

# Detecting Between-Groups Heteroscedasticity in Moderated Multiple Regression With a Continuous Predictor and a Categorical Moderator: A Monte Carlo Study

SAGE Open  
 January–March 2016: 1–14  
 © The Author(s) 2016  
 DOI: 10.1177/2158244015621115  
[sgo.sagepub.com](http://sgo.sagepub.com)  


Patrick J. Rosopa<sup>1</sup>, Amber N. Schroeder<sup>2</sup>, and Jessica L. Doll<sup>3</sup>

## Abstract

Moderated multiple regression (MMR) is frequently used to test moderation hypotheses in the behavioral and social sciences. In MMR with a categorical moderator, between-groups heteroscedasticity is not uncommon and can inflate Type I error rates or reduce statistical power. Compared with research on remedial procedures that can mitigate the effects of this violated assumption, less research attention has focused on statistical procedures that can be used to detect between-groups heteroscedasticity. In the current article, we briefly review such procedures. Then, using Monte Carlo methods, we compare the performance of various procedures that can be used to detect between-groups heteroscedasticity in MMR with a categorical moderator, including a heuristic method and a variant of a procedure suggested by O'Brien. Of the various procedures, the heuristic method had the greatest statistical power at the expense of inflated Type I error rates. Otherwise, assuming that the normality assumption has not been violated, Bartlett's test generally had the greatest statistical power when direct pairing occurs (i.e., when the group with the largest sample size has the largest error variance). In contrast, O'Brien's procedure tended to have the greatest power when there was indirect pairing (i.e., when the group with the largest sample size has the smallest error variance). We conclude with recommendations for researchers and practitioners in the behavioral and social sciences.

## Keywords

moderated multiple regression, heteroscedasticity, heterogeneity of variance, statistical assumptions

Testing for the equality of regression slopes is frequently conducted in the behavioral and social sciences. Evidence of this can be found in research on differential prediction (Aguinis, Culpepper, & Pierce, 2010; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; Saad & Sackett, 2002) and analysis of covariance (Fox, 2008; Huitema, 1980; Rutherford, 1992). Testing for the equality of regression slopes is equivalent to testing whether the relationship between a continuous outcome and a continuous predictor differs depending on a third variable—a moderator (Saunders, 1956; Stone-Romero & Liakhovitski, 2002).

The study of moderator variables, in general, is important for theory development and knowledge cumulation in education, management, industrial-organizational psychology, and related disciplines. Consistent with this, Hall and Rosenthal (1991) noted,

If we want to know how well we are doing in the biological, psychological, and social sciences, an index that will serve us

well is how far we have advanced in our understanding of the moderator variables of our field. (p. 447)

Although a variety of procedures exist for detecting the effects of continuous and categorical moderators (Stone-Romero & Liakhovitski, 2002; Zedeck, 1971), researchers have noted that moderated multiple regression (MMR) has become the major procedure for testing hypotheses involving categorical moderators (Aguinis, 2004; Overton, 2001; Sackett & Wilk, 1994; Shieh, 2009).

<sup>1</sup>Clemson University, SC, USA

<sup>2</sup>Western Kentucky University, Bowling Green, USA

<sup>3</sup>Coastal Carolina University, Conway, SC, USA

## Corresponding Author:

Patrick J. Rosopa, Associate Professor, Department of Psychology, College of Behavioral, Social, and Health Sciences, Clemson University, 418 Brackett Hall, Clemson, SC 29634-1355, USA.  
 Email: prosopa@clemson.edu



Creative Commons CC-BY: This article is distributed under the terms of the Creative Commons Attribution 3.0 License (<http://www.creativecommons.org/licenses/by/3.0/>) which permits any use, reproduction and distribution of

the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Regrettably, in MMR with a categorical moderator, it is not uncommon to violate the homoscedasticity assumption (see Aguinis & Pierce, 1998; DeShon & Alexander, 1996; Overton, 2001), which can lead to inflated Type I errors or reduced statistical power (DeShon & Alexander, 1996; Ng & Wilcox, 2010; Overton, 2001). More specifically, in MMR, the form of heteroscedasticity that can manifest is one in which the error variance differs across the levels of a categorical moderator (e.g., gender; for a review, see Aguinis, 2004; DeShon & Alexander, 1996; Ng & Wilcox, 2010; Rosopa, Schaffer, & Schroeder, 2013; Wilcox, 1997), or stated another way, between-groups heteroscedasticity exists (Ng & Wilcox, 2010).

Based on a review of three journals (*Academy of Management Journal*, *Journal of Applied Psychology*, and *Personnel Psychology*) from 1987 to 1999, Aguinis, Peterson, and Pierce (1999) identified 87 articles that reported at least one test for the equality of regression slopes. Out of 117 tests, Aguinis and his colleagues found that at least 39% of these violated the assumption. The implication of this finding is that researchers might have wrongly concluded that an interaction exists in the population when it does not (Type I error) or that an interaction does not exist in the population when it does (Type II error). In either case, "substantive research conclusions can be erroneous, theory development can be hindered, and incorrect decisions can be made . . ." (p. 319).

Although there exist a number of remedial procedures (Rosopa et al., 2013) that can be used to mitigate the effects of between-groups heteroscedasticity in MMR, including the use of statistical approximations (Alexander & Govern, 1994; DeShon & Alexander, 1994; Shieh, 2009), robust methods (Cribari-Neto, 2004; Long & Ervin, 2000; Wilcox, 2005), and weighted least squares regression (Overton, 2001; Rosopa, 2006), less research attention has focused on statistical procedures that can be used to detect between-groups heteroscedasticity in MMR. Currently, there is no empirical research that systematically compares the various approaches that can be used to detect between-groups heteroscedasticity. Thus, consistent with recommendations by Rosopa et al. (2013), one major purpose of the present article is to compare the performance of various procedures that can be used to detect between-groups heteroscedasticity in MMR with a categorical moderator.

Although researchers across diverse disciplines (e.g., econometrics, psychology, and statistics) have suggested different approaches for detecting heteroscedasticity in general (Rosopa et al., 2013), some procedures are sensitive to non-normality. A robust approach by O'Brien (1979, 1981), however, has been recommended for use in ANOVA. Thus, another purpose of the present article is to suggest a variation of O'Brien's procedure that can be used for instances in which a researcher is interested in testing for the equality of regression slopes.

Our article is divided into four major sections. First, we formally define the model used in MMR with a categorical

moderator. Second, we describe between-groups heteroscedasticity and its biasing effects. Third, we review various procedures that can be used to detect between-groups heteroscedasticity, including O'Brien's (1979, 1981) procedure. Fourth, we describe the results of a Monte Carlo simulation designed to assess the relative performance of various procedures that can be used to detect between-groups heteroscedasticity.

## **MMR With a Continuous Predictor and a Categorical Moderator**

When testing for the equality of  $k$  regression slopes using MMR, a continuous outcome ( $y$ ) is modeled as a function of a continuous predictor ( $x$ ), a categorical moderator ( $z$ ) (indexed by  $k - 1$  regressors, that is,  $z_1, z_2, \dots, z_{k-1}$ ), and the two-way interaction between  $x$  and  $z$  (indexed by  $k - 1$  product terms between  $x$  and the regressors). Population parameters are denoted by Greek letters such as  $\beta$  and  $\sigma^2$ , for example, to differentiate them from sample estimates using a circumflex, that is,  $\hat{\beta}$  and  $\hat{\sigma}^2$ , respectively.

When  $k = 2$ , the full linear model for the  $i$ th observed response in the  $j$ th group can be expressed as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_{1ij} + \beta_3 x_{ij} z_{1ij} + e_{ij}, \quad (1)$$

for  $i = 1, 2, \dots, n_j$ ,  $j = 1, 2, \dots, k$ , where  $n_j$  is the  $j$ th sample size;  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  are unstandardized regression coefficients; and  $e_{ij}$  is the  $i$ th residual in the  $j$ th group (an estimate of  $\varepsilon_{ij}$ , an unknown population error). Although Equation 1 is a fixed effects regression model, when the predictors are treated as random variables, the estimates of  $E(y_{ij})$  can be viewed as conditional on the specific values of the predictors, that is,  $E(y_{ij} | x_{ij}, z_{1ij})$  (Bauer & Curran, 2005; Rencher, 2000).

More generally, for  $k \geq 2$ , the full linear model for the  $N = \sum_{j=1}^k n_j$  observations (with the number of terms  $p = 2k - 1$ ) can be compactly expressed in matrix form. See the appendix for the general form, including the assumptions that specify that the model is linear, all relevant terms are included in the model, and  $\varepsilon_{ij}$ s are constant and uncorrelated.

Note that normally distributed  $\varepsilon_{ij}$ s are neither required nor assumed for the linear model to be valid (Rencher, 2000). However, when the normality assumption is invoked for statistical inferences (e.g., confidence intervals, hypothesis tests), this implies that the  $y_{ij}$ s (and  $\varepsilon_{ij}$ s) are independent. When  $k = 2$ , the test for the equality of independent slopes is distributed as  $t$  with  $df = N - 4$ . When  $k > 2$ , the test statistic is distributed as an  $F$  random variable (see the appendix).

In Equation 1, the  $N$  residuals ( $e_{ij}$ ) are assumed to have a diagonal  $N \times N$  covariance matrix given by

$$\text{cov}(\mathbf{e}) = \sigma_e^2 \mathbf{I}_N = \begin{pmatrix} \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_e^2 \end{pmatrix}, \quad (2)$$

(Fox, 2008; Rencher, 2000). Note the common variance on the main diagonal in Equation 2. Heteroscedasticity, in contrast, is said to exist when these variances are no longer equal. This can be denoted by

$$\text{cov}(\mathbf{e}) = \sigma_{e_m}^2 \mathbf{I}_N = \begin{pmatrix} \sigma_{e_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{e_2}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{e_N}^2 \end{pmatrix}, \quad (3)$$

where  $\sigma_{e_m}^2 \neq \sigma_{e_{m'}}^2$  for some  $m$  and  $m'$  such that  $m = 1, 2, \dots, N$ . As an example of heteroscedasticity, assume that  $k = 2$  where the first  $n_1$  observations that are from Group 1 have a variance 2 times larger (i.e.,  $2\sigma^2$ ) than the remaining  $n_2$  observations that are from Group 2; heteroscedasticity exists such that the  $N$  diagonal elements of Equation 3 would assume two different values— $2\sigma^2$  or  $\sigma^2$ . This is an example of between-groups heteroscedasticity (Ng & Wilcox, 2010).<sup>1</sup>

### Between-Groups Heteroscedasticity and Its Biasing Effects

Extant research has found that between-groups heteroscedasticity can affect statistical inferences (e.g., increased Type I or Type II error rates) and these effects are nontrivial (DeShon & Alexander, 1996; Dretzke, Levin, & Serlin, 1982; Ng & Wilcox, 2010; Overton, 2001).

The error variance in the  $j$ th group ( ${}_j\sigma_e^2$ ) can be expressed as

$${}_j\sigma_e^2 = {}_j\sigma_y^2 \left(1 - {}_j\rho_{yx}^2\right), \quad (4)$$

where  ${}_j\sigma_y^2$  and  ${}_j\rho_{yx}$ , respectively, are the variance of  $y$  in the  $j$ th group and the correlation coefficient between  $y$  and  $x$  in the  $j$ th group (DeShon & Alexander, 1996; Dretzke et al., 1982). When between-groups homoscedasticity exists,  ${}_1\sigma_e^2 = \dots = {}_k\sigma_e^2$ .

Inspection of Equation 4 shows that if  ${}_j\sigma_y^2$  (or  ${}_j\rho_{yx}$ ) is constant across the  $k$  groups, then any difference in  ${}_j\rho_{yx}$  (or  ${}_j\sigma_y^2$ ) across the  $k$  groups will result in between-groups heteroscedasticity, unless values for  ${}_j\sigma_y^2$  and  ${}_j\rho_{yx}$  are such that they “balance out” so as to satisfy the homoscedasticity assumption. Moreover, when population regression slopes actually differ, the assumption is likely to be violated (Overton, 2001). This is evident in the following expression, after substituting  ${}_j\rho_{yx} = {}_j\beta_{y,x}({}_j\sigma_x / {}_j\sigma_y)$  into Equation 4:

$${}_j\sigma_e^2 = {}_j\sigma_y^2 - \left({}_j\beta_{y,x}\right)^2 {}_j\sigma_x^2, \quad (5)$$

where  ${}_j\beta_{y,x}$  and  ${}_j\sigma_x^2$ , respectively, are the slope based on the regression of  $y$  on  $x$  in the  $j$ th group and the variance of  $x$  in the  $j$ th group. Thus, when slopes are unequal,  ${}_j\sigma_y^2$  and  ${}_j\sigma_x^2$  in each group must have values that offset one another so as to allow  ${}_1\sigma_e^2 = \dots = {}_k\sigma_e^2$ . As research suggests, when testing for the equality of regression slopes, violating the between-groups homoscedasticity assumption is not uncommon (Aguinis & Pierce, 1998; DeShon & Alexander, 1996; Overton, 2001; Wilcox, 1997).

This violated assumption has biasing effects on the Type I error rates and the statistical power of MMR whether subgroup sample size ( $n_j$ ) is equal or unequal across the categorical moderator. Although with equal  $n_j$ s and equal  ${}_j\sigma_x^2$  across groups, some argue that Type I error rates perform “acceptably well” (Dretzke et al., 1982, p. 376), when equal  ${}_j\sigma_x^2$  across groups is untenable, Type I error rates become conservative, which can reduce the power of MMR (DeShon & Alexander, 1996). However, the power of MMR does not suffer greatly when  $n_j$ s are equal and error variances do not differ considerably (Alexander & DeShon, 1994).

With between-groups heteroscedasticity and unequal  $n_j$ s, however, the effects are much more severe. Type I error rates and statistical power “can be either gross underestimates or severe overestimates depending on the pattern of sample sizes relative to the pattern of error variances” (DeShon & Alexander, 1996, p. 270). More precisely, when the larger  ${}_j\sigma_e^2$  is paired with the larger  $n_j$  (direct pairing), statistical tests based on MMR become conservative. This results in actual Type I error rates less than the nominal level and, ceteris paribus, power is decreased. Conversely, when the larger  ${}_j\sigma_e^2$  is paired with the smaller  $n_j$  (indirect pairing), statistical tests based on MMR become liberal. This results in actual Type I error rates greater than the nominal level and, ceteris paribus, power is increased (albeit illegitimately; see, for example, DeShon & Alexander, 1996; Overton, 2001).<sup>2</sup>

To exacerbate matters, unequal  $n_j$ s are quite common in the behavioral and social sciences for a number of reasons. One reason is that attrition may result in unbalanced data (Shadish, Cook, & Campbell, 2002), such as in randomized experiments where participants in some conditions fail to complete an outcome measure. Another reason is that the population from which a researcher purposively samples could be disproportionate across subpopulations of the characteristic of interest (e.g., race; Shadish et al., 2002). This commonly occurs in the validation of personnel selection instruments (see, for example, Hattrup & Schmitt, 1990; Hunter, Schmidt, & Hunter, 1979). In addition, in longitudinal studies or in the analysis of archival data, missing values can lead to unequal  $n_j$ s across variables of interest (Schafer & Graham, 2002).

Overall, the biasing effects of between-groups heteroscedasticity on Type I error rates and statistical power can have implications on both theory development and practice in the behavioral and social sciences (Aguinis & Pierce, 1998; Oswald, Saad, & Sackett, 2000; Rosopa et al., 2013). For example, assume that sample sizes are unequal between two independent groups (e.g., male vs. female) and between-groups heteroscedasticity exists such that the larger error variance is paired with the group with the larger sample size (i.e., direct pairing). Furthermore, assume that the researcher/practitioner failed to detect a hypothesized slope difference between groups (i.e., between males and females) that actually exists in the population. Stated differently, the failure to detect a hypothesized moderating effect that exists in the population might be due to the influence of between-groups heteroscedasticity. As detailed in a review by Aguinis and Pierce (1998), inflated Type I error rates could lead to the publication of specious results. This seems plausible considering that, for decades, researchers have noted the problem of failing to detect hypothesized moderators using MMR (Aguinis & Stone-Romero, 1997; McClelland & Judd, 1993; Zedeck, 1971).

As noted above, researchers have identified a number of alternatives to MMR when between-groups heteroscedasticity exists. For example, DeShon and Alexander (1996) conducted a comprehensive Monte Carlo study evaluating the relative performance of various statistical approximations, with two statistical approximations (viz.,  $A$  and  $J$  approximations) having the most promise. With a dichotomous moderator, Overton (2001) suggested a weighted least squares approach for MMR. In addition, some researchers have recommended certain robust estimators regardless of the form of heteroscedasticity (Cribari-Neto, 2004; Long & Ervin, 2000).

Because violation of the between-groups homoscedasticity assumption can afflict the Type I error rates and power of MMR, it would be useful to assess whether this assumption has been violated. The following section considers this issue.

## A Review of Procedures for Detecting Between-Groups Heteroscedasticity

An issue seldom raised by researchers or practitioners in the context of MMR is how to detect violations of the between-groups homoscedasticity assumption. Although Aguinis (2004) explained that there are two methods (to be noted below) for evaluating whether the assumption has been violated, any procedure that can be used to test the equality of  $k$  independent variances could potentially be used to detect between-groups heteroscedasticity in MMR with a categorical moderator. Some procedures involve the variances of the residuals, whereas others may use some other measure of dispersion (Boos & Brownie, 2004; Conover, Johnson, & Johnson, 1981). Some procedures are used specifically in the

context of ANOVA (e.g., Brown & Forsythe, 1974). Another procedure is used primarily in regression models in econometrics (e.g., Breusch & Pagan, 1979). In addition, as noted below, a rule-of-thumb has also been recommended in MMR with a categorical moderator (see DeShon & Alexander, 1996). However, the relative performance of these and other procedures described below has not been examined.

In addition, although a number of studies have compared various tests for homogeneity of variances specifically in ANOVA (see, for example, Conover et al., 1981; Martin & Games, 1977), we could not find any studies involving MMR with between-groups heteroscedasticity and the effects of direct and indirect pairing. For example, a simulation conducted by Conover et al. (1981) involved a one-way ANOVA with four independent groups, and they included only direct pairing conditions when  $N = 80$ . Thus, because neither sample size nor type of pairing was manipulated, the effect of these factors could not be examined. Games, Winkler, and Probert (1972) empirically investigated the robustness of various tests for homogeneity of variances to violations of the normality assumption in the context of ANOVA with three independent groups. However, because sample sizes were always equal across groups, pairing of error variances was not considered. Boos and Brownie (2004) reviewed the two-sample case and a one-way ANOVA, but did not report simulation results. In a simulation by Sarkar, Kim, and Basu (1999), they included a one-way ANOVA with three independent groups and considered both direct and indirect pairing for  $Ns$  as large as 120.

As noted above, Aguinis (2004) mentioned two methods for detecting between-groups heteroscedasticity. One was a heuristic method suggested by DeShon and Alexander (1996). The second was a statistical test by Bartlett (1937). In the sections that follow, we describe these and other procedures that could be used to detect between-groups heteroscedasticity.

### Heuristic Method

DeShon and Alexander (1996) described a heuristic method to signal whether the between-groups heteroscedasticity assumption has been violated to such a degree as to unduly influence the results of MMR analyses. Specifically, when a researcher calculates the variance of the residuals separately within each of the  $k$  groups, the ratio of the largest estimated variance to the smallest estimated variance should not exceed 1.5. This ratio is computationally simple and does not require specialized software.

Note that the heuristic method is not a statistical test, but rather a rule-of-thumb and its statistical performance, in terms of Type I error or power, has not been examined. As a rule-of-thumb, the heuristic method may not possess the desirable property of being robust at any Type I error rate ( $\alpha$ ). That is, regardless of  $\alpha$  (e.g., .01 or .05), a researcher would conclude that heteroscedasticity exists if the ratio (based on

sample estimates of two variances) exceeds 1.5. However, the heuristic method was included in the present simulation to assess its performance relative to other procedures.

### Bartlett

Bartlett (1937) developed a procedure that can be used to test for homogeneity of variances by conducting a transformation of the variances. To use this procedure, this test involves transforming the variances of the residuals across the levels of  $z$ . To test the null hypothesis that  $\sigma_e^2 = \dots = \sigma_e^2$ , we calculate

$$s^2 = \frac{\sum_{j=1}^k v_j s_j^2}{\sum_{j=1}^k v_j},$$

$$c = 1 + \frac{1}{3(k-1)} \left[ \sum_{j=1}^k \frac{1}{v_j} - \frac{1}{\sum_{j=1}^k v_j} \right],$$

and

$$u = \left( \sum_{j=1}^k v_j \right) \ln s^2 - \sum_{j=1}^k v_j \ln s_j^2,$$

where  $s_j^2$  is an estimate of the variance of the  $e_{ij}$ s in the  $j$ th group with degrees of freedom equal to  $v_j = n_j - 1$ . The test statistic,  $u/c$ , is approximately distributed as chi-square with degrees of freedom equal to  $(k-1)$ . The null hypothesis is rejected if  $u/c > \chi_{\alpha, k-1}^2$ . As suggested by DeShon and Alexander (1996), this procedure can be used to test whether between-groups heteroscedasticity exists. Although Games et al. (1972) found that this test had greater power than a number of other tests, it has been noted that this procedure is sensitive to departures from normality (Box, 1953; Levene, 1960).

### Brown and Forsythe

To detect heteroscedasticity in the context of ANOVA, Brown and Forsythe (1974) suggested conducting a one-way ANOVA on the absolute value of the residuals around the group median instead of the mean (cf. Levene, 1960). Based on simulations conducted by Conover et al. (1981), tests for homogeneity of variances based on the median tend to control Type I error rates better than tests based on the mean. Brown and Forsythe's procedure is relatively straightforward and appears to be less affected by skewed data in unbalanced designs than other procedures, while still providing adequate statistical power (Lix, 1996). In addition, because of its computational ease, it may be a very practical procedure for

researchers and practitioners (Boos & Brownie, 2004; Conover et al., 1981).

### Score

The score test, developed independently in the econometrics (Breusch & Pagan, 1979) and statistics (Cook & Weisberg, 1983) literature, can be used to detect various forms of heteroscedasticity. For example, the score test can be used to test whether error variances differ as a function of continuous predictors, categorical predictors, or predicted values. This procedure requires two regression analyses. In the first analysis, the sum of squares error ( $SSE$ ) from the regression equation of interest is required (see the numerator of Equation A4 in the appendix). Then, in a second regression analysis, the squared residuals from the first analysis are regressed on the variables believed to be the cause of the heteroscedasticity (e.g., the categorical moderator), and the sum of squares regression ( $SSR$ ) is calculated. The test statistic for the score test,  $(SSR/2) / (SSE/N)^2$ , is asymptotically distributed as chi-square with degrees of freedom equal to the number of variables used to predict the squared residuals.

Although the score test is not frequently used in the behavioral sciences, this procedure was included in the present study because of its flexibility. In addition, because the components needed for the statistical test are based on two regression equations (i.e., customized syntax or a stand-alone program is not required), this procedure would be generally accessible for a wide variety of users.

### O'Brien

Analogous to testing for the main and interactive effects in ANOVA, O'Brien (1979, 1981) developed a procedure that could be used to test for the main and interactive effects of the variances in the cells of one-way and factorial designs. This robust procedure has been recommended even when the normality assumption is violated (Maxwell & Delaney, 2000; O'Brien, 1979, 1981). An especially lucid description of the procedure can be found in Maxwell and Delaney (2000).

Because O'Brien's (1979, 1981) procedure is limited to designs that have only categorical predictors (e.g., one-way and factorial ANOVAs), it would be useful to generalize this method to designs that include categorical and continuous predictors. Below, we describe how O'Brien's (1979, 1981) method can be used where hypotheses involving the equality of regression slopes are being tested. Here, we focus on a dichotomous moderator.<sup>3</sup>

The modified procedure requires three steps. The first step is to calculate the residuals ( $e_{ij}$ ) from the full model (see Equation 1). Then, for each group, we calculate

$$r_j^2 = \frac{\sum_{i=1}^{n_j} (e_{ij})^2}{n_j - 2}. \quad (6)$$

The second step involves a transformation of each of the individual residuals. This calculation is achieved using the following equation:

$$e'_{ij} = \frac{n_j(n_j - 1.5)(e_{ij})^2 - .5r_j^2(n_j - 2)}{2(n_j - 2)}. \quad (7)$$

To check the calculations, the average of the  $e'_{ij}$ s in Equation 7, within each group, should equal the corresponding value in Equation 6. Specifically,  $\bar{e}'_{i1} = r_1^2$  and  $\bar{e}'_{i2} = r_2^2$ .

The third step is to conduct a two-independent-samples  $t$  test on the transformed residuals from Equation 7, using the categorical moderator ( $z$ ) (e.g., female vs. male, treatment group vs. control group) as the grouping variable. If the results of this test are statistically significant at some predetermined  $\alpha$ , then there is evidence to suggest that between-groups heteroscedasticity exists.<sup>4</sup>

In the following sections, we describe the design and results of a Monte Carlo study used to compare the performance (viz., Type I error and statistical power) of the five procedures described above—heuristic method, Bartlett's (1937) test, Brown and Forsythe's (1974) test, score test, and modified O'Brien's (1979, 1981) test.

## Method

We used Monte Carlo methods (Robert & Casella, 2004) to evaluate the performance of five procedures that can be used to detect between-groups heteroscedasticity in MMR with a dichotomous moderator. Note that the nominal  $\alpha$  for all tests was .05. The manipulated parameters of our  $5 \times 3 \times 8 \times 2 \times 5$  research design resulted in 1,200 conditions. Each of the manipulated parameters is described next.

### Manipulated Parameters

**Total sample size.** Five levels of  $N$  were used in the present study. These levels were 60, 120, 180, 240, and 300. The  $N$ s for the present study overlap with those used in previous research on MMR (e.g., Aguinis & Stone-Romero, 1997; DeShon & Alexander, 1996) and bracket the  $N$ s typically encountered in validation studies (Salgado, 1998).

**Subgroup sample size.** Sample size within groups was systematically manipulated using the following three ratios ( $n_1:n_2$ ): (a) 1:1, (b) 1:2, and (c) 1:3. For example, when  $N = 120$ , the subgroup sample sizes, based on the three ratios, were (a)  $n_1 = n_2 = 60$ , (b)  $n_1 = 40$  and  $n_2 = 80$ , and (c)  $n_1 = 30$  and  $n_2 = 90$ .

**Between-groups heteroscedasticity.** Between-groups heteroscedasticity assumed eight levels, which involved the ratios of the population error variance in each group ( ${}_1\sigma_e^2 : {}_2\sigma_e^2$ ). Specifically, the ratios of the population error variances were (a) 1:1, (b) 1:1.25, (c) 1:1.5, (d) 1:1.75, (e) 1:2, (f) 1:2.5, (g) 1:3, and (h) 1:4. Note that the ratio 1:1 represents

homoscedasticity because the population error variances are the same between groups, and the remaining ratios represent increasing levels of between-groups heteroscedasticity. These levels bracket the heuristic approach suggested by DeShon and Alexander (1996), and we did not manipulate the ratio of the error variances beyond these eight levels (e.g., 1:6) because such ratios would result in very high rejection rates, making it difficult to distinguish differences in performance among the procedures.

**Type of pairing.** Depending on whether the larger error variance ( ${}_j\sigma_e^2$ ) is paired with a small versus a large  $n_j$ , the ability of MMR to detect a hypothesized moderator can result in inflated (or conservative) Type I error rates or reduced statistical power (DeShon & Alexander, 1996; Ng & Wilcox, 2010; Overton, 2001; Shieh, 2009). Although our focus is not on the power of MMR analyses but on the performance of the five procedures to detect between-groups heteroscedasticity, we felt that it would be useful to assess whether the performance of these five procedures differed depending on the type of pairing. We considered direct pairing (i.e., largest  ${}_j\sigma_e^2$  paired with the largest  $n_j$ ) and indirect pairing (i.e., largest  ${}_j\sigma_e^2$  paired with the smallest  $n_j$ ; see DeShon & Alexander, 1996; Overton, 2001).

**Effect size.** In the present study, the MMR effect size ( $f^2$ ) was also manipulated. Although varying the size of the moderating effect was not the focus of our study, we felt that it was useful to determine whether the moderator effect size influenced the performance of the various procedures to detect between-groups heteroscedasticity. We used the modified effect size by Aguinis, Beatty, Boik, and Pierce (2005). Based on their 30-year review of research involving MMR with a categorical moderator in applied psychology and allied fields, the median effect size was .002.

Thus, in the present study, the levels of the manipulated effect size were .001, .002, .005, .01, and .02. These levels included the median effect size reported by Aguinis et al. (2005). Although Cohen (1988) labeled  $f^2 = .02$  as a small effect, in the review by Aguinis et al., they found that this was the effect size at which studies in applied psychology and management had an average power level of .84 to detect such an effect. Beyond this effect size, the power of the usual MMR test for equality of regression slopes would exceed typical recommended levels for power.

### Data Generation

For each condition, data generation and statistical analyses were conducted in *R*—a free, open source, statistical software package (Culpepper & Aguinis, 2011; R Development Core Team, 2011). For the  $j$ th group,  $n_j$  observations of bivariate normal data with population means of 0 were generated using the *mvtnorm* function in the *MASS* library in *R*. The population variances for  $x$  ( ${}_j\sigma_x^2$ ) were set to 1.  ${}_j\sigma_e^2$  assumed

**Table 1.** Empirical Type I Error Rates as a Function of Sample Size and Subgroup Proportions.

N	Heuristic	Bartlett	Brown-Forsythe	Score	O'Brien
			$n_1:n_2 = 1:1$		
60	.2864	.0496	.0448	.0474	.0462
120	.1184	.0468	.0448	.0458	.0448
180	.0550	.0484	.0474	.0482	.0504
240	.0302	.0498	.0514	.0492	.0502
			$n_1:n_2 = 1:2$		
60	.3362	.0532	.0466	.0490	.0494
120	.1518	.0516	.0524	.0508	.0514
180	.0708	.0480	.0448	.0446	.0490
240	.0400	.0524	.0510	.0518	.0522
			$n_1:n_2 = 1:3$		
60	.3714	.0598	.0486	.0466	.0510
120	.1956	.0590	.0536	.0510	.0538
180	.1036	.0492	.0498	.0460	.0460
240	.0560	.0474	.0446	.0460	.0460

Note. N = total sample size;  $n_1$  = sample size in Group 1;  $n_2$  = sample size in Group 2;  $f^2 = .002$ .

values as described above.  ${}_1\beta_{y,x} = 0.5$  whereas  ${}_2\beta_{y,x}$  was allowed to differ so as to equal one of the specified values for  $f^2$ . Equation 5 was used to solve for  ${}_j\sigma_x^2$ . Then, Equation 4 was used to solve for  ${}_j\rho_{yx}$ . Note that to compute  ${}_1\beta_{y,x}$ , the Solver function in MS Excel 2010 was used. This function can minimize or maximize a formula by changing user-specified cells. Alternatively, it can be used to set a formula to a specified value (e.g.,  $f^2 = .002$ ) by changing user-specified cells. More precisely, given the just-noted parameters and equalities, Solver was used to find the value for  ${}_2\beta_{y,x}$  (referred to as the cell to be changed in MS Excel) that would result in a specific  $f^2$  (referred to as the target cell in MS Excel). The target cell contained the formula by Aguinis et al. (2005). Note that all default options were used in the function. However, the precision option was set to  $1 \times 10^{-17}$ . Although the Solver function could not find a solution for  ${}_2\beta_{y,x}$  that results in an  $f^2$  exactly equal to the manipulated value, the difference was minuscule and, therefore, was retained. For example, for one condition with parameters as specified above, where  $f^2$  should equal .002, the Solver function found the value for  ${}_2\beta_{y,x}$  that results in an  $f^2 = .00199995713664686$ .<sup>5</sup>

We also conducted a series of accuracy checks to ensure that the data we generated conformed to the various parameters that we manipulated. In addition, we checked our data generation algorithm against similar conditions considered by DeShon and Alexander (1996), Dretzke et al. (1982), and Overton (2001).

On each simulated data set, the five procedures (viz., heuristic method; Bartlett's, 1937, test; Brown and Forsythe's, 1974, test; the score test; and the modified O'Brien's, 1979, 1981, test) were used to test whether between-groups heteroscedasticity existed. For each condition, there were 5,000 replications. The proportion of times that the null hypothesis

was rejected within a condition was recorded for each procedure.

## Results

The performance of the five procedures are compared below in terms of Type I error rate and power. Due to space limitations, we do not present the results of all 1,200 conditions. Because the pattern of results were the same regardless of the size of the moderating effect, we present results when  $f^2 = .002$ , the median effect size based on the 30-year review by Aguinis et al. (2005). Note that the complete set of results and R code can be obtained from the first author.

### Type I Error

For the conditions in which homoscedasticity existed (i.e.,  ${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1$ ), the average Type I error rate for Bartlett's (1937) test, Brown and Forsythe's (1974) test, the score test, and the modified O'Brien's (1979, 1981) test were similar to the nominal alpha (i.e., .05; see Table 1). The procedure that appeared to perform poorly was the heuristic method. Although the other four procedures appear to control Type I error at the nominal level, the heuristic method was not robust (Serlin, 2000), with empirical rejection rates typically much greater than .05. For example, when  $N = 60$  and subgroup sample sizes were equal, the empirical Type I error rate for the heuristic method was .2864 whereas the other four procedures had empirical Type I error rates near .05.

Although the heuristic method is not a formal statistical test, but simply a rule-of-thumb, and given that sampling error will affect the estimate of the residual variance in Group 1 and the estimate of the residual variance in Group 2, when  $N = 60$  and subgroup sample sizes are equal, it appears that

**Table 2.** Statistical Power as a Function of Sample Size and Degree of Between-Groups Heteroscedasticity When Subgroup Sample Sizes Are Equal.

N	Heuristic	Bartlett	Brown-Forsythe	Score	O'Brien
			${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$		
60	.3682	.0906	.0718	.0874	.0788
120	.2480	.1372	.1204	.1358	.1268
180	.1892	.1744	.1522	.1726	.1666
240	.1612	.2260	.2062	.2250	.2258
			${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.5$		
60	.5326	.1992	.1572	.1960	.1740
120	.4928	.3322	.2784	.3292	.3100
180	.4890	.4620	.3996	.4596	.4446
240	.4856	.5864	.5166	.5840	.5768
			${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:2$		
60	.7810	.4464	.3616	.4390	.3896
120	.8630	.7446	.6536	.7410	.7144
180	.9128	.9044	.8424	.9032	.8860
240	.9390	.9656	.9322	.9654	.9626
			${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:3$		
60	.9636	.8236	.7310	.8198	.7532
120	.9942	.9822	.9628	.9820	.9774
180	.9996	.9996	.9970	.9996	.9988
240	1.0000	1.0000	.9998	1.0000	1.0000

Note. N = total sample size;  ${}_1\sigma_e^2$  = population error variance in Group 1;  ${}_2\sigma_e^2$  = population error variance in Group 2;  $f^2 = .002$ .

due to chance alone, 28.64% of the time the heuristic method would signal that between-groups heteroscedasticity exists when it does not. Perhaps not surprisingly, this inflated Type I error rate becomes increasingly worse as the sample size of the subgroups becomes more disproportionate. That is, the effect of sampling error on the estimates of the subgroup residual variance is exacerbated. For example, when  $N = 60$  and  $n_1:n_2 = 1:2$ , the empirical Type I error rate for the heuristic method was .3362 and when  $n_1:n_2 = 1:3$ , the empirical Type I error rate was further inflated to .3714. Notably, the empirical Type I error rates for the other four procedures remained near .05.

Another interesting result regarding the heuristic method is that as  $N$  increases its empirical Type I error rate decreases. For example, when  $N = 120$  and subgroup sample sizes were equal, the empirical Type I error rate for the heuristic method was .1184 and when  $N = 240$ , the empirical Type I error rate decreased to .0302. This is due to the fact that as  $N$  increases, the sampling error associated with estimating the residual variance in each group also decreases. Thus, as  $N$  increases, the two estimated variances are much more precise estimates of the population ratio ( ${}_1\sigma_e^2 : {}_2\sigma_e^2$ ) equal to 1. Again, for the other four procedures, the empirical Type I error rates remained near .05 regardless of  $N$  and subgroup sample sizes.

### Statistical Power

In this section, empirical rejection rates when the homoscedasticity assumption is violated (i.e., heteroscedasticity exists) are presented (see Tables 2-4). Table 2 presents results when subgroup sample sizes are equal (i.e.,  $n_1:n_2 = 1:1$ ).

Table 3 and Table 4 present results when subgroup sample sizes are unequal,  $n_1:n_2 = 1:2$  and  $n_1:n_2 = 1:3$ , respectively. There were notable differences in the performance of the five procedures in terms of power, which we describe across equal subgroup sample sizes, and both the direct and indirect pairing conditions.

**Equal subgroup sample sizes.** Recall that when subgroup sample sizes are equal, direct versus indirect pairing does not apply because the pairing of the larger (or the smaller) error variance with the group with the larger (or smaller) sample size is a non-issue because sample sizes are the same. In Table 2, with the exception of the heuristic method, four procedures had power that increased monotonically as  $N$  increased and as the degree of between-groups heteroscedasticity increased. Although Bartlett's (1937) test tended to be the most powerful and Brown and Forsythe's (1974) test tended to be the least powerful, it appears that when subgroup sample sizes are equal, there is relatively little difference in the power of these four procedures.

For the heuristic method, although it had the greatest power of all five procedures, recall that it had very inflated Type I error rates (see Table 1). Thus, the increased power comes at the cost of inflated Type I error rates. The heuristic method had power that increased monotonically as the degree of between-groups heteroscedasticity increased. For example, in Table 2, assuming  $N = 120$ , the heuristic method had power equal to .2480 when between-groups heteroscedasticity ( ${}_1\sigma_e^2 : {}_2\sigma_e^2$ ) was 1:1.25, and increased to .4928 when the degree of between-groups heteroscedasticity

**Table 3.** Statistical Power as a Function of Sample Size, Degree of Between-Groups Heteroscedasticity, and Type of Pairing When Subgroup Proportions ( $n_1:n_2$ ) Are 1:2.

N	Direct pairing					Indirect pairing				
	Heuristic	Bartlett	Brown-Forsythe	Score	O'Brien	Heuristic	Bartlett	Brown-Forsythe	Score	O'Brien
60	.4242	.1002	.0780	.0772	.0562	.3650	.0722	.0572	.0776	.0928
120	.3024	.1336	.1176	.1148	.0968	.2448	.1222	.1020	.1328	.1444
180	.2366	.1710	.1560	.1518	.1360	.1886	.1552	.1388	.1672	.1846
240	.1928	.2118	.1868	.1940	.1752	.1582	.2006	.1764	.2142	.2260
	${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.5$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:2$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:3$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.5$	
60	.5562	.1812	.1404	.1378	.0980	.5078	.1706	.1354	.1910	.2066
120	.5354	.3072	.2616	.2736	.2272	.4700	.2940	.2458	.3146	.3346
180	.5148	.4312	.3694	.4010	.3578	.4644	.4096	.3522	.4314	.4434
240	.5124	.5444	.4922	.5134	.4722	.4590	.5284	.4628	.5460	.5610
	${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:2$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:3$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.5$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:2$	
60	.8100	.4298	.3390	.3570	.2614	.7392	.3934	.3116	.4234	.4386
120	.8636	.7002	.6184	.6624	.5858	.8268	.6982	.6022	.7194	.7236
180	.9196	.8774	.8166	.8562	.8200	.8786	.8526	.7996	.8628	.8678
240	.9386	.9474	.9046	.9388	.9232	.9202	.9398	.9008	.9436	.9488
	${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:3$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.5$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:2$		${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$	
60	.9716	.7808	.6708	.7148	.5672	.9306	.7554	.6512	.7772	.7688
120	.9940	.9758	.9470	.9712	.9422	.9882	.9686	.9392	.9716	.9730
180	.9998	.9992	.9962	.9990	.9968	.9978	.9974	.9910	.9976	.9976
240	1.0000	1.0000	.9998	1.0000	1.0000	.9994	.9994	.9986	.9994	.9994

Note. N = total sample size;  $n_1$  = sample size in Group 1;  $n_2$  = sample size in Group 2;  ${}_1\sigma_e^2$  = population error variance in Group 1;  ${}_2\sigma_e^2$  = population error variance in Group 2;  $f^2 = .002$ .

increased to 1:1.5, and increased to .8630 when the degree of between-groups heteroscedasticity increased to 1:2.

Interestingly, when the degree of between-groups heteroscedasticity is fixed (e.g., 1:1.25), the power of the heuristic method did not increase monotonically as  $N$  increased. Recall that when subgroup sample sizes were equal,  $N = 60$ , and homoscedasticity was satisfied, the empirical Type I error rate was .2864 (see Table 1). Then, for a fixed  $N$  (e.g., 60), as the degree of between-groups heteroscedasticity increased, power increased (see Table 2). Because at larger  $N$ s, the empirical Type I error rate of the heuristic method was always decreasing (cf. Table 1), power then increases as heteroscedasticity increases albeit at a much lower starting point due to the lower Type I error rate.

**Direct pairing.** In Table 3, when there was direct pairing, the heuristic method was generally the most powerful of the five procedures when  $N \leq 180$ . Otherwise, the most powerful procedure was Bartlett's (1937) test followed by the score test, Brown and Forsythe's (1974) test, and the modified O'Brien (1979, 1981). For these four procedures, power increased monotonically as  $N$  increased and as the degree of between-groups heteroscedasticity increased. Consistent with Table 2, at milder levels of direct pairing ( ${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$ ), the power of the heuristic method decreased as  $N$  increased.

Recall that as  $N$  increased, the empirical Type I error rate for the heuristic method decreased. Thus, at smaller  $N$ s, the heuristic method had more of a power advantage to start because of its inflated Type I error rate. This relative power advantage at smaller  $N$ s tended to decrease as  $N$  increased because of the lower and lower Type I error rates. This is unlike the other four procedures, which are statistical tests that have their minimum value (i.e., lower asymptote) at alpha regardless of  $N$  or subgroup sample sizes.

In Table 4, with direct pairing, the trends were similar to Table 3. However, compared with Table 3, because of the increasingly disproportionate subgroup sample sizes in Table 4, power generally decreased. However, the rank order of the various procedures remained the same. Excluding the heuristic method, which was the most powerful due to its inflated Type I error rate, Bartlett's (1937) test continued to be the most powerful and the modified O'Brien (1979, 1981) was still the least powerful.

**Indirect pairing.** For all five procedures, power was lower when there was indirect pairing versus direct pairing. Notably, of the four procedures that were able to control Type I error rate at the nominal level (viz., Bartlett's, 1937, test; Brown and Forsythe's, 1974, test; the score test; and the modified O'Brien's, 1979, 1981, test), the rank order of these

**Table 4.** Statistical Power as a Function of Sample Size, Degree of Between-Groups Heteroscedasticity, and Type of Pairing When Subgroup Proportions ( $n_1:n_2$ ) Are 1:3.

N	Direct pairing				Indirect pairing					
	Heuristic	Bartlett	Brown-Forsythe	Score	O'Brien	Heuristic	Bartlett	Brown-Forsythe	Score	O'Brien
${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.25$										
60	.4592	.0964	.0824	.0578	.0442	.4046	.0756	.0556	.0818	.1088
120	.3356	.1240	.1082	.0924	.0736	.2572	.1042	.0856	.1166	.1408
180	.2742	.1594	.1416	.1268	.1020	.2070	.1410	.1234	.1558	.1820
240	.2224	.1872	.1632	.1580	.1292	.1762	.1756	.1538	.1894	.2188
${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:1.5$										
60	.5900	.1662	.1394	.1036	.0698	.4854	.1418	.1032	.1644	.2020
120	.5522	.2804	.2392	.2212	.1682	.4662	.2582	.2104	.2862	.3304
180	.5448	.3892	.3458	.3380	.2844	.4560	.3554	.2998	.3838	.4224
240	.5354	.4862	.4256	.4380	.3824	.4682	.3732	.3106	.3994	.4364
${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:2$										
60	.8010	.3668	.2906	.2398	.1530	.6924	.3202	.2454	.3580	.4186
120	.8722	.6490	.5590	.5664	.4664	.7812	.6064	.5160	.6388	.6708
180	.9022	.8084	.7436	.7650	.6960	.8466	.7816	.7074	.8052	.8278
240	.9310	.9102	.8590	.8906	.8490	.8900	.8902	.8378	.9024	.9160
${}_1\sigma_e^2 : {}_2\sigma_e^2 = 1:3$										
60	.9688	.7104	.5938	.5690	.3822	.8992	.6682	.5598	.7082	.7392
120	.9948	.9602	.9144	.9380	.8734	.9736	.9276	.8824	.9366	.9476
180	.9994	.9960	.9878	.9928	.9836	.9942	.9894	.9774	.9920	.9938
240	.9998	.9996	.9980	.9990	.9988	.9992	.9992	.9980	.9996	.9994

Note. N = total sample size;  $n_1$  = sample size in Group 1;  $n_2$  = sample size in Group 2;  ${}_1\sigma_e^2$  = population error variance in Group 1;  ${}_2\sigma_e^2$  = population error variance in Group 2;  $f^2 = .002$ .

procedures changed when there was indirect pairing. When there was indirect pairing (see Table 3 and Table 4), the modified O'Brien (1979, 1981) tended to be the most powerful, followed by the score test, Bartlett's (1937) test, and Brown and Forsythe's (1974) test.

It deserves noting that the heuristic method had the greatest statistical power of all five procedures because of its inflated Type I error rate. However, as its Type I error rate decreased with increasing  $N$ , the heuristic method has power that becomes similar to the other four procedures (see Table 4).

## Discussion

Because between-groups heteroscedasticity is a problem in MMR analyses with categorical moderators, the present study compared the performance of various procedures that could be used to detect this statistical violation. As noted above, research has focused primarily on remedial procedures that can be used when between-groups heteroscedasticity exists. However, we felt that it was also important to compare different ways of detecting between-groups heteroscedasticity that have not been previously examined empirically in MMR with a dichotomous moderator (viz., heuristic method; Bartlett's, 1937, test; Brown and Forsythe's,

1974, test; score test; and modified O'Brien's, 1979, 1981, test). By comparing various procedures, we hoped to offer some initial recommendations for researchers and practitioners in the behavioral and social sciences.

A number of key findings can be gleaned from our study. In general, Bartlett's (1937) test is the most powerful in detecting between-groups heteroscedasticity when sample sizes are equal or when direct pairing occurs, thus providing empirical support for the recommendation offered by DeShon and Alexander (1996). It is noteworthy, however, that when there is indirect pairing (i.e., the largest  $j\sigma_e^2$  paired with the smallest  $n_j$ ), the modified O'Brien test (1979, 1981) appears to be the most powerful procedure. This suggests that multiple statistical procedures may be necessary when diagnosing between-groups heteroscedasticity, such that one procedure should be used for direct pairing and a different procedure should be used for indirect pairing.

The score test performed well across conditions, typically with the second highest power levels. Perhaps due to its origins in econometrics and statistics, it does not appear to be well known in the psychology literature and related fields. However, the score test may still be a very attractive alternative for researchers because of its flexibility to detect heteroscedasticity of various forms, including between-groups heteroscedasticity.

Brown and Forsythe's (1974) test was the least powerful across conditions. It deserves noting, however, that this procedure was developed to be robust against violations of normality. Because normality was not manipulated in the present study, it is plausible that under conditions of non-normality, in which research has suggested that Bartlett's (1937) test performs poorly (Box, 1953; Levene, 1960), Brown and Forsythe's test could potentially outperform Bartlett's test. Similarly, because O'Brien's (1979, 1981) method has been found to be robust when the normality assumption is violated, it is possible that, under conditions of non-normality, the modified O'Brien could outperform Bartlett's test even in the direct pairing conditions.

As  $N$  increases and the degree of between-groups heteroscedasticity increases, the differences in power among the five procedures are not substantial. For the conditions considered in the present study, it appears that for  $Ns \geq 240$ , it generally makes little difference which procedure is used, especially if there is a high degree of between-groups heteroscedasticity.

The present study demonstrated that O'Brien's (1979, 1981) procedure can be extended to designs beyond one-way and factorial ANOVA to include continuous predictors. The modified procedure controlled Type I error at the nominal level and had power levels comparable with, and in some cases greater than, other procedures.

The heuristic method had very poor properties. Admittedly, it is not a statistical test. Thus, it may not be reasonable to expect the heuristic method to be robust. Note that the empirical rejection rates (i.e., Type I error and power) for the heuristic method are unaffected by whether  $\alpha = .01, .05$ , or  $.10$ . Thus, at any alpha, for the 1,200 conditions considered in the present study, the heuristic method would have the same rejection rates. To counteract its inflated Type I error rate, and interpolating from Table 1, the heuristic method may be recommended for use when  $k = 2$  and  $N > 200$ .

### Recommendations for Research and Practice

A few recommendations for research and practice can be identified. First, when testing for the equality of regression slopes, it is important that researchers and practitioners evaluate whether the homoscedasticity assumption has been satisfied. Consistent with Rosopa et al. (2013), the residuals from Equation 1 (for the two-group case, specifically) or Equation A1 in the appendix (for two or more groups, more generally) should be calculated. Then, the sample-based variance of these residuals can be calculated separately for each group. Assuming that  $N > 200$ , a simple ratio of the largest to the smallest residual variance can be calculated. In addition, direct pairing exists if the largest group has the largest residual variance; alternatively, indirect pairing exists if the largest group has the smallest residual variance. As subgroup sample sizes become increasingly disproportionate, it becomes increasingly

important to know whether direct pairing or indirect pairing exists.

Second, based on the results of the present study, Bartlett's (1937) test is the most powerful for detecting between-groups heteroscedasticity when the normality assumption is satisfied and direct pairing exists. However, when there is indirect pairing, the modified O'Brien's (1979, 1981) test should be used. Notably, if subgroup sample sizes are approximately equal, it makes little difference which statistical procedure is used because the differences in statistical power are generally small.

Third, if between-groups heteroscedasticity is detected, an alternative procedure should be used instead of ordinary least squares regression. To mitigate the biasing effects of between-groups heteroscedasticity, Rosopa et al. (2013) discussed a number of procedures including weighted least squares regression and heteroscedasticity-consistent covariance matrices.

### Conclusion

The present study adds incrementally to the extant literature on statistical procedures that can be used to detect between-groups heteroscedasticity in MMR with categorical moderators. It appears that different procedures may be needed to optimally detect between-groups heteroscedasticity when there is direct pairing (viz., Bartlett's, 1937, test) versus indirect pairing (viz., modified O'Brien's, 1979, 1981, test). This is a finding unique to this study. Moreover, because the heuristic method has never been empirically examined, the present simulation results are the first to note that this method has very inflated Type I error rates and it may be best to use this method when  $N > 200$  to counteract the inflated Type I error rates. In addition to comparing the performance of various procedures, we proffered a modification to O'Brien's (1979, 1981) method, which can be added to the statistical tools used by researchers and practitioners in the behavioral and social sciences.

### Appendix

The full linear model for the  $N = \sum_{j=1}^k n_j$  observations (with the number of terms  $p = 2k - 1$ ) can be compactly expressed in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (A1)$$

where  $\mathbf{y}$  is an  $N \times 1$  response vector,  $\mathbf{X}$  is an  $N \times (p + 1)$  model matrix,  $\boldsymbol{\beta}$  is a  $(p + 1) \times 1$  vector of unstandardized regression coefficients, and  $\mathbf{e}$  is an  $N \times 1$  residual vector. In addition, it is assumed that the first-order and second-order moments of  $\mathbf{e}$  have  $E(\mathbf{e}) = \mathbf{0}$  and  $\text{cov}(\mathbf{e}) = \sigma_e^2 \mathbf{I}_N$ , respectively (where  $\mathbf{0}$  = a null vector,  $\sigma_e^2$  = the common variance, and  $\mathbf{I}_N$  = an identity matrix of order  $N$ ; Schott, 2005).

The best linear unbiased estimator of the parameters in Equation A1 is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (\text{A2})$$

Although  $\mathbf{X}$ , in Equation A2, can be partitioned differently, for convenience:

$$\mathbf{X} = [\mathbf{j} \ \mathbf{x} \ \mathbf{D}_z \ \mathbf{D}_{xz}], \quad (\text{A3})$$

where  $\mathbf{j}$  is an  $N \times 1$  vector of 1s,  $\mathbf{x}$  is an  $N \times 1$  vector for the continuous predictor,  $\mathbf{D}_z$  is an  $N \times (k - 1)$  matrix of regressors, and  $\mathbf{D}_{xz}$  is an  $N \times (k - 1)$  matrix of product terms between  $\mathbf{x}$  and the regressors in  $\mathbf{D}_z$ . Based on Equation A2 and the constant variance assumption (i.e., homoscedasticity), an unbiased estimator of  $\sigma_e^2$  can be expressed as

$$\hat{\sigma}_e^2 = \frac{(\mathbf{y} - \hat{\mathbf{X}}\hat{\beta})'(\mathbf{y} - \hat{\mathbf{X}}\hat{\beta})}{N - p - 1} = \frac{SSE}{N - p - 1}, \quad (\text{A4})$$

where  $SSE$  = sum of squared errors. Moreover, when  $\mathbf{e}$  is normally distributed, maximum likelihood estimators of  $\beta$  and  $\sigma_e^2$ , respectively, are given by Equation A2 and  $SSE / N$  (Rencher, 2000).

Although  $\mathbf{X}$  in Equation A3 represents the full model matrix, a full-and-reduced linear model approach can be used to construct the test of whether the  $k$  population regression slopes are equal (Rencher, 2000). The reduced model matrix ( $\mathbf{X}_{\text{Reduced}}$ ) excludes  $\mathbf{D}_{xz}$ . Then,  $\beta_{\text{Reduced}}$  becomes a  $(k + 1) \times 1$  vector of regression coefficients.

Assuming that  $\mathbf{e}$  is normally distributed, the test for the equality of regression slopes is conducted using an  $F$  ratio. It assesses whether the decrease in the  $SSE$  from a reduced ( $SSE_{\text{Reduced}}$ ) to a full ( $SSE_{\text{Full}}$ ) model is statistically significant. The  $F$  random variable can be expressed as

$$F = \frac{(SSE_{\text{Reduced}} - SSE_{\text{Full}})/df_1}{SSE_{\text{Full}}/df_2}, \quad (\text{A5})$$

where  $df_1$  = the number of terms omitted from the full model and  $df_2$  = the error degrees of freedom for the full model. It is worth noting that an equivalent general linear hypothesis test can be conducted using the full model in Equation A3 (see Equation 8.27 in Rencher, 2000). If  $F > F(1 - \alpha, df_1, df_2)$  (where  $\alpha$  = Type I error rate), then the null hypothesis of equal regression slopes is rejected; stated differently,  $z$  moderates the relation between  $y$  and  $x$ . Otherwise, the null hypothesis of equal regression slopes cannot be rejected. These procedures are described in greater detail in numerous texts (see Cohen, Cohen, West, & Aiken, 2003; Fox, 2008; Maxwell & Delaney, 2000; Neter, Kutner, Nachtsheim, & Wasserman, 1996). Note that when  $k = 2$ , the test of the moderating effect based on the  $F$  ratio in Equation A5 is equivalent to a two-tailed  $t$  test with  $df_2 = N - 4$ .

## Authors' Note

Portions of this article were presented at the 75th annual conference of the Psychometric Society in Athens, Georgia, and the 26th annual

conference of the Society for Industrial and Organizational Psychology in Chicago, Illinois.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research and/or authorship of this article.

## Notes

1. In the literature, between-groups heteroscedasticity has also been labeled Type II heteroscedasticity (Wilcox, 1997) and heterogeneity of subgroup error variance (Aguinis & Pierce, 1998). Because heteroscedasticity can take many forms (Fox, 2008; Long & Ervin, 2000; Rosopa, Schaffer, & Schroeder, 2013), we adopt the term “between-groups heteroscedasticity” (Ng & Wilcox, 2010) because we believe that it concisely describes the specific form that can manifest in moderated multiple regression (MMR) with a categorical moderator.
2. Historically, in the univariate two-sample case involving independent means, heteroscedasticity has been referred to as the Behrens–Fisher problem (Miller, 1997; Rencher, 1998) after the researchers who proposed approximate solutions in the 1920s and 1930s. In addition, numerous approximate solutions have been proposed for the multivariate version of the Behrens–Fisher problem (Kim, 1992; Rencher, 1998).
3. Note that by modifying O'Brien's (1979, 1981) procedure, it can also be used to detect between-groups heteroscedasticity when the categorical moderator has more than two levels (e.g., Treatment A, Treatment B, vs. Control; varying levels of race or ethnicity). These formulas are available from the first author. However, for simplicity in our exposition, we focus on a dichotomous moderator (see Equation 1).
4. For a more robust test, Welch's (1938) procedure can be applied to the transformed residuals from Equation 7.
5. The values for various parameters in one of the experimental conditions were as follows:  $N = 120$ ,  $n_1 = 40$ ,  $n_2 = 80$ ,  ${}_1\sigma_e^2 = 1.25$ ,  ${}_2\sigma_e^2 = 1$ ,  ${}_1\sigma_x = {}_2\sigma_x = 1$ ,  ${}_1\beta_{yx} = 0.5$ ,  ${}_2\beta_{yx} = 0.598045431309419$ ,  ${}_1\sigma_y = 1.22474487139159$ ,  ${}_2\sigma_y = 1.16518596709284$ ,  ${}_1\rho_{yx} = 0.408248290463863$ ,  ${}_2\rho_{yx} = 0.513261786701358$ , and  $f^2 = .00199995713664686$ . Note that, given the values of the manipulated parameters (i.e.,  $N$ ,  $n_1, n_2$ ,  ${}_1\sigma_e^2 : {}_2\sigma_e^2$ , and type of pairing),  $f^2$  was not precisely equal to .002. We refer the reader to DeShon and Alexander (1996) and Dretzke et al. (1982) for a similar overview of the data generation process.

## References

- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford.
- Aguinis, H., Beatty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, 90, 94–107. doi:10.1037/0021-9010.90.1.94

- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, 95, 648-680. doi:10.1037/a0018714.
- Aguinis, H., Peterson, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods*, 2, 315-339.
- Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods*, 1, 296-314. doi:10.1177/109442819813002
- Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, 52, 192-206. doi:10.1037/0021-9010.82.1.192
- 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. doi:10.1037/0033-2909.115.2.308
- Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, 19, 91-101. doi:10.3102/10769986019002091
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society, A160*, 268-282. doi:10.1098/rspa.1937.0109
- Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, 40, 373-400. doi:10.1207/s15327906mbr4003\_5
- Boos, D. D., & Brownie, C. B. (2004). Comparing variances and other measures of dispersion. *Statistical Science*, 19, 571-578. doi:10.1214/088342304000000503
- Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40, 318-335. doi:10.2307/2333350
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47, 1287-1294. doi:10.2307/1911963
- Brown, M. B., & Forsythe, A. B. (1974). Robust test for the equality of variances. *Journal of the American Statistical Association*, 69, 364-367. doi:10.2307/2285659
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, 23, 351-361. doi:10.2307/1268225
- Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70, 1-10. doi:10.2307/2335938
- Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis*, 45, 215-233. doi:10.1016/S0167-9473(02)00366-3
- Culpepper, S. A., & Aguinis, H. (2011). R is for revolution: A cutting-edge, free, open source statistical package. *Organizational Research Methods*, 14, 735-740. doi:10.1177/1094428109355485
- 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. doi:10.1177/0013164494054002007
- 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. doi:10.1037/1082-989X.1.3.261
- Dretzke, B. J., Levin, J. R., & Serlin, R. C. (1982). Testing for regression homogeneity under variance heterogeneity. *Psychological Bulletin*, 91, 376-383. doi:10.1037/0033-2909.91.2.376
- Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd ed.). Thousand Oaks, CA: SAGE.
- Games, P. A., Winkler, H. B., & Probert, D. A. (1972). Robust tests for homogeneity of variance. *Educational and Psychological Measurement*, 32, 887-909.
- Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs*, 58, 437-448. doi:10.1080/03637759109376240
- Hattrup, K., & Schmitt, N. (1990). Prediction of trades apprentices' performance on job sample criteria. *Personnel Psychology*, 43, 453-466. doi:10.1111/j.1744-6570.1990.tb02392.x
- Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York, NY: Wiley.
- 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. doi:10.1037/0033-2909.86.4.721
- Kim, S.-J. (1992). A practical solution to the multivariate Behrens-Fisher problem. *Biometrika*, 79, 171-176. doi:10.1093/biomet/79.1.171
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics* (pp. 278-292). Stanford, CA: Stanford University Press.
- Lix, L. M. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "F" test. *Review of Educational Research*, 66, 579-619. doi:10.2307/1170654
- Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54, 217-224. doi:10.2307/2685594
- Martin, C. G., & Games, P. A. (1977). ANOVA tests for homogeneity of variance: Nonnormality and unequal sample sizes. *Journal of Educational Statistics*, 2, 187-206. doi:10.2307/1164993
- Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Lawrence Erlbaum.
- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376-390. doi:10.1037/0033-2909.114.2.376
- Miller, R. G. (1997). *Beyond ANOVA: Basics of applied statistics*. London, England: Chapman & Hall.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago, IL: McGraw-Hill.

- Ng, M., & Wilcox, R. R. (2010). Comparing the regression slopes of independent groups. *British Journal of Mathematical and Statistical Psychology*, 63, 319-340. doi:10.1348/000711009X456845
- O'Brien, R. G. (1979). A general ANOVA method for robust tests of additive models for variances. *Journal of the American Statistical Association*, 74, 877-880. doi:10.2307/2286416
- O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin*, 89, 570-574. doi:10.1037//0033-2909.89.3.570
- Oswald, F. L., Saad, S., & Sackett, P. R. (2000). The homogeneity assumption in differential prediction analysis: Does it really matter? *Journal of Applied Psychology*, 85, 536-541. doi:10.1037//0021-9010.85.4.536
- Overton, R. C. (2001). Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods*, 6, 218-233. doi:10.1037//1082-989X.6.3.218
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>
- Rencher, A. C. (1998). *Multivariate statistical inference and applications*. New York, NY: Wiley.
- Rencher, A. C. (2000). *Linear models in statistics*. New York, NY: Wiley.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York, NY: Springer.
- Rosopa, P. J. (2006, May). An alternative solution for heterogeneity of variance across categorical moderators in moderated multiple regression. In D. Newman (Chair), *Testing interaction effects: Problems and procedures*. Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods*, 18, 335-351. doi:10.1037/a0032553
- Rutherford, A. (1992). Alternatives to traditional analysis of covariance. *British Journal of Mathematical and Statistical Psychology*, 45, 197-223.
- Saad, S., & Sackett, P. R. (2002). Investigating differential prediction by gender in employment-oriented personality measures. *Journal of Applied Psychology*, 87, 667-674. doi:10.1037//0021-9010.87.4.667
- Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of source adjustment in preemployment testing. *American Psychologist*, 49, 929-954. doi:10.1037/0003-066X.49.11.929
- Salgado, J. F. (1998). Sample size in validity studies of personnel selection. *Journal of Occupational and Organizational Psychology*, 71, 161-164.
- Sarkar, S., Kim, C., & Basu, A. (1999). Tests for homogeneity of variances using robust weighted likelihood estimates. *Biometrical Journal*, 41, 857-871.
- Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, 16, 209-222.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177. doi:10.1037//1082-989X.7.2.147
- Schott, J. R. (2005). *Matrix analysis for statistics* (2nd ed.). Hoboken, NJ: Wiley.
- Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods*, 5, 230-240. doi:10.1037/1082-989X.5.2.230
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shieh, G. (2009). Detection of interactions between a dichotomous moderator and a continuous predictor in moderated multiple regression with heterogeneous error variance. *Behavior Research Methods*, 41, 61-74. doi:10.3758/BRM.41.1.61
- Stone-Romero, E. F., & Liakhovitski, D. (2002). Strategies for detecting moderator variables: A review of conceptual and empirical issues. *Research in Personnel and Human Resources Management*, 21, 333-372. doi:10.1016/S0742-7301(02)21008-7
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350-362. doi:10.2307/2332010
- Wilcox, R. R. (1997). Comparing the slopes of two independent regression lines when there is complete heteroscedasticity. *British Journal of Mathematical and Statistical Psychology*, 50, 309-317.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). New York, NY: Elsevier.
- Zedeck, S. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin*, 76, 295-310. doi:10.1037/h0031543

## Author Biographies

**Patrick J. Rosopa** is an associate professor of industrial-industrial organizational psychology in the Department of Psychology at Clemson University. He has co-authored a book titled *Statistical Reasoning in the Behavioral Sciences* (6th ed., 2010, Wiley). His research has been published in such outlets as *Psychological Methods*, *Organizational Research Methods*, *Human Resource Management Review*, *Personality and Individual Differences*, and *Scandinavian Journal of Psychology*.

**Amber N. Schroeder** is an assistant professor in the Department of Psychological Sciences at Western Kentucky University. Her research interests focus primarily on (a) the use of social media in employment settings, (b) the examination of negative employee and organizational behavior, and (c) the assessment of employee personality and culture and their impact on work outcomes. Her research has been published in such outlets as *Psychological Bulletin*, *Psychological Methods*, *Journal of Occupational Health Psychology*, and *Journal of Managerial Psychology*.

**Jessica L. Doll** is an assistant professor in the Department of Management and Decision Sciences at Coastal Carolina University. Her research interests include workplace romances, impression management and political skill, and cross-cultural and gender differences in selection and engagement. She has presented at national and international refereed conferences and her research has been published in *Journal of Managerial Psychology* and *Journal of Organizational Behavior Management*.