Appraisal of the Homogeneity of Error Variance Assumption and Alternatives to Multiple Regression for Estimating Moderating Effects of Categorical Variables

HERMAN AGUINIS
University of Colorado at Denver

SCOTT A. PETERSEN
U.S. Military Academy at West Point

CHARLES A. PIERCE
Montana State University

Homogeneity of within-subgroup error variance is a necessary condition for using multiple regression to estimate moderating effects of categorical variables. A 12-year review of Academy of Management Journal, Journal of Applied Psychology, *and* Personnel Psychology *indicates that the assumption is violated in 40% to 60% of cases. The authors reanalyze published research to illustrate how violating the assumption may change substantive conclusions. To remedy this situation, they develop and present a computer program (i.e., ALTMMR) that (a) assesses whether a data set suffers from heterogeneity of error variance and (b) computes alternative inferential statistics to the traditional multiple regression* F *test when heterogeneity exists. ALTMMR, which can also be used as a teaching tool, was written in Java and is executable using an Internet Web browser or as a stand-alone application.*

Using multiple regression to estimate the effects of categorical moderator variables (e.g., gender, ethnicity) involves a regression equation that examines the relationship between a predictor (e.g., job satisfaction) and moderator with a criterion (e.g., organizational citizenship behaviors). The equation includes first the continuous predictor and the categorical moderator and then a third variable consisting of their product, which carries information regarding the continuous predictor by categorical moderator interaction. The moderated multiple regression (MMR) equation is the following:

Authors' Note: The views expressed herein are those of the authors and do not purport to reflect the position of the U.S. Military Academy, the Department of the Army, or the Department of Defense. This article is based in part on the master's thesis of Scott A. Petersen, completed at the University of Colorado at Denver under the supervision of Herman Aguinis. Portions of this article were presented as part of a symposium

Organizational Research Methods, Vol. 2 No. 4, October 1999 315-339

$$\hat{Y} = a + b_1 X + b_2 Z + b_3 X \cdot Z, \tag{1}$$

where $\hat{Y}$ is the predicted value for the criterion, a is the least squares intercept, b_1 is the least squares estimate of the population regression coefficient for the continuous predictor X, b_2 is the least squares estimate of the population regression coefficient for the categorical moderator Z, and b_3 is the least squares estimate of the population regression coefficient for the product between X and Z (Cohen & Cohen, 1983). Rejecting the null hypothesis that $\beta_3 = 0$ indicates that Z moderates the relationship between X and Y. Stated differently, the slope of Y on X differs across values of Z (e.g., women and men).

Pervasive Use of MMR in Organizational Science

MMR seems to be the method of choice for evaluating moderating effects of categorical variables in human resources management, organizational behavior, applied psychology, and education (Sackett & Wilk, 1994; Stone, 1988). In addition to estimating effects of categorical moderator variables in general, MMR is specifically used to assess gender- and ethnicity-based differential slopes in preemployment testing. This is a special case of Equation (1) in which Y is a measure of performance (e.g., supervisory ratings), X is a preemployment test (e.g., general cognitive abilities), and Z is gender or ethnicity. The presence of a moderating effect of Z suggests that the test predicts performance differentially for the various gender- or ethnicity-based subgroups (e.g., male and female, majority and minority).

The widespread use of MMR in organizational science to assess effects of categorical moderators in general, as well as differential slopes in preemployment testing in particular, might be a result of the implicit recommendations found in the *Standards for Educational and Psychological Testing* (American Psychological Association [APA], 1985) and the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 1987). In fact, the March 1998 draft of the revised *Standards* (APA, 1998) includes a statement endorsing the use of MMR (Fairness in Testing and Test Use, Standard 7.6):

> When studies of differential prediction of a criterion for members of different subgroups are conducted, the reports should include regression equations (or an appropriate equivalent) computed separately for each group or treatment under consideration or an analysis in which the group or treatment variables are entered as moderator variables. (p. 14)

(J. M. Cortina, Chair) at the meeting of the Society for Industrial and Organizational Psychology, Atlanta, Georgia, April 1999. Portions of the research reported in this article were supported by a grant to Scott A. Petersen from the Graduate Research Opportunities Program at the University of Colorado at Denver. We thank Richard P. DeShon (Michigan State University), Kurt Kraiger (University of Colorado at Denver), William Wolfe (University of Colorado at Denver), and members of the Behavioral Science Research Group for comments regarding the research reported herein. 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; e-mail: haguinis@castle.cudenver.edu.

Table 1
Number of Published Articles in Three Selected Journals Using MMR to Estimate Moderating Effects of Categorical Variables (January 1987 to April 1999)

	Journal Title			
	AMJ	JAP	PP	*Total*
Total number of studies using MMR to assess categorical moderators	17	57	13	87
Number of studies with race or ethnicity as a moderator[a]	2	20	7	29
Number of studies with race or ethnicity as a moderator using performance as a criterion	0	4	4	8

Note. *AMJ* = *Academy of Management Journal*, *JAP* = *Journal of Applied Psychology*, *PP* = *Personnel Psychology.*
a. Any criterion (e.g., job satisfaction, union commitment).

In short, given the endorsement of the draft of the revised version of the *Standards*, it is likely that the use of MMR will continue to be pervasive.

Use of MMR to Assess Effects of Categorical Moderator Variables: Selective Literature Review

To illustrate how frequently MMR is used to assess effects of categorical moderator variables in organizational science, we conducted a review of articles published from January 1987 to April 1999 in *Academy of Management Journal* (*AMJ*), *Journal of Applied Psychology* (*JAP*), and *Personnel Psychology* (*PP*). We selected these three journals because they are among the most influential publications devoted to empirical research in management and applied psychology (Starbuck & Mezias, 1996). Thus, this review does not intend to be comprehensive and only attempts to investigate whether three of the most highly regarded journals that regularly publish articles relevant to organizational science show a high occurrence of MMR as a technique to assess moderating effects of categorical variables. Results of this review are summarized in Table 1.

Table 1 shows that 87 articles published in three of the most influential journals in organizational science in the past 12-year period used MMR to assess effects of categorical moderators. Of these 87 articles, 29 included gender or ethnicity as the categorical moderator. Note that these numbers refer to articles. Typically, each article included several MMR tests. In short, a large number of published articles have used MMR to assess effects of categorical moderator variables.

Problems With the Use of MMR and Homogeneity of Error Variance Assumption

Several researchers (e.g., Aguinis & Stone-Romero, 1997; Bobko & Russell, 1994; Stone-Romero, Alliger, & Aguinis, 1994) have pointed out that a number of methodological and statistical artifacts affect the statistical power of MMR. Stated differently, these factors lead researchers to make the sample-based conclusion that there is no moderating effect (e.g., a preemployment test predicts performance equally well for women and men), when in fact there is a moderating effect in the population (e.g., the test overpredicts performance for men) (Aguinis & Pierce, 1998b; Aguinis, Pierce, & Stone-Romero, 1994). Obviously, the low power of MMR tests, as well as the subsequent failure to detect a moderating effect, is a hindrance to theory development and the advancement of knowledge. Not detecting existing moderating effects because of a low-power statistical tool is a luxury that organizational science researchers cannot afford, especially those making decisions that directly affect thousands of individuals' lives (e.g., staffing decision making).

Some factors known to affect the power of MMR to detect categorical moderator variables are the following: (a) predictor range restriction (Aguinis & Stone-Romero, 1997), (b) predictor and criterion reliability (Bohrnstedt & Marwell, 1978), (c) criterion scale coarseness (Aguinis, Bommer, & Pierce, 1996; Russell & Bobko, 1992), (d) predictor and criterion artificial dichotomization or polychotomization (Stone-Romero & Anderson, 1994), (e) magnitude of the moderating effect, and (f) different sample sizes across moderator-based subgroups (Stone-Romero et al., 1994; see Aguinis, 1995, for a review of these factors).

In addition to the aforementioned statistical and methodological artifacts, violating the homogeneity of within-subgroup error variance assumption has been identified as a factor that can affect the power of MMR to detect categorical moderator variables (Alexander & DeShon, 1994; Dretzke, Levin, & Serlin, 1982). Analogous to ANOVA, MMR assumes that the variance in Y that remains after predicting Y from X is equal across k moderator-based subgroups. In each subgroup, this value is estimated by the mean square residual from the regression of Y on X:

$$\sigma^2_{e(i)} = \sigma^2_{y(i)}(1 - \rho^2_{xy(i)}), \tag{2}$$

and the homogeneity of error variance assumption is met when $\sigma^2_{e(i)} = \ldots = \sigma^2_{e(k)}$.

Consequences of Violating the Homogeneity of Error Variance Assumption

A recent review article in *Organizational Research Methods* described the homogeneity of error variance assumption and the critical consequences of violating it on MMR-based conclusions (Aguinis & Pierce, 1998a). Depending on population and sample characteristics, violating the assumption (a) increases or decreases Type I error rates (e.g., a researcher mistakenly concludes that a moderating effect exists) and (b) increases Type II error rates (i.e., a researcher mistakenly concludes that a moderating effect does not exist).

Regarding Type I error rates, Dretzke et al.'s (1982) simulation showed that Type I 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 (i.e., smaller X-Y correlation coefficient; cf. Equation (2)). For example, in a situation with subgroup ns (rs) of 50 (.25) and 100 (.75), the observed Type I error rate using MMR's F test was .18 for a nominal α of .05. In addition, Type I error rates are also affected under conditions of equal subgroup sample sizes. More specifically, Type I error rates become overly conservative when the *X* variance is dissimilar across subgroups. For example, a simulation by Aguinis, Boik, and Pierce (1998) showed that in a situation with $k = 2$, with subgroup *n*s (*r*s, SD_X) of 200 (.50, 2.0) and 200 (.10, 4.0), the observed Type I error rate using MMR's *F* test was .02 for a nominal α of .05.

Regarding the inflation of Type II error rates, Alexander and DeShon (1994) found that when the subgroup with the larger sample size is associated with the larger error variance (i.e., the smaller *X*-*Y* correlation), statistical power is lowered markedly. For example, in a case where $k = 2$, with subgroup *n*s (*r*s) of 20 (.20) and 40 (.50), the empirical rejection rate (power) of the ordinary *F* test was 1.00 (i.e., 100% correct rejections). Yet, when the larger *n* was paired with the smaller correlation (i.e., when *n*s [*r*s] were 20 [.50] and 40 [.20]), power dropped to approximately .79. As noted by Aguinis and Pierce (1998a), this specific scenario in which the subgroup with the larger *n* is paired with the smaller correlation coefficient is the most typical situation in validation research in a variety of organizational settings (e.g., industrial, educational, and military). Typically, the majority subgroup (e.g., Whites, men) is more numerous than the minority subgroup (e.g., African Americans, women), and the majority subgroup presents a validity coefficient that is smaller than that of the minority subgroup (e.g., Hattrup & Schmitt, 1990).

In sum, violating the homogeneity of error variance assumption can make MMR's *F* test results erratic and untrustworthy. Type I error rates can be inflated or overly conservative, and power can be lowered markedly. Consequently, substantive research conclusions can be erroneous, theory development can be hindered, and incorrect decisions can be made regarding whether preemployment test scores predict performance differentially across ethnicity- and gender-based subgroups. In short, MMR's *F* test is misleading and therefore should not be used in the absence of homogeneity of error variance.

Lack of Awareness of Homogeneity of Error Variance Assumption in Organizational Science

The homogeneity of error variance assumption was singled out as an important issue almost 50 years ago. Gulliksen and Wilks (1950) suggested that a test for differences in standard errors of the estimate (i.e., square root of within-subgroup error variances across moderator-based subgroups) should be conducted before testing for differences in slopes across subgroups. Moreover, Gulliksen and Wilks recommended that tests for inequality of slopes not be conducted in situations involving heterogeneity of standard errors of the estimate across subgroups.

Despite the fact that Gulliksen and Wilks (1950) discussed the homogeneity of error variance assumption almost half a century ago, and Dretzke et al. (1982) contended that the assumption is usually assessed in Aptitude × Treatment interaction research, there is little evidence that MMR users in organizational science have paid any attention to the issue. Only one (i.e., Stewart, Carson, & Cardy, 1996) of the 87 articles using MMR to assess moderating effects of categorical variables described in Table 1 reported having conducted such an assessment. In addition, only 8 (i.e., 9.2%) provided the descriptive statistics necessary for a reader to independently assess compliance with the assumption. In short, a conclusion from our review is that researchers do not seem to be aware of the homogeneity of error variance assumption or of the dramatic effects that violating the assumption has on statistical power and conclusions regarding the presence of moderating effects.[1]

Assessing Compliance With the Homogeneity of Error Variance Assumption

A plethora of tests have been developed to assess compliance with the homogeneity of error variance assumption in ANOVA models. Many of these tests can be modified to assess compliance with the assumption in an MMR context. Based on error rate comparisons of several of these tests by Gartside (1972) and Games, Winkler, and Probert (1972), DeShon and Alexander (1996) concluded that Bartlett's (1937) *M* test (see Appendix A for equations) is one of the most flexible and powerful tests available. Both Gartside and Games et al. found that in simulation conditions applicable to typical research conditions in management and other organizational sciences (e.g., three or fewer subgroups, unequal subgroup sample sizes), Bartlett's *M* test adhered most closely to nominal Type I error rates and demonstrated the highest statistical power rates. However, as noted by DeShon and Alexander, Bartlett's test was outperformed when the variables examined deviated from normality.

A second procedure to assess whether variances are heterogeneous across subgroups is to use DeShon and Alexander's (1996) 1.5 rule of thumb. DeShon and Alexander conducted extensive Monte Carlo simulations and concluded that the power of MMR is not adversely affected until the error variance of the subgroup with the largest error variance is approximately 1.5 times greater than that of the subgroup with the smallest error variance.

In sum, Aguinis and Pierce's (1998a) review concluded that Bartlett's (1937) *M* test and DeShon and Alexander's (1996) empirically derived 1.5 rule of thumb are the indicators of choice to assess whether the homogeneity of error variance assumption is violated.

Alternatives to MMR's *F* Test in the Presence of Heterogeneity of Error Variance

As noted above, MMR's *F* test provides erroneous results regarding the presence of moderator variables when the homogeneity of error variance assumption is violated (Aguinis & Pierce, 1998a). Assuming that a researcher uses any of the assessment techniques described above (i.e., Bartlett's [1937] test or DeShon & Alexander's [1996] rule of thumb) and concludes that error variances are heterogeneous, a practical concern is what to do next. Researchers still need to conduct an inferential test to make a conclusion regarding the moderating effect hypothesis, but MMR's *F* test is likely to yield misleading results.

Fortunately, the situation is surmountable. Three parametric and one nonparametric alternative for evaluating subgroup regression slope differences have been examined for cases of variance heterogeneity. The parametric alternatives correct for degrees of freedom associated with the more common MMR *F* test. The nonparametric alternative does not require that the homogeneity assumption be met. The four alternatives most thoroughly investigated via Monte Carlo simulations are the following: (a) the Welch-Aspin F^* approximation (Dretzke et al., 1982), (b) James's (1951) second-order approximation (*J*) (DeShon & Alexander, 1994; see Appendix B for equations), (c) Alexander's normalized *t* approximation (*A*) (Alexander & Govern, 1994; see Appendix C for equations), and (d) the nonparametric chi-square test *U* (Marascuilo, 1966).

Aguinis and Pierce (1998a) reviewed Monte Carlo studies that investigated the power rates of each of these tests. To summarize their review, the *A* and *J* statistics are the best alternatives to the more traditional MMR *F* test in conditions of heterogeneous error variances for evaluating subgroup slope differences. Even though the *U* statistic is not as susceptible to nonnormality as the other parametric tests, it has little utility. This test performs inadequately for small or moderately large sample sizes and requires intercepts across the (maximum of two) subgroups to be equal (DeShon & Alexander, 1996; Wilcox, 1988). The F^* has no advantage over the *A* or *J* statistics and is the most susceptible to normality deviations. The *A* statistic is simpler to compute and is slightly more robust to deviations from normality. However, one important advantage of *J* is its greater statistical power in small-sample situations. Finally, the nonparametric *U* statistic is not as powerful as the *A* and *J* parametric options.

The Need for a User-Friendly Tool to Assess Assumption Compliance and Compute Alternative Statistics

Despite the fact that researchers are likely to make erroneous conclusions regarding moderating effect hypotheses when using MMR's *F* test in the presence of heterogeneous subgroup error variances, only one of the 87 articles described in Table 1 reported that assumption compliance had been verified. Moreover, despite the fact that recent Monte Carlo studies have uncovered the properties of the various alternatives to MMR in the presence of heterogeneity, not a single article of the 87 included in Table 1 addressed the issue or used any of the alternative techniques available. A possible explanation is researchers' lack of awareness regarding violating the assumption and its consequences. We believe the Aguinis and Pierce (1998a) review might help remedy the awareness issue. However, even if researchers are aware of the impact of heterogeneity on MMR-based conclusions, they still face the practical concerns of how to estimate whether they are facing a heterogeneous variance situation and, if variances are heterogeneous, how to compute the alternative *A* and *J* statistics to be used in lieu of MMR. More specifically, none of the most widely used statistical packages (e.g., SPSS, SAS, Minitab) include procedures to allow researchers to perform these computations. And, as shown in Appendixes A, B, and C, the complexity of the equations to compute *M*, *A*, and *J* makes it difficult to obtain these statistics by hand or even using a spreadsheet program.

In response to this need, next we describe a computer program (ALTMMR) that addresses the above practical issues faced by researchers who test hypotheses regarding categorical moderator variables using MMR. As explained next, ALTMMR (a) enables assessment of compliance with the homogeneity of subgroup error variance assumption and (b) computes alternative inferential statistics to be used to assess the effects of categorical moderator variables when the assumption is violated. Such a tool will help researchers make more valid conclusions regarding moderating effect hypotheses.

Program Description

ALTMMR was written in Java using Microsoft Visual J++ (Version 1.1). Java is an increasingly popular programming language because of its flexibility for World Wide Web (WWW) applications and its ability to create appealing graphical interfaces. Also, we chose Java to enhance the number of potential program users. Most computers today have a WWW browser preinstalled (e.g., Internet Explorer, Netscape Navigator); therefore, a Java applet has the potential to reach the largest number of users regardless of operating system platform (e.g., Windows 95/98/NT, Macintosh, OS2). In addition to allowing execution using a WWW browser, Java applets can be used as stand-alone applications. Thus, the program is not limited only to those who have a WWW browser.

We developed ALTMMR in the following two versions: (a) a browser applet version and (b) a stand-alone version. The program requires minimal input (i.e., sample descriptive statistics rather than raw data) and outputs four statistics (ratio of the largest to the smallest error variance to be compared with DeShon and Alexander's [1996] 1.5 rule of thumb, Bartlett's [1937] *M*, James's [1951] *J*, and Alexander's *A* [Alexander & Govern, 1994]). In addition, the program was designed as an educational tool, so it includes extensive associated text-based information regarding MMR, the homogeneity of error variance assumption, and how to interpret the program's output. This information is available via hyperlinks found before the user is requested to input the necessary sample-based information. Thus, instructors could use ALTMMR online to teach about MMR and MMR's assumptions and perform demonstrations of how specific data sets violate the error variance assumption. Moreover, instructors could conduct online demonstrations of how, when the assumption is violated, results from MMR's *F* test are inconsistent with the more appropriate *A* and *J* statistics.

Input

ALTMMR prompts the user for the necessary information in two steps. First, the user enters the number of moderator-based subgroups to be compared (k) and selects the nominal alpha level in a dialog window (the alpha level is only necessary for James's [1951] *J*; precise *p* values are calculated for the *M* and *A* statistics). Second, a new dialog window is displayed in which the user enters information for each subgroup (e.g., women and men).

Assessment of variance heterogeneity. ALTMMR assesses whether the homogeneity of within-subgroup error variance assumption is violated by computing Bartlett's (1937) *M* statistic and the error variance ratio to be compared to DeShon and Alexander's (1996) rule of thumb described above. The user must provide (a) the number of subgroups (k), (b) the standard deviation of the criterion Y for each subgroup (i.e., $s_{y(i)}$), (c) the correlation between the criterion Y and the predictor X for each subgroup (i.e., $r_{xy(i)}$), and (d) the sample size for each subgroup (for calculation of the degrees of freedom). Based on this information, the program computes Bartlett's *M* and the precise associated *p* value of rejecting the null hypothesis that the variances are equal (Appendix A shows the equations involved in computing the *M* statistic).

To compute the error variance ratio, ALTMMR uses $s^2_{y(i)}$s and $r_{xy(i)}$s as estimates for $\sigma^2_{y(i)}$s and $\rho_{xy(i)}$s, respectively. Then, each subgroup error variance is computed using Equation (2). When more than two subgroups are evaluated, the program selects the largest error variance ratio from all possible pairwise combinations of error variances. The resulting error variance ratio is compared to DeShon and Alexander's (1996) 1.5 rule of thumb.

Computation of alternatives to MMR. To compute the *J* and *A* statistics, ALTMMR prompts users to input the standard deviations of the predictor *X* for each subgroup (i.e., $s_{x(i)}$). *J*, its associated adjusted critical value, and *A* are calculated using the equations shown in Appendixes B and C, respectively. Note that the equations presented in Appendixes B and C use unstandardized regression weights (i.e., *b*s), not correlations. Thus, ALTMMR computes each subgroup *b* as follows using the information already entered:

$$b_{(i)} = r_{xy(i)} \sqrt{\frac{s^2_{y(i)}}{s^2_{x(i)}}}.$$

Also, note that a precise *p* value cannot be calculated using James's (1951) equations. The thrust of this approximation is in adjusting the critical value of the chi-square distribution to correct for infinite degrees of freedom (DeShon & Alexander, 1994). Thus, the value of *J* is not referenced to the chi-square distribution directly (as is the *A* statistic). The adjusted critical value for *J* is calculated from an initial critical value based on the actual degrees of freedom for the sample.

Output

Figure 1 illustrates the output screen produced by ALTMMR for hypothetical data with two moderator-based subgroups. As shown on this illustrative output screen, ALTMMR first outputs the user-input values to enable verification. Second, the program provides information regarding compliance with the homogeneity of error variance assumption as follows: (a) the (largest) ratio of error variance and (b) the value of Bartlett's (1937) *M* statistic and its associated *p* value regarding the null hypothesis of homogeneity of variance. In this hypothetical example, both indicators suggest that the error variances are heterogeneous across subgroups.

In addition to providing the homogeneity-related results, the program includes conditional statements to enable appropriate interpretations of the values related to the assessment of homogeneity—that is, whether the ratio meets DeShon and Alexander's (1996) rule of thumb and whether Bartlett's (1937) test indicates homogeneity or heterogeneity. In addition, if these indicators suggest heterogeneous error variances, the output screen indicates so.

Note that the program is based on an empirical (i.e., DeShon & Alexander's [1996] rule of thumb) and a theoretical (i.e., significance test associated with the *M* statistic) criterion to determine whether the assumption has been violated. However, violation of the assumption is a matter of degree rather than a binary outcome. Consequently, if (a) the error variance ratio falls just short of the 1:1.5 critical value, (b) the *p* value associated with the *M* statistic falls just short of the preestablished Type I error rate (e.g., .05), or (c) there is a discrepancy between the two results, then MMR users are advised to do the following. We suggest that, in addition to reporting MMR's *F* test, research-

```
The data you input were..

      Subgroup   n    sx    sy    rxy
         1       50   5.0   5.0   0.25
         2      100   3.0   9.0   0.60

-----Homogeneity of Subgroup Error Variance Information-----
DeShon & Alexander's rule of thumb for homogeneity is NOT met. (The
Error Variance ratio is 1:2.21)

Bartlett's Test indicates heterogeneous error variance (M = 9.2746,
p  < .0023).

Since both statistics indicate heterogeneous error variance, a
necessary assumption of MMR is NOT met, and the alternative
statistics below are more accurate indicators of a moderating
effect.

        ------------Alternative Statistics---------------
James's Test INDICATES A MODERATING EFFECT! (p < 0.05) J = 30.6806,
and J(critical) = 3.906

Alexander's Test INDICATES A MODERATING EFFECT! (A = 27.6078, p =
0.0)

Thus, both statistics indicate that the regression slopes for the
subgroups are different.
```

Figure 1: Sample Screen Output for Illustrative Use of ALTMMR

ers report the *A* and *J* values provided by ALTMMR. If all three values lead to the same conclusion (i.e., presence or absence of a moderating effect), researchers can be confident about their results. On the other hand, if results of the three tests do not converge, researchers should report results regarding compliance with the assumption and acknowledge that results must be replicated before definitive conclusions are made. Obviously, this second scenario introduces uncertainty and does not allow researchers to make a decisive conclusion regarding their moderating effect hypothesis. However, the same scenario takes place when the statistical power of MMR's *F* test is low. More precisely, Aguinis and Stone-Romero (1997) suggested that when an MMR analysis is conducted at low levels of power, researchers should acknowledge that results must be replicated before concluding that a null hypothesis of no moderating effect is tenable.

In addition to the homogeneity-related results, Figure 1 shows that ALTMMR displays results regarding the *J* and *A* statistics (i.e., alternatives to MMR in the case of heterogeneity). Conditional statements also provide information regarding the interpretation of these values. That is, the user is provided with the adjusted critical value for *J*, the precise *p* value for *A*, and whether these values indicate the presence of the hypothesized moderating effect.

In sum, the output screen provides information regarding (a) the assessment of compliance with the homogeneity of subgroup error variance assumption, (b) alternatives to MMR, and (c) information regarding these results that aids researchers with interpretation.

Using ALTMMR to Assess Frequency of Assumption Violation in Published Research

As noted above, when the homogeneity assumption is violated, substantive research conclusions based on the traditional MMR F test can be highly inaccurate. However, Alexander and DeShon (1994) noted that "there is very little information available on either the extent or the magnitude of this assumption violation in actual research" (p. 312). This situation was still the case 2 years later, which prompted DeShon and Alexander (1996) to state that "there is very little empirical information on the extent of error variance heterogeneity in the literature" (p. 273). More recently, Aguinis and Pierce's (1998a) review addressed conceptual issues surrounding the assumption but did not assess empirically the frequency of assumption violation. A frequent violation of the assumption in studies published in influential organizational science journals would make a compelling case for the need to check for assumption violation and computation of alternative statistics to MMR as needed.

In response to this research need, we used ALTMMR to assess the frequency of assumption violation in the 87 articles published in *AMJ*, *JAP*, and *PP* described in Table 1. As noted above, only 8 (i.e., 9.2%) provided the descriptive statistics necessary to independently assess compliance with the assumption. However, as is typical in the MMR literature, many of these articles reported more than one MMR analysis. Thus, we were able to check compliance with the homogeneity assumption for a total of 117 MMR tests. Table 2 shows results of this reanalysis.

Based on results summarized in Table 2, the following two findings are worth noting. First, there is a high degree of agreement between the two criteria used to assess assumption violation: Of 117 tests, DeShon and Alexander's (1996) rule of thumb and Bartlett's (1937) M statistic provided conflicting results in 22 cases (i.e., agreement of 81.20%). Moreover, an examination of the 22 cases for which the criteria provided conflicting results shows that when one of the criteria showed violation, the other one fell just short of also suggesting violation. For example, line 124 in Table 2 summarizing results from Melamed, Ben-Avi, Luz, and Green's (1995) test of the Monotony × Gender interaction shows that Bartlett's M statistic suggests violation (i.e., $M = 16.83$, $p < .05$), and DeShon and Alexander's rule of thumb does not. However, the error variance ratio is 1:1.41, just short of the 1:1.50 empirically derived cutoff score. Similar values were obtained for the remaining 21 cases for which the two criteria provided conflicting results. Of these 21 cases, 19 correspond to Mael (1995), and in all instances, the error variance ratio was 1:1.30 or greater. For the remaining 2 cases in which the two criteria provided conflicting results, the rule of thumb indicated violation, and the M statistic fell just short of being statistically significant.

The second noteworthy finding from Table 2 is that the assumption is violated more frequently than one would have anticipated. Tallying results for which both criteria agreed on whether there was a violation, Table 2 shows that the assumption was violated in 46 instances (i.e., 39.32%). If we count results for which at least one criterion indicated that the assumption was violated, this number increases to 68 (i.e., 58.12%). Thus, violation of the assumption occurred in approximately 40% to 60% of the MMR tests.

(text continued on p. 331)

Table 2
Checks of Compliance With Homogeneity of Error Variance Assumption in Articles From Table 1

Author(s)	*Criterion*	*Predictor*	*Moderator*	*Error Variance Ratio*	M
Easterling and Leventhal (1989)	Worry about cancer	Perceived risk	Family cancer history category	**1:2.66**	**13.29**
Gibbons, Helweg-Larsen, and Gerrard (1995)	Sexual willingness	Condom intention	Nationality (Danish, American)	1:1.13	1.15
Gibbons et al. (1995)	Sexual willingness	Parental influence	Nationality (Danish, American)	1:1.12	0.95
Gibbons et al. (1995)	Sexual willingness	Friend influence	Nationality (Danish, American)	1:1.08	0.39
Gibbons et al. (1995)	Sexual willingness	Prevalence estimate	Nationality (Danish, American)	1:1.07	0.33
Gibbons et al. (1995)	Sexual willingness	Prototype	Nationality (Danish, American)	1:1.02	0.02
Gibbons et al. (1995)	Smoking	Parent influence	Nationality (Danish, American)	**1:2.35**	**59.23**
Gibbons et al. (1995)	Smoking	Friend influence	Nationality (Danish, American)	**1:2.52**	**69.77**
Gibbons et al. (1995)	Smoking	Prevalence estimate	Nationality (Danish, American)	**1:2.56**	**71.85**
Gibbons et al. (1995)	Smoking	Prototype	Nationality (Danish, American)	**1:2.69**	**79.96**
Hattrup and Schmitt (1990)	Weighted task performance measures	Differential Aptitude Test/ Employee Aptitude Test combination	Ethnicity	**1:1.75**	**7.20**
Hattrup and Schmitt (1990)	Weighted task performance measures	Differential Aptitude Test/ Employee Aptitude Test combination	Gender	1:1.00	0.00
Hattrup and Schmitt (1990)	Weighted task performance measures	Alternate Aptitude Test	Ethnicity	**1:2.04**	**12.05**
Hattrup and Schmitt (1990)	Weighted task performance measures	Alternate Aptitude Test	Gender	1:1.10	0.13
Hattrup and Schmitt (1990)	Task performance measures	Differential Aptitude Test/ Employee Aptitude Test combination	Ethnicity	**1:1.55**	**4.29**
Hattrup and Schmitt (1990)	Task performance measures	Differential Aptitude Test/ Employee Aptitude Test combination	Gender	1:1.18	0.33

Hattrup and Schmitt (1990)	Task performance measures	Alternate Aptitude Test	Ethnicity	**1:1.82**	**8.31**
Hattrup and Schmitt (1990)	Task performance measures	Alternate Aptitude Test	Gender	1:1.01	0.00
Latack, Josephs, Roach, and Levine (1987)	Program satisfaction	Age	Gender	1:1.02	0.01
Latack et al. (1987)	Program satisfaction	Age	Gender	1:1.02	0.01
Latack et al. (1987)	Program satisfaction	Education	Gender	1:1.07	0.09
Latack et al. (1987)	Program satisfaction	Realistic expectations	Gender	1:1.03	0.01
Latack et al. (1987)	Program satisfaction	Anxiety	Gender	1:1.03	0.02
Latack et al. (1987)	Program satisfaction	Affirmative action support	Gender	1:1.04	0.03
Latack et al. (1987)	Program satisfaction	Coworker acceptance	Gender	1:1.15	0.35
Latack et al. (1987)	Program satisfaction	Job assignments	Gender	1:1.03	0.02
Latack et al. (1987)	Program satisfaction	Organizational acceptance	Gender	1:1.11	0.21
Latack et al. (1987)	Union satisfaction	Age	Gender	**1:2.51**	**19.39**
Latack et al. (1987)	Union satisfaction	Education	Gender	**1:2.36**	**16.65**
Latack et al. (1987)	Union satisfaction	Realistic expectations	Gender	**1:2.51**	**19.43**
Latack et al. (1987)	Union satisfaction	Anxiety	Gender	**1:2.27**	**14.94**
Latack et al. (1987)	Union satisfaction	Affirmative action support	Gender	**1:2.45**	**18.25**
Latack et al. (1987)	Union satisfaction	Coworker acceptance	Gender	**1:2.28**	**15.13**
Latack et al. (1987)	Union satisfaction	Job assignments	Gender	**1:2.76**	**24.08**
Latack et al. (1987)	Union satisfaction	Organizational acceptance	Gender	**1:3.11**	**31.07**
Latack et al. (1987)	Job satisfaction	Age	Gender	1:1.00	0.00
Latack et al. (1987)	Job satisfaction	Education	Gender	1:1.01	0.00
Latack et al. (1987)	Job satisfaction	Realistic expectations	Gender	1:1.09	0.14
Latack et al. (1987)	Job satisfaction	Anxiety	Gender	1:1.05	0.04
Latack et al. (1987)	Job satisfaction	Affirmative action support	Gender	1:1.00	0.00
Latack et al. (1987)	Job satisfaction	Coworker acceptance	Gender	1:1.12	0.24
Latack et al. (1987)	Job satisfaction	Job assignments	Gender	1:1.06	0.05
Latack et al. (1987)	Job satisfaction	Organizational acceptance	Gender	1:1.00	0.00
Latack et al. (1987)	Performance	Age	Gender	1:1.01	0.00
Latack et al. (1987)	Performance	Education	Gender	1:1.13	0.28
Latack et al. (1987)	Performance	Realistic expectations	Gender	1:1.08	0.10

(continued)

Table 2 Continued

Author(s)	*Criterion*	*Predictor*	*Moderator*	*Error Variance Ratio*	M
Latack et al. (1987)	Performance	Anxiety	Gender	1:1.11	0.22
Latack et al. (1987)	Performance	Affirmative action support	Gender	1:1.23	0.82
Latack et al. (1987)	Performance	Coworker acceptance	Gender	1:1.12	0.24
Latack et al. (1987)	Performance	Job assignments	Gender	1:1.11	0.19
Latack et al. (1987)	Performance	Organizational acceptance	Gender	1:1.11	0.20
Latack et al. (1987)	Probability of finishing	Age	Gender	**1:6.44**	**39.17**
Latack et al. (1987)	Probability of finishing	Education	Gender	**1:6.26**	**38.20**
Latack et al. (1987)	Probability of finishing	Realistic expectations	Gender	**1:6.26**	**38.20**
Latack et al. (1987)	Probability of finishing	Anxiety	Gender	**1:6.18**	**37.78**
Latack et al. (1987)	Probability of finishing	Affirmative action support	Gender	**1:6.15**	**37.63**
Latack et al. (1987)	Probability of finishing	Coworker acceptance	Gender	**1:2.93**	**15.75**
Latack et al. (1987)	Probability of finishing	Job assignments	Gender	**1:6.25**	**38.14**
Latack et al. (1987)	Probability of finishing	Organizational acceptance	Gender	**1:6.22**	**38.01**
Latack et al. (1987)	Union commitment	Age	Gender	1:1.21	0.64
Latack et al. (1987)	Union commitment	Education	Gender	1:1.16	0.38
Latack et al. (1987)	Union commitment	Realistic expectations	Gender	1:1.41	1.95
Latack et al. (1987)	Union commitment	Anxiety	Gender	1:1.20	0.58
Latack et al. (1987)	Union commitment	Affirmative action support	Gender	1:1.41	1.90
Latack et al. (1987)	Union commitment	Coworker acceptance	Gender	1:1.19	0.53
Latack et al. (1987)	Union commitment	Job assignments	Gender	1:1.26	0.91
Latack et al. (1987)	Union commitment	Organizational acceptance	Gender	1:1.01	0.00
Mael (1995)	Swimming ability	Hours of exercise	Ethnicity (White, African American)	1:1.31	**4.83**
Mael (1995)	Swimming ability	Preference for individual sports	Ethnicity (White, African American)	1:1.31	**5.03**
Mael (1995)	Swimming ability	Physical aptitude exam	Ethnicity (White, African American)	1:1.33	**5.36**
Mael (1995)	Swimming ability	Rugged, outdoors activities	Ethnicity (White, African American)	1:1.31	**4.90**
Mael (1995)	Swimming ability	Body mass index	Ethnicity (White, African American)	1:1.30	**4.79**
Mael (1995)	Swimming ability	Body weight	Ethnicity (White, African American)	1:1.31	**4.89**
Mael (1995)	Swimming ability	Age learned to swim	Ethnicity (White, African American)	1:1.44	**8.71**
Mael (1995)	Swimming ability	Boy/Girl Scout	Ethnicity (White, African American)	1:1.33	**5.59**

Mael (1995)	Swimming ability	Television watching	Ethnicity (White, African American)	1:1.40	**7.54**
Mael (1995)	Swimming ability	Music lessons	Ethnicity (White, African American)	1:1.32	**5.15**
Mael (1995)	Swimming ability	Social studies favorite subject	Ethnicity (White, African American)	1:1.36	**6.33**
Mael (1995)	Swimming ability	High school rank	Ethnicity (White, African American)	1:1.37	**6.48**
Mael (1995)	Swimming ability	Nights spent at home per week	Ethnicity (White, African American)	1:1.30	**4.69**
Mael (1995)	Swimming ability	Weekly hours spent on homework	Ethnicity (White, African American)	1:1.33	**5.44**
Mael (1995)	Swimming ability	Religious attendance	Ethnicity (White, African American)	1:1.35	**6.00**
Mael (1995)	Swimming ability	Preferred number of social friends	Ethnicity (White, African American)	1:1.30	**4.69**
Mael (1995)	Swimming ability	Earlier part-time work	Ethnicity (White, African American)	1:1.34	**5.68**
Mael (1995)	Swimming ability	More summers worked	Ethnicity (White, African American)	1:1.31	**4.97**
Mael (1995)	Swimming ability	Birth order	Ethnicity (White, African American)	1:1.35	**6.00**
Mael (1995)	Age of learning to swim	Hours of exercise	Ethnicity (White, African American)	**1:2.77**	**102.01**
Mael (1995)	Age of learning to swim	Preference for individual sports	Ethnicity (White, African American)	**1:2.80**	**104.85**
Mael (1995)	Age of learning to swim	Physical aptitude exam	Ethnicity (White, African American)	**1:2.77**	**101.95**
Mael (1995)	Age of learning to swim	Rugged, outdoors activities	Ethnicity (White, African American)	**1:2.83**	**107.22**
Mael (1995)	Age of learning to swim	Body mass index	Ethnicity (White, African American)	**1:2.80**	**104.88**
Mael (1995)	Age of learning to swim	Body weight	Ethnicity (White, African American)	**1:2.81**	**105.06**
Mael (1995)	Age of learning to swim	Boy/Girl Scout	Ethnicity (White, African American)	**1:2.74**	**99.95**
Mael (1995)	Age of learning to swim	Television watching	Ethnicity (White, African American)	**1:2.74**	**99.82**
Mael (1995)	Age of learning to swim	Music lessons	Ethnicity (White, African American)	**1:2.83**	**106.84**
Mael (1995)	Age of learning to swim	Social studies favorite subject	Ethnicity (White, African American)	**1:2.74**	**99.46**
Mael (1995)	Age of learning to swim	High school rank	Ethnicity (White, African American)	**1:2.76**	**101.28**
Mael (1995)	Age of learning to swim	Nights spent at home per week	Ethnicity (White, African American)	**1:2.81**	**105.73**
Mael (1995)	Age of learning to swim	Weekly hours spent on homework	Ethnicity (White, African American)	**1:2.81**	**105.28**
Mael (1995)	Age of learning to swim	Religious attendance	Ethnicity (White, African American)	**1:2.74**	**99.30**
Mael (1995)	Age of learning to swim	Preferred number of social friends	Ethnicity (White, African American)	**1:2.80**	**104.60**

(continued)

Table 2 Continued

Author(s)	*Criterion*	*Predictor*	*Moderator*	*Error Variance Ratio*	M
Mael (1995)	Age of learning to swim	Earlier part-time work	Ethnicity (White, African American)	**1:2.79**	**104.01**
Mael (1995)	Age of learning to swim	More summers worked	Ethnicity (White, African American)	**1:2.78**	**103.31**
Mael (1995)	Age of learning to swim	Birth order	Ethnicity (White, African American)	**1:2.69**	**95.79**
Melamed, Ben-Avi, Luz, and Green (1995)	Job satisfaction	Subjective monotony	Gender	1:1.14	2.23
Melamed et al. (1995)	Psychological distress	Subjective monotony	Gender	1:1.41	**16.83**
Melamed et al. (1995)	Sick days	Subjective monotony	Gender	**1:1.93**	**63.37**
Pulakos and Schmitt (1995)	Performance	Interview	Gender	1:1.18	1.28
Pulakos and Schmitt (1995)	Performance	Interview	Ethnicity	1:1.11	0.32
Srinivas and Motowidlo (1987)	Performance rating dispersion	Type A score	Workload category (high, low)	**1:2.71**	**14.13**
Tubbs (1993)	Performance	Absolute goal discrepancy	Goal level category (easy, hard)	1:1.20	0.49
Tubbs (1993)	Performance	Self-report composite test	Goal level category (easy, hard)	1:1.43	1.94
Tubbs (1993)	Performance	Direct commitment	Goal level category (easy, hard)	**1:1.51**	2.56
Tubbs (1993)	Performance	Hollenbeck, Williams, and Klein (1989) commitment measure	Goal level category (easy, hard)	**1:1.61**	3.40
Tubbs (1993)	Performance	Effort-related commitment	Goal level category (easy, hard)	1:1.37	1.50
Tubbs (1993)	Performance	Motivational force	Goal level category (easy, hard)	**1:1.73**	**4.44**
Tubbs (1993)	Performance	Personal goal	Goal level category (easy, hard)	1:1.15	0.31

Note. Error variance ratio = ratio of largest to smallest subgroup error variance; *M* = Bartlett's (1937) statistic. Error variance ratios larger than 1:1.5 are noted in bold type and suggest that the homogeneity of error variance assumption has been violated. Statistically significant *M* values (i.e., $p < .05$) are noted in bold type and suggest that the homogeneity of error variance assumption has been violated.

It should be noted that articles reviewed in Table 2 were published in three of the most influential journals in human resources management, organizational behavior, and applied psychology. *AMJ*, *JAP*, and *PP* are known for enforcing rigorous methodological standards. Thus, if between approximately 40% and 60% of MMR tests reported in these journals violate the homogeneity assumption, we can assume that this number is at least as high for tests reported in other journals in organizational science.

Illustration of Change in Substantive Research Conclusions

As noted above, only one of the 87 articles described in Table 1 reported that an assessment of compliance with the homogeneity assumption had been conducted. In addition, none of these 87 articles computed alternative statistics to MMR's *F* test in the presence of variance heterogeneity. Moreover, articles did not typically report sufficient information for readers to independently assess assumption compliance and compute alternative statistics if needed. Table 2 shows that fewer than 10% of the articles shown in Table 1 reported the descriptive statistics needed to assess compliance with the assumption. In addition, virtually none reported the necessary information to compute the *A* and *J* statistics. Thus, we located and contacted all authors of articles shown in Table 2 directly by telephone or e-mail to obtain this information. The vast majority of these authors contacted did not have access to their data for reasons ranging from a change in affiliation and losing of data in the move to a damaged data storage device (e.g., hard drive, backup tape). Thus, given this situation, we can only illustrate how violating the assumption changes substantive research conclusions.

One illustration of how violation of the homogeneity assumption changes substantive research conclusions is a study by Mael (1995). Mael investigated, among other issues, whether the relationship between the criterion "age at which one learned to swim" and the predictor "number of summers worked" was moderated by ethnicity. As shown on line 120 in Table 2, the assumption was violated. More precisely, the error variance ratio was 1:2.78 and $M = 103.31$, $p < .05$. Mael reported that results of MMR's *F* test indicated that ethnicity was in fact a moderator. However, a perusal of the descriptive statistics reported by Mael shows that the group with the largest sample size (i.e., Whites) was paired with the larger predictor-criterion correlation coefficient. As described above, Dretzke et al. (1982) showed that this direct pairing of n with r typically leads to inflated Type I error rates. At any rate, because the assumption was violated, it was inappropriate to use MMR's *F* test. Thus, we proceeded to compute the appropriate *A* and *J* statistics using ALTMMR. Results indicated that ethnicity was not a moderator; $A = 2.59$, $p > .05$, and $J = 2.62$, $p > .05$. Based on Dretzke et al.'s results, it is possible that the statistically significant moderating effect reported by Mael is a product of a Type I error.

Another illustration of how violation of the homogeneity assumption affects substantive organizational research conclusions is a study by Latack, Josephs, Roach, and Levine (1987). Latack et al. investigated various factors hypothesized to affect the career path of women in nontraditional occupations. As shown in Table 2, they used MMR to investigate the possible moderating effect of gender in a sample of carpenter apprentices. Line 46 in Table 2 shows that one such MMR test included coworker acceptance as a predictor and union satisfaction as a criterion. Coworker acceptance was operationalized as the extent to which carpenter apprentices are accepted by fellow apprentices and journeymen, feelings of fitting in, and extent of teasing and harassment experienced (reverse coded), whereas union satisfaction was operationalized as the extent to which carpenter apprentices were satisfied with what the union provided in contract negotiation, job security, service to members, improved wages, and so forth. As shown on line 46 in Table 2, the homogeneity assumption was violated; the error variance ratio was 1:2.28 and $M = 15.13$, $p < .01$. Latack et al. conducted an MMR analysis and found that the F test was not statistically significant and therefore concluded that there was no moderating effect of gender on the coworker acceptance–union satisfaction relationships. Note, however, that because the assumption was violated, it was inappropriate to use MMR's F test to test this relationship. Moreover, the descriptive statistics used to investigate this relationship were such that the group with the smaller sample size (i.e., women) was paired with the greater predictor-criterion correlation coefficient. As noted above, violating the assumption in an inverse pattern of the sample size–correlation coefficient situation leads to a reduction of statistical power (Alexander & DeShon, 1994). Thus, we used ALTMMR to compute the more appropriate A and J statistics.

Results showed that contrary to the conclusion derived by using the inappropriate F test, there was a moderating effect of gender; $A = 4.49$, $p < .05$, and $J = 4.20$, $p < .05$. More precisely, the coworker acceptance–union satisfaction slope was steeper for women than for men. That is, the same degree of coworker acceptance led to greater union satisfaction for women than for men.

Obviously, a change in substantive conclusions has implications for theory development. In some cases, a change in substantive conclusions can also affect the implementation of organizational interventions. Take, for instance, the case of Latack et al. (1987) described above addressing factors hypothesized to promote women's success in nontraditional occupations. The moderating effect of gender can be explained by the fact that women in nontraditional occupations have different expectations than men (Aguinis & Adams, 1998). They anticipate and expect to be teased and ridiculed, whereas men do not (Latack et al., 1987). Thus, they expect to not be easily and rapidly accepted. Consequently, a similarly moderate level of perceived acceptance for women and men may just meet men's expectations, but it may far exceed women's expectations. In turn, women's exceeded expectations regarding coworker acceptance may create positive affect that spills over to greater union satisfaction scores as compared to men's reported union satisfaction. Future research could explore this post hoc explanation.

Finally, we emphasize that we do not intend to devalue Mael's (1995) and Latack et al.'s (1987) studies. Rather, our goal is to illustrate that violating the homogeneity assumption may change substantive research conclusions and affect theory development in organizational science. Given the widespread violation of the homogeneity assumption (cf. Table 2), it is likely that substantive conclusions of numerous other studies might change if the appropriate A and J statistics were used instead of the inappropriate F test in assessing the presence of a categorical moderating effect.

Program Availability

The browser-based and stand-alone versions of the program are available free of charge. There are several ways to obtain them. If one has access to the Internet, the browser version can be executed online using a Java-capable Web browser (e.g., Microsoft Internet Explorer, Version 4.0 or later; Netscape Navigator, Version 3.0 or later) at http://members.aol.com/imsap/altmmr.html. Web page links are also available at this site to download both the browser and stand-alone versions for offline execution. Alternatively, for those without WWW access, the program can be sent electronically as an attachment file by e-mailing a request to the first author. Finally, for those without Internet access, the program can be obtained by sending a blank formatted diskette and a self-addressed, stamped envelope to the first author.

Concluding Remarks

Aguinis and Pierce (1998a) reviewed the homogeneity of within-subgroup error variance assumption and the implications of violating the assumption for MMR-based conclusions. In this article, we reviewed published research in three of the most influential journals in human resources management, organizational behavior, and applied psychology and found that the assumption was violated in 40% to 60% of cases. Moreover, as reviewed by Aguinis and Pierce and illustrated in this article, in the presence of heterogeneity of variance, results based on MMR cannot be trusted. Because results based on MMR's *F* test become highly erratic, MMR should not have been used in these studies. Despite Aguinis and Pierce's review, knowledge of the assumption may not be sufficient for researchers to have the ability to check compliance and implement alternatives. Currently, none of the major statistical packages include procedures to accomplish these goals. ALTMMR is a user-friendly program that can be executed on any platform and allows researchers to (a) check compliance with the assumption and (b) compute alternatives to MMR that can be used in the presence of heterogeneity. In addition, ALTMMR is an educational tool regarding MMR in general and provides hyperlinks to text-based information. In closing, violating the homogeneity of error variance assumption can lead researchers to make incorrect conclusions regarding moderating effects. However, we believe that the assumption will not be routinely checked, and alternatives will not be implemented unless a tool is readily available and easy to use. ALTMMR helps to address these needs.

APPENDIX A
Computation of Bartlett's (1937) *M* Statistic

Bartlett's (1937) *M* statistic is approximately distributed as chi-square with $k - 1$ degrees of freedom when sample size in each of the moderator-based subgroups $n_k - 1 \geq 3$. Given that k = number of subgroups, n_k = number of observations in each subgroup, s^2 = subgroup variance on the criterion, and v = degrees of freedom from which s^2 is based, the *M* statistic is computed as

$$M = \frac{\left(\sum_i v_i\right)\log_e\left(\frac{\sum_i v_i s_i^2}{\sum_i v_i}\right) - \sum_i v_i \log_e s_i^2}{1 + \frac{1}{3(k-1)}\left(\Sigma_i \frac{1}{v_i} - \frac{1}{\sum_i v_i}\right)}. \tag{A1}$$

For unconditional subgroup variances, substituting $\sigma^2_{e(i)}$ from Equation (2) for s_i^2 yields

$$M = \frac{\left(\sum_i v_i\right) \log_e \left(\frac{\sum_i v_i \sigma^2_{e(i)}}{\sum_i v_i}\right) - \sum_i v_i \log_e \sigma^2_{e(i)}}{1 + \frac{1}{3(k-1)}\left(\Sigma_i \frac{1}{v_i} - \frac{1}{\sum_i v_i}\right)}. \tag{A2}$$

APPENDIX B
Computation of James's (1951) *J* Statistic

To test for differential slopes, *J* is computed using Equation (B1) (Alexander & Govern, 1994):

$$J = \sum_{i=1}^{k} \left[\frac{(b_i - b^+)^2}{s^2_{b_i}}\right]. \tag{B1}$$

The components of Equation (B1) are obtained by implementing the following steps:

1. Determine the squared standard error ($s^2_{b_i}$), where $s^2_{b_i} = \frac{\left(1 - r_i^2\right) s^2_{Y_i}}{\left(n_i - 2\right) s^2_{X_i}}$.
2. Define a weight for each regression weight (i.e., b_i) such that $\sum w_i = 1$: $w_i = \frac{1/s^2_{b_i}}{\sum_{i=1}^{k} 1/s^2_{b_i}}$.
3. The variance-weighted estimate of the common regression slope (b^+) then becomes $b^+ = \sum_{i=1}^{k} w_i b_i$.

Once the *J* statistic has been computed, the adjusted critical value (*c*) for a chi-square distribution with $k - 1$ degrees of freedom and nominal Type I error α is determined as follows:

1. Let $v_i = n_i - 2$.
2. $R_{st} = \sum_{i=1}^{k} \left[\frac{w_i^t}{v_i^s}\right]$ (note that values for *s* and *t* are only those required in the calculation of $h(\alpha)$ below).
3. $\chi_{2.s} = \frac{c^s}{\prod_{q=1}^{s} (k + 2s - 3)}$, where $\prod_{q=1}^{S}$ denotes the product of each term from 1 to *s* (note that *s* represents the multiplier to provide the values of $\chi_{2 \cdot s}$ [i.e., for $\chi_2, \chi_4, \chi_6, \chi_8$, *s* is 1, 2, 3, 4, respectively]).
4. $T = \sum_{i=1}^{k} \frac{(1 - w_i)^2}{v_i}$.
5. $h(\alpha)$, adjusted for infinite degrees of freedom, is calculated as follows:

$$
h(\alpha) = c + \frac{1}{2}\left(3\chi_4 + \chi_2\right)T + \left[
\begin{array}{l}
\left(\frac{1}{16}\right)\left(3\chi_4 + \chi_2\right)^2 \frac{1-(k-3)}{c}T^2 + \left(\frac{1}{2}\right)(3\chi_4 + \chi_2)\times \\
\left[
\begin{array}{l}
\left(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2\right) + \\
\left(2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2\right)\times\left(\chi_2 - 1\right) + \\
\left(\frac{1}{4}\right)\left(-R_{12} + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2\right)\times \\
\left(3\chi_4 - 2\chi_2 - 1\right)
\end{array}
\right] \\
\left(R_{23} - 3R_{22} + 3R_{21} - R_{20}\right)\left(5\chi_6 + 2\chi_4 + \chi_2\right) + \\
\left(\frac{3}{16}\right)\left(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20}\right)\times \\
\left(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2\right) + \\
\left(\frac{1}{16}\right)\left(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2\right)\times \\
\left(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2\right) + \\
\left(\frac{1}{4}\right)\left(-R_{22} + R_{11}^2\right)\left(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2\right) + \\
\left(\frac{1}{4}\right)\left(R_{23} - R_{12}R_{11}\right)\left(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2\right)
\end{array}
\right]
$$

The null hypothesis (i.e., H_0: $\beta_1 = . . . = \beta_k$) is rejected when $J > h(\alpha)$.

APPENDIX C
Computation of Alexander's (Alexander & Govern, 1994) *A* Statistic

Calculation of the *A* statistic is similar to the computation for *J* described in Appendix B. To test for differences in slopes, the *A* statistic is calculated using Equation (C1) and referenced to the chi-square distribution with $(k - 1)$ degrees of freedom:

$$A = \sum_{i=1}^{k} z_i^2, \tag{C1}$$

where the following steps are required:

1. Determine the squared standard error ($S_{b_i}^2$), define a weight for each regression weight (i.e., b_i), and determine the variance-weighted estimate of the common regression slope (b^+) as in Steps 1 through 3 in Appendix B.
2. Define a one-sample *t* statistic for each subgroup where

$$t_i = \frac{b_i - b^+}{S_{b_i}}.$$

3. Square each *t* statistic and transform it by calculating z_i, where

$$z_i = c + \frac{c^3 + 3c}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855c)}{(10b^2 + 8bc^4 + 1000b)},$$

and

$$a = v_i - .5;\ b = 48a^2;\ c = \sqrt{\left[a \ln\left(\frac{1 + t_i^2}{v_i}\right)\right]}, \text{ where } v_i = n_k - 2.$$

4. Square these z_is and sum to determine *A* as shown in Equation (C1), and reference to the chi-square distribution with $k - 1$ degrees of freedom.

The null hypothesis (i.e., H_0: $\beta_1 = \ldots = \beta_k$) is rejected when $A > h(\alpha)$.

Note

1. A reviewer noted that many of the studies included in our 1987 to 1999 review predate articles concerning the issue of homogeneity of error variance in moderated multiple regression (MMR), and therefore our review might underestimate the attention paid to the assumption. We agree with this comment as it applies to the work by Alexander and colleagues (Alexander & DeShon, 1994; DeShon & Alexander, 1996) and the review by Aguinis and Pierce (1998a). However, as noted above, the issue of homogeneity of error variance in the context of MMR dates back to the work of Gulliksen and Wilks (1950). Also, Dretzke et al.'s (1982) *Psychological Bulletin* article was published before the 1987 lower bound for our review. Thus, although there has been a recent renewed interest in the topic among methodologists, the issue of the homogeneity of error variance assumption in the context of MMR dates back to the 1950s.

References

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

Aguinis, H., & Adams, S.K.R. (1998). Social-role versus structural models of gender and influence use in organizations: A strong inference approach. *Group and Organization Management*, *23*, 414-446.

Aguinis, H., Boik, R. J., & Pierce, C. A. (1998, April). *Estimating the statistical power of differential prediction analysis*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

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., & 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., & 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.

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.

Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, *19*, 91-101.

American Psychological Association (APA). (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association (APA). (1997). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society*, *A160*, 268-282.

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

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

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

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.

Easterling, D. V., & Leventhal, H. (1989). Contribution of concrete cognition to emotion: Neutral symptoms as elicitors of worry about cancer. *Journal of Applied Psychology, 74*, 787-796.

Games, P. A., Winkler, H. B., & Probert, D. A. (1972). Robust tests for homogeneity of variance. *Educational and Psychological Measurement*, *32*, 887-909.

Gartside, P. A. (1972). A study of methods for comparing several variances. *Journal of the American Statistical Association*, *67*, 342-346.

Gibbons, F. X., Helweg-Larsen, M., & Gerrard, M. (1995). Prevalence estimates and adolescent risk behavior: Cross-cultural differences in social influence. *Journal of Applied Psychology*, *80*, 107-121.

Gulliksen, H., & Wilks, S. S. (1950). Regression tests for several samples. *Psychometrika*, *15*, 91-114.

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

Hollenbeck, J. R., Williams, C. R., & Klein, H. J. (1989). An empirical examination of the construct validity of a self-report measure of goal commitment. *Journal of Applied Psychology, 74*, 18-23.

James, G. S. (1951). The comparison of several groups of observations when the ratios of population variances are unknown. *Biometrika*, *38*, 324-329.

Latack, J. C. , Josephs, S. L., Roach, B. L., & Levine, M. D. (1987). Carpenter apprentices: Comparison of career transitions for men and women. *Journal of Applied Psychology, 72*, 393-400.

Mael, F. (1995). Staying afloat: Within group swimming proficiency for Whites & Blacks. *Journal of Applied Psychology*, *80*, 479-490.

Marascuilo, L. A. (1966). Large-sample multiple comparisons. *Psychological Bulletin*, *65*, 280-290.

Melamed, S., Ben-Avi, I., Luz, J., & Green, M. S. (1995). Objective and subjective work monotony: Effects on job satisfaction, psychological distress, and absenteeism in blue collar workers. *Journal of Applied Psychology*, *80*, 29-42.

Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel Psychology*, *48*, 289-308.

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

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

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

Srinivas, S., & Motowidlo, S. J. (1987). Effects of raters' stress on the dispersion and favorability of performance ratings. *Journal of Applied Psychology*, *72*, 247-251.

Starbuck, B., & Mezias, J. (1996). Journal impact ratings. *The Industrial-Organizational Psychologist*, *33*(4), 101-105.

Stewart, G. L., Carson, K. P., & Cardy, R. L. (1996). The joint effects of conscientiousness and self-leadership training on employee self-directed behavior in a service setting. *Personnel Psychology*, *49*, 143-164.

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.

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.

Tubbs, M. E. (1993). Commitment as a moderator of the goal-performance relation: A case for clearer construct definition. *Journal of Applied Psychology, 78*, 86-97.

Wilcox, R. R. (1988). A new alternative to the ANOVA *F* and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology*, *41*, 109-117.

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.

Scott A. Petersen (http://www.dean.usma.edu/bsl/mgtmain/petersen.htm) is a captain in the U.S. Army serving as an instructor of organizational behavior at the U.S. Military Academy. He received an M.A. in industrial/organizational psychology and an M.B.A. from the University of Colorado at Denver. His current research interests include leadership development, team development and assessment, and research methods.

Charles A. Pierce (http://www.montana.edu/wwwpy/cppage.html) is an assistant professor of social and 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.