the power of MMR to detect dichotomous moderating effects. In designing and conducting the simulation, practical considerations led to choices concerning (a) the nature of the data generated by the simulation, and (b) the values assumed by the manipulated parameters.

Regarding the nature of the generated variables, it should be acknowledged that this simulation is limited by (a) the use of the correlation model in which variances are equal in each group (i.e., X and Y were generated as normal deviates), and (b) the assumption that selection takes place in a top-down fashion. Regarding Point (a), the ratio of Y to X variance was held constant across the two moderator-based subgroups. However, as Alexander and DeShon's (1994) recent simulation demonstrated, the presence of a moderating effect of a dichotomous variable causes a violation of the assumption of homogeneity of error variances. The typical effect of this violation is a decrease in power to detect a moderating effect, especially in the presence of (a) unequal sample sizes across moderator-based subgroups, and (b) small total sample size (see Aguinis & Pierce, 1996, for a review). Interestingly, as noted by Alexander and DeShon, the violation of the homogeneity of error variance assumption is virtually guaranteed given that (a) the Y to X variance ratio is equal across subgroups, and (b) a hypothesized moderating effect exists. This has implications for the results of our study. Specifically, because our simulation mimicked actual situations in which MMR is used to detect the effects of dichotomous moderators using MMR, it also emulated the systematic violation of this assumption. The net effect of this violation would be to reduce MMR power. However, we do not regard this as an important threat to the validity of our conclusions. The reason for this is that the generated power estimates were derived under conditions of heterogeneity of error variance that parallel those that exist in empirical research aimed at the detection of dichotomous moderator variables (cf. DeShon & Alexander, 1994).

Regarding Point (b), the range restriction manipulation was implemented in a top-down fashion. Admittedly, selection may not always be performed in a strict top-down fashion. For example, some top-ranked applicants may not accept a job offer (Murphy, 1986), or personnel specialists may use a banding approach to selection (Aguinis, Cortina, & Goldberg, 1997; Cascio, Outtz, Zedeck, & Goldstein, 1991). Nevertheless, our choice was guided by two considerations. First, strict top-down selection is the typical procedure mimicked in Monte Carlo studies (e.g., Aguinis & Whitehead, 1997; Millsap, 1989). Second, although selection may not be implemented in a strict top-down fashion, selection systems are typically structured to use top-down selection. Admittedly, the obtained MMR power rates are valid for top-down selection situations and departures from this situation may affect the precision of specific power rates shown in Tables 2 and 3. However, the overall conclusions of our research remain unchanged: MMR power is influenced by the main and interactive effects of such variables as predictor variable range restriction, sample sizes across moderator-based subgroups, and unequal proportions across moderator-based subgroups.

As regards the manipulation of parameter values, we had to make certain choices because the design of the simulation required generating 1,000 samples for each of the cells of the design. Even with the manipulation of only five variables, our study's design involved 648 separate conditions. Performing the simulations required a very large number of calculations and a substantial amount of computing time. As a consequence of limits in the range of values of the parameters used in our study, some of the manipulated variables did not cover the full range of possible values. Thus, questions might be raised regarding the construct validity of our simulation because of what Cook and Campbell (1979) referred to as the confounding of constructs with levels of constructs. For example, our study used only two N values (i.e., 60 and 300). Consequently, our results regarding the effects of N on MMR power cannot be generalized across the full range of values that N might assume in a given empirical study. Although we recognize that our choice of values for manipulated variables may somewhat limit the generalizability of our findings, we do not regard this as a serious threat to our study's conclusions. The reason for this is that, in planning our simulation, we chose values for the manipulated parameters that were consistent with those that are likely to be found in many research contexts (e.g., validation research in personnel selection). Thus, we believe that our results can be generalized to numerous research contexts in which research and practitioner psychologists use MMR to test for the effects of dichotomous moderator variables. In addition, they can be used to anticipate the impact of the manipulated artifacts on MMR power.

Research Needs

There are at least two issues that remain unresolved and deserve future research attention. First, the present simulations used the same range ratio values for both moderator-based subgroups. In the absence of a moderating effect in the population, differential degrees of range restriction across subgroups may result in an artificial increase in Type I error (Jaccard et al., 1990). However, the degree to which Type I error will be artificially increased needs to be examined empirically. Alternatively, in the presence of a moderating effect in the population, differential range restriction may either increase the Type I error rate or yield lower statistical power as compared with an equal degree of range restriction across subgroups, depending on (a) whether the relationship between X and Y is stronger in one subgroup than in the other, and (b) whether the lowest range ratio occurs in the subgroup having the strongest predictor-criterion relationship (cf. Alexander & DeShon, 1994). Therefore, additional simulations are needed to investigate the degree to which either α is artificially increased or power is reduced by differences in subgroup proportions and differences in effect sizes.

Second, our study used statistical power as the dependent variable of interest. The choice of power as the dependent variable was guided by the frequently reported failure to detect hypothesized moderating effects and the need to investigate factors that affect power (Cronbach, 1987; Smith & Sechrest, 1991). An alternative approach would be to focus on Type I error rates and to investigate other artifacts under which MMR might lead researchers to conclude that there are moderating effects when, in reality, there are no such effects on the population. Although some investigators have argued that research from this perspective is necessary (e.g., Dunlap & Kemery, 1988), the manipulated variables in the present simulations did not lead to undue increases in Type I error rates (see Footnote 1). Nevertheless, as previously stated, future research should investigate the extent to which α is unduly increased in the presence of differential predictor variable range restriction.

A Closing Comment

The results of this study showed that MMR power is influenced by the main and interactive effects of such variables as predictor variable range restriction, sample sizes across moderator-based subgroups, and unequal proportions across moderator-based subgroups. Consequently, we urge researchers to be more sensitive to these methodological artifacts in studies aimed at estimating moderating (i.e., interaction) effects using MMR. Unless they are, it seems highly likely that many existing moderating effects will go undetected in future studies. As a result, researchers may erroneously conclude that theories hypothesizing the interactive effects of independent variables are invalid and practitioners may inappropriately use personnel selection tests that predict performance differentially for various subgroups.

References

Aguinis, H. (1994). A QuickBASIC program for generating correlated multivariate random normal scores. *Educational and Psychological Measurement, 54,* 687–689.

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

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., Cortina, J. M., & Goldberg, E. (1997). *A new procedure for computing equivalence bands in personnel selection.* Manuscript submitted for publication.

Aguinis, H., & Pierce, C. A. (1996, April). Heterogeneity of error variance and differential prediction: Clarifications, implications, and solutions. In M. J. Burke (Chair), *Quantitative advances in personnel selection research and practice.* Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

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., & Whitehead, R. (1997). *Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization.* Manuscript submitted for publication.

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.

Alliger, G. M. (1992). Generating correlated bivariate random normal standard scores in QuickBASIC. *Educational and Psychological Measurement, 52,* 107–108.

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated regression strategy: An alternative to differential analysis. *Personnel Psychology, 31,* 223–241.

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

Box, G. E. P., & Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics, 29,* 610–613.

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.

Callender, J. C., & Osburn, H. G. (1988). Unbiased estimation of sampling variance of correlations. *Journal of Applied Psychology, 73,* 312–315.

Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance, 4,* 233–264.

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: Erlbaum.

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

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

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

Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The "Big Five" personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology, 45,* 119–140.

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

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

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

DeShon, R. P., & Alexander, R. A. (1994, April). *Power of the test for regression slope differences in differential prediction research.* Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Nashville, TN.

Dunbar, S. B., & Novick, M. R. (1988). On predicting success in training for men and women: Examples from Marine Corps clerical specialties. *Journal of Applied Psychology, 73,* 545–550.

Dunlap, W. P., & Kemery, E. R. (1988). Effects of predictor intercorrelations and reliabilities on moderated multiple regression. *Organizational Behavior and Human Decision Processes, 41,* 248–258.

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

Guilford, J. P., & Fruchter, B. (1973). *Fundamental statistics in psychology and education* (5th ed.). New York: McGraw-Hill.

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

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

Hays, W. L. (1988). *Statistics* (2nd ed.). Orlando, FL: Holt, Rinehart & Winston.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86,* 721–735.

Jaccard, J. J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression* (University Paper Series on Quantitative Applications in the Social Sciences, Report No. 07-072). Newbury Park, CA: Sage.

Kelley, T. L. (1923). *Statistical method.* New York: Macmillan.

Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971). Research design and validity assessment. *Personnel Psychology, 24,* 247–274.

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

Linn, R. L. (1983). Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement, 20,* 1–15.

Linn, R. L. (1994). Fair test use: Research and policy. In M. G.

Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 363–375). Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison Wesley.

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

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

Millsap, R. E. (1989). Sampling variance in the correlation coefficient under range restriction: A Monte Carlo study. *Journal of Applied Psychology, 74,* 456–461.

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.

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

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

Otis, A. S. (1922). A method of inferring the change in a coefficient of correlation resulting from a change in the heterogeneity of the group. *Journal of Educational Psychology, 13,* 293–294.

Pearson, K. (1903). Mathematical contributions to the theory of evolution: Part 11. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society, Series A, 200,* 1–66.

Pierce, J. L., Gardner, D. G., Dunham, R. B., Cummings, L. L. (1993). Moderation by organization-based self-esteem of role condition-employee response relationships. *Academy of Management Journal, 36,* 271–288.

Russell, C. J., Settoon, R. P., McGrath, R. N., Blanton, A. E., Kidwell, R. E., Lohrke, F. T., Scifres, E. L., & Danforth, G. W. (1994). Investigator characteristics as moderators of personnel selection research: A meta-analysis. *Journal of Applied Psychology, 79,* 163–170.

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

Sagie, A., & Koslowsky, M. (1993). Detecting moderator variables with meta-analysis: An evaluation and comparison of techniques. *Personnel Psychology, 46,* 629–640.

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

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsh, H. R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology, 38,* 697–798.

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

Smith, B., & Sechrest, L. (1991). Treatment of Aptitude × Treatment interactions. *Journal of Consulting and Clinical Psychology, 59,* 233–244.

Society for Industrial and Organizational Psychology. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.

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

Stone, E. F., & Hollenbeck, J. R. (1989). Clarifying some controversial issues surrounding statistical procedures for detecting moderator variables: Empirical evidence and related matters. *Journal of Applied Psychology, 74,* 3–10.

Stone-Romero, E. F., Alliger, G. M., & Aguinis, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects 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.

Thorndike, R. L. (1947). *Research problems and techniques* (Report No. 3, AAF Aviation Psychology Program Research Reports). Washington, DC: U.S. Government Printing Office.

Thorndike, R. L. (1949). *Personnel selection.* New York: Wiley.

Trattner, M. H., & O'Leary, B. S. (1980). Sample sizes for specified statistical power in testing for differential validity. *Journal of Applied Psychology, 65,* 127–134.

Walsh, J. P., & Kosnik, R. D. (1993). Corporate raiders and their disciplinary role in the market for corporate control. *Academy of Management Journal, 36,* 671–700.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Wohlgemuth, E., & Betz, N. E. (1991). Gender as a moderator of the relationship of stress and social support to physical health in college students. *Journal of Counseling Psychology, 38,* 367–374.

Xie, J. L., & Johns, G. (1995). Job scope and stress: Can job scope be too high? *Academy of Management Journal, 38,* 1288–1309.

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

(Appendix follows on next page)

Appendix

Computer Programs and Key Accuracy Checks

Computer Programs

The simulations were conducted using QuickBASIC 4.5 programs adapted from Aguinis (1994), Alliger (1992), and Stone-Romero et al. (1994). For the present study, a subroutine named "TrivarGenerate" was written to generate the point biserial correlation between the continuous predictor X and the dichotomous moderator Z. This subroutine generates random samples of size N for the variables X, Y, and Z from a population with specified correlations $\rho_{YX(1)}$, $\rho_{YX(2)}$, and ρ_{XZ}. First, for Subgroup 1, TrivarGenerate uses the Box–Muller method (Box & Muller, 1958) to generate random normal values for variables X and Y and the correlation $\rho_{YX(1)}$. Then, using the same method, it generates normal values for variables X and Y and the correlation between these two variables for cases in Subgroup 2 (i.e., $\rho_{YX(2)}$).

To manipulate the degree of predictor intercorrelation ρ_{XZ}, a constant K was added to values of X for the cases for which $Z = 2$. Because X values are normally distributed (mean zero, unit variance), the constant K is equivalent to the mean X score for cases for which $Z = 2$ (i.e., $\bar{X}_{Z=2}$). To compute $K = \bar{X}_{Z=2}$, a t-statistic equivalent to the specified point biserial correlation coefficient (ρ_{XZ}) for a predefined N is obtained with Equation 8.15.1 presented by Hays (1988, p. 311). Note that K takes on a different value for each combination of values of total sample size, sample size in Subgroup 1, sample size in Subgroup 2, and ρ_{XZ}.

Once samples of values for variables X, Y, and Z are generated with specified correlations $\rho_{YX(1)}$, $\rho_{YX(2)}$, and ρ_{XZ}, range ratios are manipulated using conditional statements of the type if $X(j)$ < = SEL THEN GOTO, where SEL is the truncation point on variable X. This branching statement specifies that only X scores above SEL should be retained (and the accompanying scores on Z and Y). If the X score selected randomly is smaller than SEL, then the conditional loop directs the program to generate a new value, until it generates an X score that is larger than SEL. Because X is normally distributed, the values of SEL were −.84, .25, and .84, normal scores marking off the top 80%, 40%, and 20% of the distribution. For the cells in which there was no selection, the command line referring to SEL was omitted, and scores were generated from the full population range.

Key Accuracy Checks

The accuracy of the programs was thoroughly examined through several manipulation checks. The strategy used was to compute statistics from the generated sampling distributions and to compare them with (a) their expected values (a function of population parameters specified in the program), (b) expected standard errors of relevant statistics, and (c) results obtained using programs written by other researchers. For example, to test the manipulation of $\rho_{XZ} = .30$, the r_{XZ} values from 400 samples of $N = 68$ ($n_1 = n_2 = 34$) were recorded. The mean sample correlation was .296, and the standard deviation of the r_{XZ}s was .109. This standard deviation corresponds to an expected S_r of .110 (i.e., $E(S_r) = [1 - \rho^2]/N^{.5}$). For the same parameters, a program used by Callender and Osburn (1988) yielded a similar S_r of .115. Analogous tests were conducted for several values of N and correlations among X, Y, and Z. The results of all such tests showed that key accuracy (validity) estimates derived from the programs were generally within ± .005 of their expected values.

Received March 1, 1996
Revision received July 29, 1996
Accepted July 30, 1996 ■