Heterogeneity of Error Variance and the Assessment of Moderating Effects of Categorical Variables: A Conceptual Review

Herman Aguinis
Charles A. Pierce

Reprinted from *Organizational Research Methods*
Volume 1, No. 3, July 1998, pp. 296-314

Heterogeneity of Error Variance and the Assessment of Moderating Effects of Categorical Variables: A Conceptual Review

HERMAN AGUINIS
University of Colorado at Denver

CHARLES A. PIERCE
Montana State University

Homogeneity of within-subgroup error (residual) variance is a necessary statistical assumption for the appropriate use of moderated multiple regression (MMR) for assessing the effects of categorical moderator variables (e.g., ethnicity, gender). We provide a conceptual review of the homogeneity of error variance assumption in the context of MMR analyses. First, we clarify issues pertaining to the violation of the homogeneity of error variance assumption and differentiate it from the homoscedasticity assumption. Second, we delineate the implications of violating the homogeneity of error variance assumption for organizational theory and practice. Finally, we critically review solutions recently proposed to mitigate the detrimental effects of violating the homogeneity of error variance assumption on conclusions regarding the effects of categorical moderator variables.

The presence of a moderating or interaction effect indicates that the relationship between a predictor X and a criterion Y is not homogenous across values of a third (moderator) variable Z. The estimation of moderating effects is becoming an increasingly critical methodological issue in several organizational disciplines, including human resources management, organizational behavior, and industrial/organizational

Authors' Note: Portions of this article were presented at the meetings of the Rocky Mountain Psychological Association, Boulder, CO, April 1995, and the Society for Industrial and Organizational Psychology, San Diego, CA, April 1996. The research reported in this article was facilitated by a Faculty Grant Award and a Summer Research and Development Stipend to Herman Aguinis from the University of Colorado at Denver. We thank Richard P. DeShon (Michigan State University), Lawrence R. James (University of Tennessee, Knoxville), and members of the Behavioral Science Research Group for their valuable comments on previous drafts. Correspondence and reprint requests should be addressed to Herman Aguinis, College of Business and Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217-3364; e-mail: haguinis@castle.cudenver.edu URL: http://www.cudenver.edu/~haguinis and http://www.montana.edu/wwwpy/cppage.html.

Organizational Research Methods, Vol. 1 No. 3, July 1998 296-314

psychology (Aguinis & Pierce, in press-b; Aguinis & Whitehead, 1997). For example, in the area of staffing decision making (i.e., personnel selection and placement), 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., minority and nonminority). Consequently, if a moderator such as ethnicity is found, there is *differential prediction* or *predictive bias*, and the preemployment test is considered to be biased for certain subgroups (Bartlett, Bobko, Mossier, & Hannan, 1978; Cleary, 1968; Society for Industrial and Organizational Psychology [SIOP], 1987).

We offer a conceptual review of an often overlooked yet pervasive issue regarding the estimation of moderating effects of categorical variables; that is, the violation of the homogeneity of within-subgroup error variance assumption in the context of moderated multiple regression analyses. First, however, we (a) briefly review the use of moderated multiple regression (MMR) to estimate moderating effects of categorical variables, (b) illustrate the assessment of moderating effects of categorical variables with the case of differential prediction research, and (c) provide a brief overview of statistical power problems with MMR to assess moderating effects.

Use of Moderated Multiple Regression to Estimate Moderating Effects of Categorical Variables

MMR is a consensually accepted statistical technique for estimating moderating effects of categorical variables such as ethnicity and gender in the organizational sciences (e.g., management, applied psychology; Cortina, 1993; Sackett & Wilk, 1994). For instance, Russell and Bobko (1992) state that "[a] simple count of the number of studies examining moderator effects in major applied psychology journals indicates that moderated regression analysis is the preferred statistical procedure for detecting interaction effects" (p. 336). MMR consists of forming a least squares regression equation (Cohen & Cohen, 1983; Saunders, 1956; Zedeck, 1971). Let Y be a continuous criterion variable (i.e., performance scores), let X be a continuous predictor variable (i.e., preemployment test scores), and let Z be a categorical predictor variable hypothesized to be a moderator (e.g., gender dummy coded 1 = males and 2 = females). Equation 1 shows the sample-based least squares regression that tests the additive model for predicting Y from X, Z, and the interaction between X and Z (i.e., moderating effect of Z) represented by the $X \times Z$ product term.

$$\hat{Y} = a + b_1X + b_2Z + b_3X{\times}Z, \tag{1}$$

where $\hat{Y}$ is the predicted value for Y, a is the least squares estimate of the intercept of the surface of best fit, b_1 is the least squares estimate of the population regression coefficient for X, b_2 is the least squares estimate of the population regression coefficient for Z, and b_3 is the least squares estimate of the population regression coefficient for the product term that carries information about the interaction between X and Z (Cohen & Cohen, 1983). Rejecting the null hypothesis of $\beta_3 = 0$ indicates the presence of a moderating or interaction effect. Stated differently, rejecting this null hypothesis indicates that the regression of Y on X is unequal across levels of Z (e.g., minority and nonminority subgroups, male and female subgroups).

Moderating Effects of Categorical Variables: The Case of Differential Prediction

In the area of human resources management, the presence of differential prediction by ethnicity or gender is ethically and legally problematic. Differential prediction indicates that the test-criterion relationship is stronger for one ethnic- or gender-based subgroup than for another. For example, assume there is gender-based differential prediction such that there is a stronger relationship (i.e., greater $b_{Y.X}$ value or steeper slope) for the female subgroup than for the male subgroup (Dunbar & Novick, 1988). In this case, because selection decisions are typically made considering scores in the middle and high portions of the X range, using one common regression equation implies that scores for females will be underpredicted and scores for males will be overpredicted. This situation, similar to Dunbar and Novick's actual results regarding nine clerical specialties in the U.S. Marine Corps, leads to predictions biased in favor of members of the male subgroup and against members of the female subgroup.

Statistical Power Problems with Moderated Multiple Regression to Assess Moderating Effects

MMR is regularly used in testing hypotheses regarding moderating effects of dichotomous variables such as ethnicity (i.e., minority and nonminority subgroups) and gender (i.e., female and male subgroups) (e.g., Cortina, Doherty, Schmitt, Kaufman, & Smith, 1992; Hattrup & Schmitt, 1990; Houston & Novick, 1987; Schmitt, Hattrup, & Landis, 1993). Despite its popularity, numerous researchers have expressed the concern that MMR may often lead to erroneous conclusions (Aguinis & Stone-Romero, 1997; Bobko & Russell, 1990, 1994; Linn, 1983; Linn & Hastings, 1984; MacCallum & Mar, 1995). For example, many theory-based, sound hypotheses involving moderated relationships are frequently not supported (Zedeck, 1971).

In response to the failures to detect hypothesized moderator variables, recent research has investigated the accuracy of MMR to estimate moderating effects under various conditions of (a) predictor and criterion variable distributions and operationalizations, (b) predictor-moderator intercorrelation (i.e., multicollinearity), and (c) sample size (Bobko & Russell, 1994; McClelland & Judd, 1993; Stone-Romero, Alliger, & Aguinis, 1994). Overall, this body of research suggests that MMR-based tests of moderator variable hypotheses are, for the most part, conducted at very low levels of statistical power (see Aguinis, 1995, for a review). Stated differently, null findings regarding moderating effects may often be due to Type II errors (i.e., erroneous dismissals of population moderating effects). Consequently, because of the concern that researchers may erroneously fail to reject false null hypotheses regarding moderating effects, some solutions have been advanced in an attempt to prevent researchers from mistakenly discarding models that encompass moderator variables (Aguinis, Bommer, & Pierce, 1996; Aguinis & Pierce, in press-a; Aguinis, Pierce, & Stone-Romero, 1994; Sackett & Wilk, 1994).

An additional factor that influences MMR-based conclusions regarding the operation of moderator variables is the violation of the homogeneity of within-subgroup error variance assumption. When this statistical assumption is violated, the estimation of moderating effects and the assessment of differential prediction becomes problematic: Researchers may commit a Type I or a Type II statistical error, depending on the specific sample and population characteristics. Thus, researchers may discover a false moderator (Type I error) or erroneously dismiss a model including a moderator variable (Type II error). Committing Type I or Type II statistical errors has consequential effects for theory building as well as for organizational practices. More specifically, committing these errors as a consequence of heterogeneity of within-subgroup error variance implies that decisions regarding the use of selection and placement tests may be incorrect: Unbiased (i.e., differential prediction-free) tests may be incorrectly judged as biased (Type I error), and biased (i.e., differential prediction-laden) tests may be incorrectly judged as unbiased (Type II error).

The issue of comparing error variances across subgroups in the context of moderated regression models is not new. 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, $\sigma_{e(i)}$) should be conducted before testing for differences in slopes across subgroups (which is at present accomplished using MMR). 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 (i.e., $\sigma_{e(1)} \neq \sigma_{e(2)}$, in the case of two subgroups). However, Gulliksen and Wilks's 48-year old recommendation does not seem to have had its intended effect. A possible reason is that, thus far, articles addressing this issue have had a more technical focus and are perhaps not accessible to most organizational researchers (e.g., Alexander & DeShon, 1994; Dretzke, Levin, & Serlin, 1982; Gulliksen & Wilks, 1950).

The fact that many organizational science scholars are not aware of the homogeneity of within-subgroup error variance assumption is illustrated by a review of the extent to which this assumption is violated in differential prediction research. Based on a review of articles published or referenced in *Journal of Applied Psychology* and *Personnel Psychology* between 1980 and 1993, and the validity data base published in the *Journal of Business and Psychology* in 1992 (Landy, 1992), DeShon and Alexander (1994b) concluded that 39 MMR tests reported in a total of 20 studies violated this assumption. Furthermore, in all but one of these 20 studies, the subgroup with the largest sample size was accompanied by the largest error variance. As described in a later section of the present article, this pairing of *n* and error variance increases the likelihood of committing a Type II error in testing for differential prediction.

The frequent violation of the homogeneity of within-subgroup error variance assumption as well as the consequential implications of this violation for theory building and staffing decision making sparked an increased interest in the consequences of violating this assumption (e.g., Alexander & DeShon, 1994; DeShon & Alexander, 1994b; Hsu, 1994). However, despite what seems to be an increased awareness regarding issues surrounding homogeneity of within-subgroup error variance, this assumption is often confused with the distinct homoscedasticity assumption. For instance, Stone and Hollenbeck (1989) found differences in residual variances across two subgroups (i.e., heterogeneity of within-subgroup error variance) and concluded, instead, that the homoscedasticity assumption had been violated. Also, some recently proposed solutions to mitigate the effects of heterogeneity of error variance (Hsu, 1994) do not seem to be appropriate and may be more detrimental than beneficial.

Accordingly, the objective of the present article is to provide organizational researchers with a conceptual review of the homogeneity of within-subgroup error variance assumption in the context of MMR, which is a consensually accepted technique for testing hypotheses regarding moderating effects of categorical variables. The more specific goals of the present article are to (a) clarify the nature of the homogeneity of within-subgroup error variance assumption and differentiate it from the homoscedasticity assumption, (b) illustrate practical implications of violating this assumption in the context of personnel selection and placement decision making in terms of Type I and Type II statistical errors, and (c) critically discuss solutions recently proposed to mitigate the detrimental effects of violating this statistical assumption on conclusions regarding the operation of moderator variables.

Homogeneity of Error Variance Assumption: Clarifications

The statistical assumptions of ordinary least squares (OLS) regression include (a) independence of observations, (b) normality of population scores, and (c) homoscedasticity (i.e., the conditional variance of a criterion variable Y is the same irrespective of values or levels of a predictor variable X or, stated differently, an equal spread of observed Y scores about predicted Y [$\hat{Y}$] scores across values or levels of X) (Cohen & Cohen, 1983; Pedhazur, 1982).

An additional statistical assumption for the use of moderated multiple regression analysis for testing moderating effects of categorical variables is homogeneity of within-subgroup error (residual) variance (Kendall & Stuart, 1979). Homogeneity of error variance exists when the variance in Y that remains after predicting Y from X is equal across moderator-based subgroups (e.g., $\sigma^2_{e(i)}$ or $\sigma^2_{Y_1-\hat{Y}_1}$ for males, $\sigma^2_{e(2)}$ or $\sigma^2_{Y_2-\hat{Y}_2}$ for females). This equality allows for the overall residual Y variance to be estimated from the mean square residual term (i.e., σ^2_e or $\sigma^2_{Y-\hat{Y}}$). In other words, homogeneity of within-subgroup error variance is achieved when the variance in Y that is unaccounted for by X is equal across moderator-based subgroups. In such situations, $\sigma^2_{Y_1-\hat{Y}_1} = \sigma^2_{Y_2-\hat{Y}_2}$. Note that this assumption is equivalent to the perhaps more familiar homogeneity of variance assumption in the context of analysis of variance (ANOVA) models.

Relationship Between the Homoscedasticity and Homogeneity of Error Variance Assumptions

Although the homoscedasticity and homogeneity of error variance assumptions have been treated as synonymous (e.g., Stone & Hollenbeck, 1989), it deserves noting that they are not equivalent. The homoscedasticity assumption applies to all OLS regression models (including MMR), whereas the homogeneity of error variance assumption applies only to MMR models. The homoscedasticity assumption appears to be well understood and is described in most statistics and research methods textbooks in the organizational sciences (e.g., Berry & Feldman, 1985). However, researchers should not assume that meeting the more familiar homoscedasticity assumption implies that the homogeneity of error variance assumption is also satisfied.

In the presence of homoscedasticity, the homogeneity of error variance assumption may or may not be satisfied.

The error variance for each of the moderator-based subgroups is

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

where $\sigma^2_{Y(i)}$ and $\rho_{XY(i)}$ are the Y variance and the $X - Y$ correlation in each moderator-based subgroup, respectively. Thus, the error variance for Subgroup 1 can be expressed as

$$\sigma^2_{e(1)} = \sigma^2_{Y(1)}\,(1 - \rho^2_{XY(1)}) = V\,[E\,(Y_1 - \hat{Y}_1) \mid X_1] + E\,[V(Y_1 - \hat{Y}_1) \mid X_1] \tag{3}$$

where E is the expectation (i.e., mean), V is the variance, and $V\,(Y_1 - \hat{Y}_1) \mid X_1$ refers to the error variance at a given X_1.

Subgroup homoscedasticity. The homoscedasticity assumption for Subgroup 1 can be verified by examining the Y_1 on X_1 regression model. Homoscedasticity is satisfied if residuals (i.e., $Y_1 - \hat{Y}_1$) are similarly distributed across various points of X_1. If one assumes that Y_1 is similarly distributed around $\hat{Y}_1$ (i.e., the mean of errors equals zero), Equation 3 reduces to

$$\sigma^2_{e(1)} = \sigma^2_{Y(1)}\,(1 - \rho^2_{XY(1)}) = E\,[V(Y_1 - \hat{Y}_1) \mid X_1]. \tag{4}$$

For Subgroup 2, a similar expression for within-subgroup error variance can be written as

$$\sigma^2_{e(2)} = \sigma^2_{Y(2)}\,(1 - \rho^2_{XY(2)}) = E\,[V(Y_2 - \hat{Y}_2) \mid X_2]. \tag{5}$$

Similar to Subgroup 1, homoscedasticity can be verified for Subgroup 2 by examining the Y_2 on X_2 regression model. Homoscedasticity is satisfied if residuals (i.e., $Y_2 - \hat{Y}_2$) are similarly distributed across various points of X_2.

Overall homoscedasticity. To assess whether the homoscedasticity assumption is satisfied for the overall regression model (what we label *overall homoscedasticity*), one needs to examine the Y on X regression model including all scores (i.e., Subgroups 1 and 2 combined). The homoscedasticity assumption is satisfied if the $Y - \hat{Y}$ residual scores are similarly distributed across various points of the X scale.

Heterogeneity of within-subgroup error variance in the presence of subgroup and overall homoscedasticity. Meeting the homoscedasticity assumption does not imply that the homogeneity of within-subgroup error variance assumption is also satisfied. For example, if the Y variances are equal across the two subgroups (i.e., $\sigma^2_{Y(1)} = \sigma^2_{Y(2)}$), and there is a stronger $X - Y$ relationship for one subgroup than the other (e.g., $\rho_{XY(1)} < \rho_{XY(2)}$), Equation 2 indicates that the error variances must differ across subgroups (i.e., $\sigma^2_{e(1)} > \sigma^2_{e(2)}$). Similarly, if the correlation coefficients are equal across subgroups but Y variances differ, this situation also leads to a systematic violation of the homogeneity of within-subgroup error variance assumption, even in the presence of subgroup and overall homoscedasticity. Finally, both the Y variances and correlations may differ across subgroups, which also results in the violation of the within-subgroup error variance assumption (unless, as shown in Equation 2, the difference in Y variances is precisely counterbalanced by the difference in correlation coefficients).

Any of the three conditions mentioned above can exist even in the presence of homoscedasticy. To illustrate that, even in the presence of homoscedasticity, differences in correlation coefficients across subgroups may lead to the violation of the homogeneity of within-subgroup error variance assumption, we offer a graphic illustration. Figure 1 shows a scatterplot of a hypothetical overall $X - Y$ relationship for males and females, in which X (a predictor variable) and Y (a criterion variable) are both continuous in nature. The data shown in Figure 1 are homoscedastic. That is, the data points shown in the graph are similarly distributed throughout the regression line.

Assume that one is interested in testing whether there is differential prediction for gender-based subgroups. Then, one would test whether the dichotomous variable gender (Z) ($Z = 1$, males; $Z = 2$, females) moderates the relationship between X and Y shown in Figure 1. Thus, two separate $X - Y$ scatterplots are drawn, one for the relationship between X and Y when $Z = 1$ (males; shown in Figure 2) and one for the relationship between X and Y when $Z = 2$ (females; shown in Figure 3). As with the data plotted in Figure 1, each of the data sets in Figures 2 and 3 is also homoscedastic. That is, each graph shows that the data points are similarly distributed throughout the regression line. However, the amount of error variance present when Y is predicted from X is clearly not equivalent for the two moderator-based subgroups (i.e., for males and females). Although the data points are similarly distributed throughout each of the two regression lines in Figures 2 and 3, the average deviation of the data points from the line is larger for males (Figure 2) than for females (Figure 3). Stated differently, the amount of error variance (σ_e^2) is not equivalent across the two moderator-based subgroups; that is, it is larger for the male (Figure 2) than for the female (Figure 3) subgroup (i.e., $\sigma^2_{e(1)} > \sigma^2_{e(2)}$).

In sum, Figures 1 to 3 clarify that even in the presence of subgroup and overall homoscedasticity, heterogeneity of within-subgroup error variance can occur. These figures illustrate that the homoscedasticity and homogeneity of error variance assumptions should not be treated synonymously and that satisfying homoscedasticity does not necessarily imply that error variances are homogeneous across moderator-based subgroups.

Are Within-Subgroup Error Variances Homogeneous or Heterogeneous?

Figures 2 and 3 illustrate a situation in which the homogeneity of within-subgroup error variance assumption seems to be violated. However, how can researchers more precisely determine whether a specific data set in hand violates the homogeneity of within-subgroup error variance assumption? As is the case with tests of other statistical assumptions, researchers have the choice of using formal statistical tests or heuristic guidelines (Weinzimmer, Mone, & Alwan, 1994).

First, regarding formal statistical tests, recall that the homogeneity of within-subgroup error variance assumption in MMR is equivalent to the homogeneity of variance assumption in ANOVA. Consequently, Bartlett's (1937) homogeneity of variance test could be directly modified from an ANOVA to an MMR context (cf. DeShon & Alexander, 1996). This can be done by simply replacing unconditional subgroup variances in the dependent variable with subgroup error variances (i.e., $\sigma^2_{e(i)}$; cf. Equation 2) in Bartlett's formulae. However, Bartlett's test is adversely affected by deviations from normality (Games, Winkler, & Probert, 1972). Consequently, a

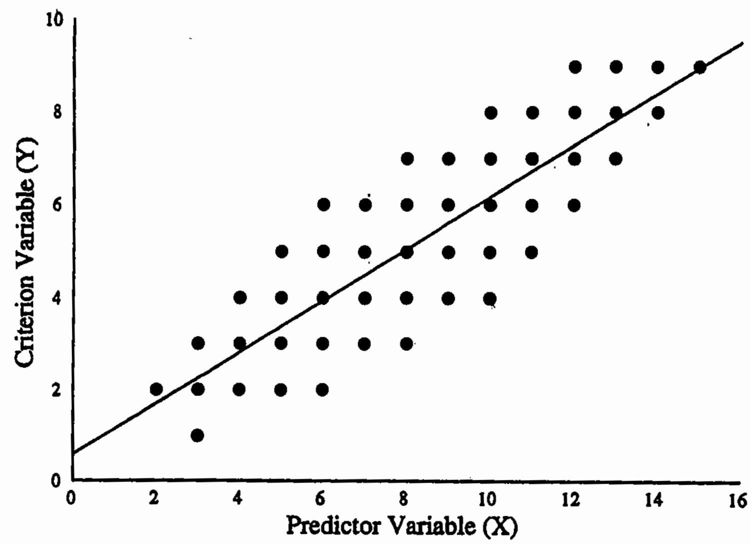

Figure 1: Scatterplot of a Hypothetical Relationship Between X and Y for Males and Females Combined

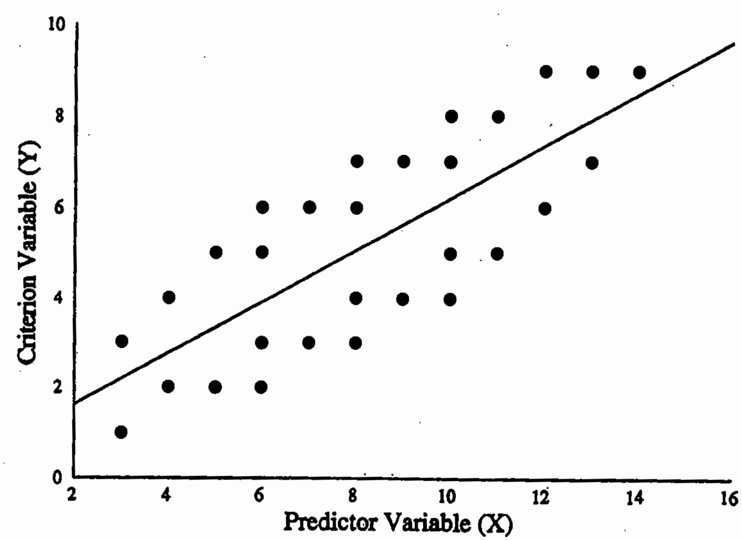

Figure 2: Scatterplot of the Same Hypothetical Relationship Between X and Y for Males Only

rejection of a null hypothesis of homogeneity of within-subgroup error variance may be due to deviations from normality and not from violating the homogeneity assumption.

A second alternative is to use heuristics or empirically derived rules of thumb. DeShon and Alexander (1996) conducted a Monte Carlo study regarding the accuracy of MMR to estimate the moderating effect of a categorical variable. They manipulated thousands of parameter values for subgroup sample sizes, heterogeneity of within-subgroup error variance, departures from Y normality within each subgroup, and within-subgroup correlation between predictor and criterion scores. Based on the

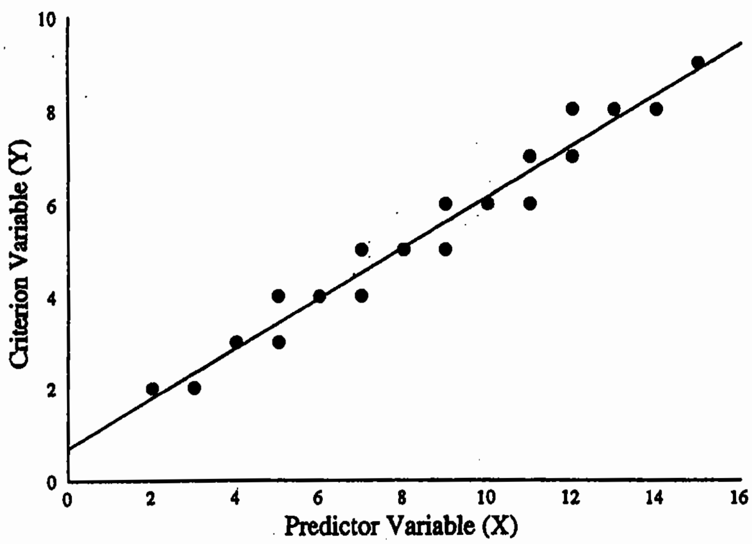

Figure 3: Scatterplot of the Same Hypothetical Relationship Between X and Y for Females Only

results of this large-scale simulation study, the general conclusion was that the F statistic used in MMR begins to be adversely affected when the error variance in one subgroup is approximately 1.5 times larger than the error variance in another subgroup. Thus, this empirically derived 1.5 rule of thumb can be used for determining whether heterogeneity of within-subgroup error variance is an artifact likely to affect MMR-based conclusions.

In our view, a combination of formal procedures (e.g., Bartlett's test) and heuristics (i.e., the 1.5 DeShon & Alexander rule of thumb) optimizes decision making regarding the potential impact of violating the homogeneity of within-subgroup error variance assumption. Moreover, we recommend that both procedures be used and that convergence be sought. Of course, if conclusions based on both the formal and heuristic methods are congruent, then MMR users would be more confident about their decision.

Violating the Homogeneity of Error Variance Assumption: Implications

Satisfying the homogeneity of within-subgroup error variance assumption may be difficult in many circumstances. For instance, if a researcher tests a false null hypothesis of no differential prediction (i.e., there actually is differential prediction in the population of scores), the assumption of homogeneity of within-subgroup error variance is almost certainly violated.

Assuming the following equalities,

$$\sigma^2_{Y_1} = \sigma^2_{Y_2};\ \sigma^2_{X_1} = \sigma^2_{X_2}, \tag{6}$$

and given that

$$\beta_{Y.X(i)} = \rho_{XY(i)}\left(\frac{\sigma_{Y(i)}}{\sigma_{X(i)}}\right), \tag{7}$$

the null hypothesis of equal subgroup slopes is identical to the null hypothesis of equal subgroup correlation coefficients. Thus, given Equation 2 and Equation 6, the assumption will always be violated when the null hypothesis is false (Alexander & DeShon, 1994).

A similar situation occurs even when the assumed equalities shown in Equation 6 are relaxed. Assuming that

$$\sigma^2_{Y_1}/\sigma^2_{X_1} = \sigma^2_{Y_2}/\sigma^2_{X_2}, \tag{8}$$

Equation 7 indicates that differences in slopes are still identical to differences in correlations across subgroups.

If Equation 6 or Equation 8 is true, Equation 2 indicates that the error terms are necessarily heterogenous. Note that the only, and perhaps rare, situation in which $\sigma^2_{e(i)}$s would not differ across subgroups when $\rho_{XY(i)}$s differ is when Equation 8 is true, and the difference in Y variances across subgroups precisely offsets the difference in $X - Y$ correlations across subgroups.

This situation creates contradictory effects on statistical power. On one hand, larger population effect sizes (i.e., differences between slopes or correlation coefficients across subgroups) results in greater power. At the same time, however, a larger effect size results in a greater violation of the homogeneity of error variance assumption which, in turn, results in lower power. As we describe below, the overall impact of these contradictory forces is a decrease in power. This decrease is particularly noticeable when there is an inverse pairing of sample size with effect size, that is, the larger subgroup n paired with the smaller correlation coefficient.

In general, violating the homogeneity of error variance assumption (i.e., having error variance heterogeneity) has important practical implications regarding the use of MMR for moderator variable detection with respect to both Type I error and statistical power rates.

Effects on Type I Error Rates: Finding "False" Moderators

Dretzke et al. (1982) and DeShon and Alexander (1996) conducted Monte Carlo investigations to ascertain the effects of violating the homogeneity of error variance assumption on Type I error rates. In the less typical validation study when sample sizes are equal across moderator-based subgroups (cf. Hunter, Schmidt, & Hunter, 1979; Hunter, Schmidt, & Rauschenberger, 1984), Dretzke et al. ascertained that Type I error rates associated with a null hypothesis of equal slopes across subgroups (i.e., $\beta_{Y.X(1)} = \beta_{Y.X(2)}$) do not seem to be artificially inflated. Alternatively, in the more typical situation of unequal subgroup sample sizes, error variance heterogeneity can result in an inflated Type I error rate when testing for moderating effects. For example, with subgroup ns (rs) of 50 (.25) and 100 (.75), the actual Type I error probability using an ordinary F test was .18 for a nominal α of .05 (Dretzke et al., 1982).

Consistent with the previous illustration, Type I error rate inflation was found to be most noticeable when the smaller subgroup sample size was paired with the larger residual variance (i.e., the smaller subgroup $X - Y$ correlation). Note, however, that Dretzke et al.'s (1982) simulation held the X variance constant across subgroups (but not the Y variance across subgroups; this is why subgroup correlations differed). DeShon and Alexander (1996) showed that Dretzke et al.'s results regarding the robustness of MMR in equal subgroup n conditions should be qualified. More precisely, DeShon and Alexander's empirical results indicate that Dretzke et al.'s conclusions hold only when the X variance is equal across subgroups. However, when the X variance is moderately unequal across subgroups, heterogeneity of error variance leads to overly conservative Type I error rates.

In sum, heterogeneity of error variance is likely to affect Type I error rates and lead to erroneous conclusions regarding moderating effects. First, Type I error rates are likely to be artificially inflated when sample sizes are unequal across subgroups. This is most noticeable when the smaller subgroup sample size is paired with the smaller validity coefficient (i.e., the larger error variance). Second, Type I error rates are also affected under conditions of equal subgroup sample sizes. Type I error rates become overly conservative when the X variance is dissimilar across subgroups. Thus, MMR users should be especially aware of inaccurate Type I error rates when (a) sample sizes are unequal across subgroups (resulting in overly liberal Type I error rates); and (b) sample sizes are equal across subgroups and X variances are unequal across subgroups (resulting in overly conservative Type I error rates).

Effects on Statistical Power: Incorrectly Dismissing Moderator Variables

To address the issue of statistical power, which was not examined in Dretzke et al.'s (1982) simulation, Alexander and DeShon (1994) conducted a Monte Carlo study and ascertained that under unequal subgroup sample size conditions, 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. This result was consistent for situations involving $k = 2$ or more subgroups. The ordinary F test for assessing a moderating effect with MMR is thus not robust to violations of the homogeneity of within-subgroup error variance assumption. However, power levels do not suffer as much when sample sizes are equal across moderator-based subgroups.

This specific scenario in which the subgroup with the larger n is paired with the smaller validity coefficient is the most typical situation in validation research in a variety of organizational settings (e.g., industrial, educational, and military) (Hunter, Schmidt, & Hunter, 1979; Valentine, 1977). Typically, the majority subgroup (e.g., Whites, males) is more numerous than the minority subgroup (e.g., African Americans, females), and the majority subgroup presents a validity coefficient that is smaller than that of the minority subgroup. Extensive meta-analytic research by Hunter, Schmidt, and colleagues has documented this pairing of subgroup n and validity coefficient in the late 1970s and early 1980s, and several more recent studies indicate that this situation is still pervasive (DeShon & Alexander, 1994b; Hattrup & Schmitt, 1990).

In sum, Alexander and DeShon's (1994) simulation results demonstrate that this most typical situation in validation research in which the subgroup with the larger n presents the smaller validity coefficient is likely to lead to a Type II error by incorrectly dismissing moderating effects. Thus, it is not surprising that the empirical evidence accumulated thus far suggests that differential prediction on the basis of, for example, cognitive abilities tests is not supported for the major ethnic subgroups (SIOP, 1987).

Summary: Conditions Leading to Incorrect Decision Making Regarding the Effects of Categorical Moderator Variables

Simulation work by Dretzke et al. (1982), Alexander and DeShon (1994), and DeShon and Alexander (1996) lead to two major conclusions regarding the impact of violating the homogeneity of within-subgroup error variance assumption on MMR-based inferences, especially for the typical validation study in which sample sizes are not equal across moderator-based subgroups. First, with respect to Type I errors, researchers are more likely to erroneously conclude that a moderating effect exists when the smaller subgroup sample size is paired with the larger residual variance (i.e., the smaller $X - Y$ correlation coefficient). Second, with respect to Type II errors, researchers are more likely to erroneously dismiss moderating effects when the larger subgroup sample size is paired with the larger residual variance (i.e., the smaller $X - Y$ correlation coefficient). Finally, reviews of the personnel selection literature suggest that the majority subgroup (i.e., larger sample size) typically has the smaller validity coefficient. Consequently, committing a Type II error may be more likely and frequent than committing a Type I error when MMR is used to test hypotheses regarding categorical moderator variables in validation research.

Alleviating Heterogeneity of Error Variance Effects: Solutions Recently Proposed

Inverse Data Transformation

Given the aforementioned implications of violating the homogeneity of within-subgroup error variance assumption, Hsu (1994) suggested a possible solution to mitigate heterogeneity and, seemingly, improve the assessment of moderating effects using MMR. Hsu asserted that data transformations are routinely conducted in social science research to meet the assumptions required by various statistical tests such as ANOVA and other methods based on the general linear model. Hence, if there exists a data transformation that alleviates the heterogeneity of within-subgroup error variance problem, this transformation should be used. According to Hsu, the inverse transformation of the criterion (i.e., $Y' = 1/Y$) provides such a solution. By obtaining Y' and conducting the subsequent MMR analysis on Y' (rather than on the original Y), heterogeneity of within-subgroup error variance is eliminated. To illustrate this contention, Hsu provided an example in which there was heterogeneity of within-subgroup error variance. After the inverse transformation was implemented on the criterion, the error variances became homogenous. Thus, subsequent moderator analysis results based on MMR were seemingly more meaningful and trustworthy because the assumption was not violated.

Inverse Data Transformation: A Remedy That Kills the Patient

We agree that the heterogeneity of within-subgroup error variance problem is pervasive in organizational science research. Moreover, the implications of this violation, as described above, are serious and consequential. However, the inverse transformation solution advocated by Hsu (1994) suffers from a limitation: It reduces heterogeneity of within-subgroup error variance but, in many situations, it may also eliminate the moderating effect. Stated differently, once the criterion Y has been transformed to $Y' = 1/Y$, the probability of detecting population moderating effects may be reduced to values close to zero.

The inverse transformation equates within-subgroup error variances by equating Y variances across subgroups. Analytically, Equation 2 shows that, once Y variances are equal, equating the error variances across moderator-based subgroups can only be conducted at the expense of also equating the $X - Y$ correlations across subgroups (i.e., eliminating the moderating effect). That is, if the variance of Y scores is constant across subgroups, the only remaining variable besides $\sigma^2_{e(i)}$ in Equation 2 is $\rho_{XY(i)}$. Thus, Equation 2 shows that the only way to achieve within-subgroup error variance homogeneity (i.e., $\sigma^2_{e(1)} = \sigma^2_{e(2)}$) is to equate $\rho_{XY(i)}$ across subgroups.

This type of transformation defeats the purpose of improving the accuracy in estimating moderating effects. More precisely, in certain situations, this transformation may guarantee that even if there is differential prediction, it will not be detected. Thus, even though a researcher is more confident that the homogeneity of within-subgroup error variance assumption is met, using MMR on the newly created Y' variable may lead to the sample-based conclusion that there is no moderating effect because the sample-based $X - Y$ correlations are now equal for all values of Z, despite the fact that the effect may be present (and of substantial magnitude) in the population.

The aforementioned analytic explanation that the moderating effect is often eliminated using the inverse transformation can be illustrated using Hsu's (1994) data. Hsu presented the example of a continuous criterion variable Y, a dichotomous predictor X, and a dichotomous moderator Z. Y was the number of tasks accomplished per hour of work after 1 week of training, X was training method ($X = 1$: Training Method 1; $X = 2$: Training Method 2), and Z was experience level ($Z = 1$: high experience, $Z = 2$: low experience). Illustrative data were available for 12 workers (three in each of the four cells of the design). These data are reproduced in Table 1.

Table 1 shows Hsu's (1994) original data for each cell together with each cell's variance and mean. Also, Table 1 shows the original criterion data (Y) and the criterion data resulting from the inverse transformation (i.e., $Y' = 1/Y$). Based on these data, we computed the Y variance, the error variance, and the correlation coefficient between training method and the criterion for each of the two levels of the moderator variable Z (i.e., experience level).

An initial perusal of the original data suggests that there may be an interaction between training method and degree of experience. More precisely, the impact of training method seems to be greater for high-experienced than for low-experienced workers. A more formal test of this interaction reveals that, indeed, training method and degree of experience interact in affecting the number of tasks performed per hour, $F(1, 8) = 3.63$, $p = .093$ (Hsu, 1994, Table 2, p. 223). Note, however, that the p value does not reach the traditional .05 level of significance due to the unusually small sample size used in Hsu's illustration.

Table 1 also includes the transformed criterion data. As a result of the inverse transformation, Y variances, residual variances, and correlation coefficients between training method and the criterion are identical across experience-based subgroups. A formal test of the interaction shows that, indeed, the transformation eliminated the effect, $F(1, 8) = .00$, $p = 1.00$ (Hsu, 1994, Table 2, p. 223).

Table 1
Illustrative Data Reproduced from Hsu: Number of Tasks Performed per Hour (Y) as a Function of Training Method (X) and Trainee's Experience Level (Z)

	Original Data (Y) *Experience Level*		*Transformed Data (Y' = 1/Y)* *Experience Level*	
	High	*Low*	*High*	*Low*
Training method 1	3.33	1.11	.30	.90
	4.00	1.17	.25	.85
	5.00	1.25	.20	.80
	$S^2 = .84^2$	$S^2 = .07^2$	$S^2 = .05^2$	$S^2 = .05^2$
	$M = 4.111$	$M = 1.179$	$M = .25$	$M = .85$
Training method 2	5.00	1.25	.20	.80
	6.66	1.33	.15	.75
	10.00	1.43	.10	.70
	$S^2 = 2.546^2$	$S^2 = .089^2$	$S^2 = .05^2$	$S^2 = .05^2$
	$M = 7.222$	$M = 1.337$	$M = .15$	$M = .75$

Source. Hsu (1994, Table 1, p. 222).
Note. For original data, when experience level is high, $S_Y^2 = 5.78$, $S_e^2 = 2.87$, and $r_{XY} = .71$; when experience level is low, $S_Y^2 = .013$, $S_e^2 = .005$, and $r_{XY} = .77$. For transformed data, when experience level is high and low, $S_Y^2 = .005$, $S_e^2 = .002$, and $r_{XY} = .78$.

In sum, because the inverse data transformation proposed by Hsu (1994) may lead MMR users to commit a Type II error and incorrectly dismiss population moderating effects, we now turn to alternative data analytic strategies that can be used to test these hypotheses in the presence of within-subgroup error variance heterogeneity.

Alternative Methods for Estimating Moderating Effects and Assessing Differential Prediction

Gulliksen and Wilks (1950) recommended against the use of inequality of slope tests across subgroups in conditions of error variance heterogeneity. However, they did not discuss any alternative procedures that could be used in these situations. Nevertheless, in the presence of error variance heterogeneity, alternative methods (i.e., procedures other than the ordinary t or F test in MMR) can be implemented. Such procedures include (a) nonparametric methods that do not require equality of error variances for moderator-based subgroups, and (b) parametric methods that include direct or indirect procedures and approximations for correcting the degrees of freedom associated with the more typical t and F parametric tests. The use of alternative nonparametric and parametric tests is appropriate in situations involving heterogeneity of error variance because, especially when sample sizes differ across subgroups, MMR-based results cannot be trusted in terms of Type I or Type II error rates.

Nonparametric methods**.** A nonparametric statistical test available to replace MMR in situations with heterogeneity of error variance is Marascuilo's (1966) U statistic. This statistic approximates a chi-square distribution and is similar to the ordinary F statistic with the exception that it uses separate error variance estimators for each of the k moderator-based subgroups (i.e., $\sigma^2_{e(1)}, \sigma^2_{e(2)}, \ldots \sigma^2_{e(k)}$) (see Dretzke et al., p. 377, and Marascuilo, 1966, for determining degrees of freedom, and for formulae and test procedures).

In general, nonparametric techniques such as Marascuilo's U and rank-transformations tests result in lower statistical power rates than do parametric tests. Thus, parametric tests are typically preferred over nonparametric methods (e.g., Olejnik & Algina, 1987). Accordingly, we now turn to parametric alternatives.

Parametric methods. Less traditional parametric methods that directly or indirectly correct the degrees of freedom associated with more traditional tests are (a) Welch-Aspin's F approximation (F^*) (Aspin, 1948; Welch, 1938), (b) James's second-order approximation (J) (DeShon & Alexander, 1994a; James, 1951), and (c) Alexander and colleagues' normalized-t approximation (A) (Alexander & Govern, 1994; DeShon & Alexander, 1996).

First, the F^* statistic approximates an F distribution. F^* is similar to Marascuilo's U statistic in that it also uses separate error variance estimators for each of the k moderator-based subgroups. However, whereas U is asymptotically distributed, F^* is based on a finite degrees of freedom (see Dretzke et al., 1982, p. 378, for formulae and test procedures). Second, the J statistic, originally developed for testing the equality of k independent means in the presence of heterogeneity of variance, was adapted by DeShon and Alexander (1994a) to test for the equality of regression slopes. The computation of J entails calculating a U statistic and then correcting the degrees of freedom used to reference U to the chi-square distribution (see DeShon & Alexander, 1994a, 1996, for formulae and test procedures). Finally, the A statistic approximates a chi-square distribution with $k - 1$ degrees of freedom (k is the number of moderator-based subgroups) and is based on a normalizing transformation of the t statistic (see DeShon & Alexander, 1994a, 1996, for formulae and test procedures). FORTRAN and SAS computer programs are available for the computation of F^*, J, and A (DeShon & Alexander, 1994a, 1996).

Relative performance of the F^*, J, and A statistics. DeShon and Alexander (1996) empirically compared the performance of the F^*, J, and A statistics under various conditions of equal and unequal subgroup sample sizes, equal and unequal error variances, departures from Y normality within each subgroup, and $X - Y$ correlations ranging from .10 to .90. The comparison of the relative performance of these tests yielded the following general results and conclusions. First, under conditions of error variance heterogeneity, J had a slight performance advantage over F^* and A when sample sizes were small (i.e., 10 to 25). Second, the performance of F^* worsened as the number of moderator-based subgroups increased from two to eight. Third, the performance of A and J was very similar across various levels of error variance heterogeneity, sample size, and number of subgroups simulated. Fourth, A was more robust to violations of Y normality within subgroups as compared to F^* and J. Fifth, unlike F^*, A was not adversely affected by an increase in the number of subgroups. Sixth, computation of A is simpler than the computation of J. Based on the aforementioned six considerations, DeShon and Alexander's (1996) general conclusion and recommendation is that inferences regarding differential prediction under conditions of error variance heterogeneity should be based on the normalized-t approximation A statistic.

Summary and Conclusions

MMR is a consensually accepted method for assessing moderating effects of categorical variables in the organizational sciences. However, heterogeneity of within-subgroup error variance is an artifact that may lead to incorrect MMR-based conclusions regarding the operation of moderator variables. Nevertheless, a literature review of the extent to which this assumption is violated indicates that management and applied psychology researchers are not aware of the issue. Accordingly, our intent was to raise awareness regarding the distinct homogeneity of error variance assumption, the consequences of violating this assumption, and alternative procedures to implement when the assumption is violated.

First, we clarified the nature of the homogeneity of within-subgroup error variance assumption and distinguished it from the homoscedasticity assumption; these assumptions are not equivalent, and the homogeneity of within-subgroup error variance assumption can be violated even in the presence of subgroup and overall homoscedasticity. Second, we delineated the consequential effects of heterogeneity of within-subgroup error variance in terms of making incorrect conclusions regarding the presence or absence of moderating effects, in general, and differential prediction in human resources management research. In the most typical validation study in which the subgroup with the larger *n* (i.e., majority subgroup) is paired with the smaller validity coefficient, heterogeneity of within-subgroup error variance leads to very low statistical power. Consequently, organizations may (unknowingly) use selection and placement tests that predict performance differentially for various ethnic- or gender-based subgroups. Third, although researchers are eager to advance solutions that mitigate the detrimental impact of violating the homogeneity of within-subgroup error variance assumption on MMR, a recently proposed inverse data transformation procedure should not be used: Even though it alleviates heterogeneity of within-subgroup error variance, it may also eliminate the moderating effect altogether. Finally, the empirical Monte Carlo evidence accumulated thus far favors the use of alternative parametric methods such as the *A* statistic in lieu of MMR for testing moderator variable hypotheses when the homogeneity of within-subgroup error variance assumption is violated.

References

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., & Pierce, C. A. (in press-a). Statistical power computations for detecting dichotomous moderator variables with moderated multiple regression. *Educational and Psychological Measurement.*

Aguinis, H., & Pierce, C. A. (in press-b). Testing moderator variable hypotheses meta-analytically. *Journal of Management.*

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

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192-206.

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

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.

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika, 35*, 88-96.

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

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

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice* (University Paper Series on Quantitative Applications in the Social Sciences Rep. No. 07-050). Beverly Hills, CA: Sage.

Bobko, P., & Russell, C. J. (1990). Variance homogeneity in interactive regression: A clarifying note about data transformations. *Journal of Applied Psychology, 75*, 95-96.

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

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

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

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.

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

DeShon, R. P., & Alexander, R. A. (1994b, 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.

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.

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.

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

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.

Houston, W. M., & Novick, M. R. (1987). Race-based differential prediction in Air Force technical training programs. *Journal of Educational Measurement, 24*, 309-320.

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

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.

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

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

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics* (Vol. 2, 4th ed.). London: Charles Griffin.

Landy, F. J. (1992). Introduction: Case studies in test validation. *Journal of Business and Psychology, 7*, 111-117.

Linn, R. L. (1983). Predictive bias as an artifact of selection procedures. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord* (pp. 27-40). Hillsdale, NJ: Lawrence Erlbaum.

Linn, R. L., & Hastings, C. N. (1984). Subgroup differentiated prediction. *Applied Psychological Measurement, 8*, 165-172.

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

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

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

Olejnik, S. F., & Algina, J. (1987). An analysis of statistical power for parametric ANCOVA and rank transform ANCOVA. *Communications in Statistics: Theory and Methods, 16*, 1923-1949.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). Fort Worth, TX: Holt, Rinehart & Winston.

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

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

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

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

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

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

Valentine, L. D. (1977). *Prediction of Air Force technical training success from ASVAB and educational background* (AFHRL-TR-77-18). Lackland Air Force Base, TX: Air Force Human Resources Laboratory, Personnel Research Division.