Journal of Applied Psychology
2006, Vol. 91, No. 4, 917–926

0021-9010/06/$12.00 DOI: 10.1037/0021-9010.91.4.917

RESEARCH REPORTS

Probing Three-Way Interactions in Moderated Multiple Regression: Development and Application of a Slope Difference Test

Jeremy F. Dawson and Andreas W. Richter
Aston University

Researchers often use 3-way interactions in moderated multiple regression analysis to test the joint effect of 3 independent variables on a dependent variable. However, further probing of significant interaction terms varies considerably and is sometimes error prone. The authors developed a significance test for slope differences in 3-way interactions and illustrate its importance for testing psychological hypotheses. Monte Carlo simulations revealed that sample size, magnitude of the slope difference, and data reliability affected test power. Application of the test to published data yielded detection of some slope differences that were undetected by alternative probing techniques and led to changes of results and conclusions. The authors conclude by discussing the test's applicability for psychological research.

Keywords: moderated multiple regression, three-way interactions, slope difference test

The concept of moderation in social science is central to testing theories that aim to explain the interactive effect of two or more variables in predicting a dependent variable. Many such contingency theories represent crucial developments in various disciplines, including applied psychology, organizational behavior, administrative science, and sociology.

A variable *Z* is a moderator variable of the relation between an independent variable *X* and a dependent variable *Y*, when the magnitude of this relation varies across levels of *Z* (Baron & Kenny, 1986; James & Brett, 1984; Zedeck, 1971). *Z* can be a continuous or a categorical variable. Frequently, the relation between *X* and *Y* may depend on more than one variable. The basic concept and rationale of moderation may then be generalized from two-way interactions to more complex three-way interactions, in which the relation between *X* and *Y* is contingent not only on *Z* but also on another moderator variable, *W* (as well as the interplay of *Z* and *W*; Jaccard & Turrisi, 2003). Such three-way interactions serve to examine the concerted interplay of several variables and can be used to test configurational theories, typologies, or more complex contingency theories (e.g., Brown, Ganesan, & Challagala, 2001; Erez & Earley, 1987; Oldham & Fried, 1987).

Detecting Interactions, Plotting Interactions, and Post Hoc Probing

Detecting Interactions

Moderated multiple regression (MMR) analysis is the method of choice to detect moderator effects in field research and is superior to strategies such as comparison of subgroup-based correlation coefficients (Stone-Romero & Anderson, 1994). The most common procedure to test two-way interactions statistically is to regress a dependent variable *Y* on the independent variable *X*, the moderator variable *Z*, and the product (interaction) term of *X* and *Z* (*XZ*). This interaction term is often entered in a separate step, although this is not essential. A significant interaction term *XZ* indicates that the effect of *X* on *Y* differs across the range of the moderator variable *Z* (Cohen & Cohen, 1975; Peters, O'Connor, & Wise, 1984; Zedeck, 1971). Although in the present article we express this in terms of an independent variable *X* and a moderator *Z*, we note that both variables *X* and *Z* might function as a moderator or as a predictor variable, leaving the distinction to theoretical reasons (Baron & Kenny, 1986; James & Brett, 1984).

In the case of three-way interactions, this procedure can then be generalized to test the effect of *X* on *Y* depending on two moderator variables, *Z* and *W*. Analogously, *Y* is regressed on the variables *X*, *Z*, and *W*; the products of each pair of variables, that is, *XZ*, *XW*, and *WZ* (often in a separate step); and the product term of all three predictor variables, that is, *XWZ* (again, often in a separate step). Similar to the two-way interaction case, the significance of the three-way interaction term indicates that the relation between *X* and *Y* varies across levels of *Z*, *W*, and/or the combination of *Z* and *W*. As with the two-way interaction scenario, we leave to theoretical reasons which variables function as moderators or the independent variable.

Jeremy F. Dawson and Andreas W. Richter, Work and Organisational Psychology Group, Aston Business School, Aston University, Birmingham, United Kingdom.

Andreas W. Richter is now at the Instituto de Empresa Business School, Madrid, Spain.

We thank Rolf van Dick for his comments on a draft of this article. We are most grateful to James Smither, Nico van Yperen, and Jessica Kwong for providing their data for this analysis. Even though we criticize the way these authors further probed three-way interactions, this criticism does not refer to the research itself. We consider their research excellent and their findings of strong scientific relevance.

Correspondence concerning this article should be addressed to Jeremy F. Dawson, Work and Organisational Psychology Group, Aston Business School, Aston University, Birmingham B4 7ET, United Kingdom. E-mail: j.f.dawson@aston.ac.uk

Plotting the Interaction

One way to further interpret the three-way interaction involves plotting the relation between *X* and *Y*, at high and low values of *Z* and *W* (for a discussion of plotting techniques, see Aiken & West, 1991; Jaccard & Turrisi, 2003). Such a plot allows a quick, visual indication of the nature of an interaction effect, and the direction of the slopes can be interpreted on the basis of face validity (see Figure 1 for an illustration). Alternatively, the relation between *X* and *Y* at high and low levels of one of the moderators (e.g., *Z*) may be plotted in two separate graphs at high and low levels of the other (e.g., *W*). Interaction plots are a useful means for illustration purposes. However, they do not allow inferences as to whether a significant three-way interaction is the result of significant differences among any two, three, or all four combinations of the two moderator variables *Z* and *W* at high and low levels; whether any difference between pairs of slopes is significant; or whether an individual slope is a significant predictor of the dependent variable. Further statistical probing is required to answer these questions.

Post Hoc Probing Techniques

The most popular procedure to further examine significant interaction terms statistically is the *pick-a-point* approach (Aiken & West, 1991; Rogosa, 1980). Researchers typically conduct this analysis by first computing simple slopes of *Y* on *X* at conditional values (e.g., high and low levels) of *Z* and then testing whether simple slopes at combinations of high and low values of *Z* and *W* differ significantly from zero in predicting the dependent variable. Calculation of confidence intervals for simple slopes (Cohen, Cohen, West, & Aiken, 2003) represents a refined advancement of this method. These tests may be interpreted such that a significant slope at high or low levels of *Z* and *W* represents a significant relation between the independent and dependent variables at those values, irrespective of what this relation is at other values of *Z* and *W*. Limitations of this technique have been discussed elsewhere (see Aiken & West, 1991; Bauer & Curran, 2005; Rogosa, 1980), including that conditional values may not reflect the entire range of a continuous moderator variable and are sometimes chosen somewhat arbitrarily. Another probing technique developed in response to this criticism is the calculation of regions of significance via the Johnson–Neyman technique (Bauer & Curran, 2005), which, to our knowledge, so far only has been developed for two-way interactions. For the purpose of this article, it is important to stress one additional limitation of the pick-a-point technique: These post hoc tests do not allow the conclusion that the impact of the independent variable on the dependent variable is significantly more different for one slope at combinations of high and low conditional values of *Z* and *W* than for any other slope. These comparisons do thus not present a *relational* test of two simple slopes but are only an *absolute* test of a single simple slope.

A commended alternative approach for post hoc probing is the use of subgroup analysis (see Aiken & West, 1991; Peters et al., 1984). The researcher may use a median split on one of the two moderator variables (e.g., *W*) and then further probe the relation of the independent variable (*X*) and the other moderator (*Z*) for each subgroup of high and low *W* separately. Further probing involves individual subgroup analysis—for instance, through the use of regression analysis. We argue that subgroup analyses bear three relevant limitations: First, they represent an artificial split into subgroups that is not in concordance with the (continuous) nature of the respective variable. Second, they limit the power of the analysis, as the sample size of subgroups is reduced. Third and foremost, they do not allow for a comparative test of slopes that do exist across the barriers of subgroups (i.e., a comparison of a slope at high *Z* and high *W* with a slope at low *Z* and low *W* or a comparison of a slope at high *Z* and low *W* with a slope at low *Z* and high *W*), as subgroups are treated as if they were separate samples. Subgroup analyses are thus restricted to examine slope differences within a subgroup only.

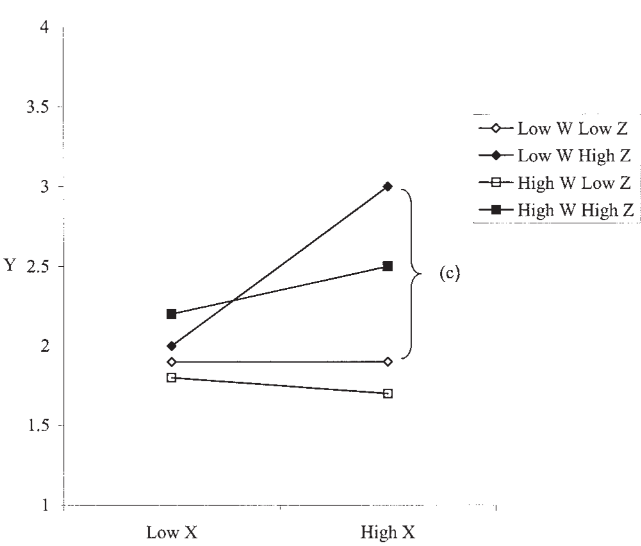

Figure 1. A three-way interaction plot illustrating simple slopes of *Y* on *X* at high and low values of *W* and *Z*. (c) = difference between simple slopes at low *W* and low *Z* and at low *W* and high *Z*.

Common Strategies in Probing Three-Way Interactions

A literature search in the PsycARTICLES, PsycINFO, and Proquest databases yielded three common strategies used by researchers to probe three-way interactions: Some authors did not further probe the significant three-way interaction term by statistical means but simply plotted the interaction and interpreted slope differences on the basis of face validity, that is, the direction of the slopes (e.g., Baker & Cullen, 1993; Schaubroeck, Ganster & Kemmerer, 1994; Schaubroeck & Merritt, 1997). As we have suggested, this approach tempts researchers to draw unjustified conclusions as to whether individual slopes differ from each other or are significant predictors of the dependent variable. A second group of researchers used the pick-a-point approach (Aiken & West, 1991; Rogosa, 1980) to examine whether simple slopes differed significantly from zero and further interpreted the interaction on the basis of interaction plots (e.g., Janssen, van de Vliert, & Veenstra, 1999). The question of which pairs of slopes differed from each other significantly was then left to speculation. A third (and perhaps most advanced) strategy was to carry out subgroup analysis on one moderator variable and subsequently run separate regression analyses for subgroups at low and high levels of this moderator (e.g., Dickson & Weaver, 1997; Kwong & Leung, 2002; Oldham & Cummings, 1996; Oldham & Fried, 1987). This approach results in ignorance of potential slope differences that might have existed across subgroups; furthermore, it uses the very type of arbitrary split (of at least one variable) that MMR is designed to overcome.

None of the three strategies allows the researcher to test each possible combination of pairs of simple slopes for statistical significance. A precise test of an explicit psychological hypothesis of slope differences, as, for instance, in the Schaubroek et al. (1994) study, is not possible if the interaction is only plotted. Also, other researchers might have refined their hypotheses had a test for slope differences indicated that some slopes differed significantly but other slopes did not (e.g., Xie, 1996). Such a significance test of the difference between simple slopes, in fact, does not yet exist (Cohen et al., 2003, p. 280).

In the remainder of the article, we first theoretically derive a formula to test slope differences for statistical significance. We then use Monte Carlo simulations to determine how sample characteristics affect the power of this test. We continue demonstrating the test's applicability by reanalyzing published data. We conclude by discussing test usefulness and limitations.

The Development of a Significance Test for Slope Differences

The development of the test formula contains four steps: First, one calculates generic formulas for simple slopes of the relation between X and Y at high and low levels of Z and W (in a similar fashion to Aiken & West, 1991). Second, one calculates the difference between any two pairs of slopes (Δslope). Third, one calculates the standard error ($SE_{\Delta\text{slope}}$) of the difference of pairs of slopes. To determine whether slopes differ from each other, it is necessary to put the slope difference in relation to its standard error. Therefore, the final step requires one to test whether the ratio of the difference between pairs of slopes and its standard error ($\Delta\text{slope}/SE_{\Delta\text{slope}}$) differs from zero.

Generic Formulas for Simple Slopes of X on Y at High and Low Levels of Z and W

The generic form of the three-way interaction regression equation can be represented as follows:

$$Y = b_0 + b_1X + b_2Z + b_3W + b_4XZ + b_5XW + b_6ZW + b_7XZW + \varepsilon.$$

A typical plot of a three-way interaction would consist of four lines. Each of these lines would show the (estimated) relation between X and Y under one of four conditions: (a) Z high, W high; (b) Z high, W low; (c) Z low, W high; and (d) Z low, W low. Normally, the high and low values of Z and W are taken to be 1 standard deviation above and below the mean values of the variables (Aiken & West, 1991). To leave this as general as possible, however, we refer to high and low values of Z and W as Z_H, Z_L, W_H, and W_L, respectively.

The lines are described by the regression equation given above, with substitution of parameter estimates for b_0 to b_7 and of Z_H, Z_L, W_H, and W_L for Z and W as appropriate. The regression equation can be rewritten as

$$Y = (b_0 + b_2Z + b_3W + b_6ZW) + (b_1 + b_4Z + b_5W + b_7ZW)\,X + \varepsilon.$$

The part of this equation in the first set of parentheses represents the intercept on a graph of Y against X; the part in the second set of parentheses represents the gradient, or slope, of the line. Therefore, the slopes of the four lines can be represented by

$$b_1 + b_4Z_H + b_5W_H + b_7Z_HW_H, \tag{1}$$

$$b_1 + b_4Z_H + b_5W_L + b_7Z_HW_L, \tag{2}$$

$$b_1 + b_4Z_L + b_5W_H + b_7Z_LW_H, \tag{3}$$

and

$$b_1 + b_4Z_L + b_5W_L + b_7Z_LW_L. \tag{4}$$

Calculating Differences Between Pairs of Slopes

The quantity of interest here is the difference between these slopes. In total, there are six pairs of slopes that may potentially be of interest for testing between. These fall into two categories: Slopes between which only one variable changes (e.g., Equations 1 and 2; W changes between high and low, but Z remains high in both cases), and slopes for which both variables change (e.g., Equations 1 and 4; both Z and W change from high to low). The difference between each pair of slopes is shown in Table 1.

Although these formulas appear somewhat complicated, if simple values of Z and W are substituted, they reduce considerably. For instance, if both variables have been standardized (i.e., centered with mean 0 and standard deviation 1) before the interaction terms are calculated and the regression performed and if the lines have been plotted for values of Z and W that are one standard deviation above and below the mean (i.e., $Z_H = W_H = 1$, and $Z_L = W_L = 1$), Difference a in Table 1 would reduce to $2(b_5 + b_7)$, and Difference e would reduce to $2(b_4 + b_5)$. Although we

Table 1
Formulas for Differences Among All Six Pairs of Slopes

Slopes	Difference	Label
1 and 2	$b_5\,(W_H - W_L) + b_7Z_H\,(W_H - W_L)$	a
1 and 3	$b_4\,(Z_H - Z_L) + b_7W_H\,(Z_H - Z_L)$	b
2 and 4	$b_4\,(Z_H - Z_L) + b_7W_L\,(Z_H - Z_L)$	c
3 and 4	$b_5\,(W_H - W_L) + b_7Z_L\,(W_H - W_L)$	d
1 and 4	$b_4\,(Z_H - Z_L) + b_5\,(W_H - W_L) + b_7\,(Z_HW_H - Z_LW_L)$	e
2 and 3	$b_4\,(Z_H - Z_L) + b_5\,(W_L - W_H) + b_7\,(Z_HW_L - Z_LW_H)$	f

recommend that such standardization take place, we continue to describe the general situation and proceed to describe the simplification achieved by the use of standardized variables.

Calculating the Standard Error of Slope Differences

To determine whether a difference in slopes is significant, it is necessary to compare the difference with its standard error. It can be shown that the standard errors of the slope differences are given by the formulas in Table 2. The formulas rely on the variances and covariances of the regression estimates b_4, b_5, and b_7; in keeping with Aiken and West's (1991) notation, we refer to the variance of regression estimate b_4 as s_{44} and the covariance of estimates b_4 and b_5 as s_{45}. For a complete derivation of these formulas, see the Appendix. Note that the formulas for the cases in which variables are standardized are far simpler than the general formulas.

A Significance Test of Differences Between Pairs of Slopes

The difference between two slopes itself represents a relation between Y and X (for a certain change in at least one of Z and W) and, as such, is similar to a slope in its own right. Therefore the ratio (slope)/(standard error[slope]) has, under the assumption that the slopes are equal, a t distribution with $(n - k - 1)$ degrees of freedom, where n is the sample size and k is the total number of predictors in the regression equation (including all the interaction terms). Thus, to test the significance of Slope Difference a in Table 1, it is necessary to test the hypothesis

$$t = \frac{(\mathrm{a})}{SE_{\mathrm{a}}} = \frac{b_5(W_H - W_L) + b_7 Z_H (W_H - W_L)}{(W_H - W_L)\sqrt{s_{55} + Z_H^2 s_{77} + 2Z_H s_{57}}} = \frac{b_5 + b_7 Z_H}{\sqrt{s_{55} + Z_H^2 s_{77} + 2Z_H s_{57}}} \neq 0,$$

Table 2
Standard Errors of Differences Between All Six Pairs of Slopes and Associated Test Statistics

Slopes	Variable	Standard error (difference)	Test statistic
a. 1 and 2	General	$(W_H - W_L)\sqrt{s_{55} + Z_H^2 s_{77} + 2Z_H s_{57}}$	$t = \frac{b_5 + b_7 Z_H}{\sqrt{s_{55} + Z_H^2 s_{77} + 2Z_H s_{57}}} \neq 0$
	Standardized	$2\sqrt{s_{55} + s_{77} + 2s_{57}}$	$t = \frac{b_5 + b_7}{\sqrt{s_{55} + s_{77} + 2s_{57}}} \neq 0$
b. 1 and 3	General	$(Z_H - Z_L)\sqrt{s_{44} + W_H^2 s_{77} + 2W_H s_{47}}$	$t = \frac{b_4 + b_7 W_H}{\sqrt{s_{44} + W_H^2 s_{77} + 2W_H s_{47}}} \neq 0$
	Standardized	$2\sqrt{s_{44} + s_{77} + 2s_{47}}$	$t = \frac{b_4 + b_7}{\sqrt{s_{44} + s_{77} + 2s_{47}}} \neq 0$
c. 2 and 4	General	$(Z_H - Z_L)\sqrt{s_{44} + W_L^2 s_{77} + 2W_L s_{47}}$	$t = \frac{b_4 + b_7 W_L}{\sqrt{s_{44} + W_L^2 s_{77} - 2W_L s_{47}}} \neq 0$
	Standardized	$2\sqrt{s_{44} + s_{77} - 2s_{47}}$	$t = \frac{b_4 + b_7}{\sqrt{s_{44} + s_{77} - 2s_{47}}} \neq 0$
d. 3 and 4	General	$(W_H - W_L)\sqrt{s_{55} + Z_L^2 s_{77} + 2Z_L s_{57}}$	$t = \frac{b_5 + b_7 Z_L}{\sqrt{s_{55} + Z_L^2 s_{77} - 2Z_L s_{57}}} \neq 0$
	Standardized	$2\sqrt{s_{55} + s_{77} - 2s_{57}}$	$t = \frac{b_5 + b_7}{\sqrt{s_{55} + s_{77} - 2s_{57}}} \neq 0$
e. 1 and 4	General	$\sqrt{(Z_H - Z_L)^2 s_{44} + (W_H - W_L)^2 s_{55} + (Z_H W_H - Z_L W_L)^2 s_{77} + 2\left[\begin{matrix}(Z_H - Z_L)(W_H - W_L)s_{45} \\ + (Z_H - Z_L)(Z_H W_H - Z_L W_L)s_{47} \\ + (W_H - W_L)(Z_H W_H - Z_L W_L)s_{57}\end{matrix}\right]}$	$t = \frac{b_4(Z_H - Z_L) + b_5(W_H - W_L) + b_7(Z_H W_H - Z_L W_L)}{\sqrt{(Z_H - Z_L)^2 s_{44} + (W_H - W_L)^2 s_{55} + (Z_H W_H - Z_L W_L)^2 s_{77} + 2\left[\begin{matrix}(Z_H - Z_L)(W_H - W_L)s_{45} \\ + (Z_H - Z_L)(Z_H W_H - Z_L W_L)s_{47} \\ + (W_H - W_L)(Z_H W_H - Z_L W_L)s_{57}\end{matrix}\right]}} \neq 0$
	Standardized	$2\sqrt{s_{44} + s_{55} + 2s_{45}}$	$t = \frac{b_4 + b_5}{\sqrt{s_{44} + s_{55} + 2s_{45}}} \neq 0$
f. 2 and 3	General	$\sqrt{(Z_H - Z_L)^2 s_{44} + (W_L - W_H)^2 s_{55} + (Z_H W_L - Z_L W_H)^2 s_{77} + 2\left[\begin{matrix}(Z_H - Z_L)(W_L - W_H)s_{45} \\ + (Z_H - Z_L)(Z_H W_L - Z_L W_H)s_{47} \\ + (W_L - W_H)(Z_H W_L - Z_L W_H)s_{57}\end{matrix}\right]}$	$t = \frac{b_4(Z_H - Z_L) + b_5(W_L - W_H) + b_7(Z_H W_L - Z_L W_H)}{\sqrt{(Z_H - Z_L)^2 s_{44} + (W_L - W_H)^2 s_{55} + (Z_H W_L - Z_L W_H)^2 s_{77} + 2\left[\begin{matrix}(Z_H - Z_L)(W_L - W_H)s_{45} \\ + (Z_H - Z_L)(Z_H W_L - Z_L W_H)s_{47} \\ + (W_L - W_H)(Z_H W_L - Z_L W_H)s_{57}\end{matrix}\right]}} \neq 0$
	Standardized	$2\sqrt{s_{44} + s_{55} - 2s_{45}}$	$t = \frac{b_4 - b_5}{\sqrt{s_{44} + s_{55} - 2s_{45}}} \neq 0$

where t follows a t distribution with $(n - k - 1)$ degrees of freedom (this is a generalization of the two-way interaction slopes test proposed by Aiken & West, 1991). Equivalent test formulas for the other slope differences, including the forms for standardized variables, are given in Table 2.

To conduct such a test, a researcher would need to run a regression analysis to test the three-way interaction (as exemplified in the introduction) but obtain the covariance matrix of the regression coefficients. In most statistics packages, such as SPSS, this function is not included by default but is available as an option. Having extracted the relevant coefficients and variances or covariances of the coefficients, one can easily calculate the relevant formula above and compare it against a table of values of the t distribution with $n - k - 1$ degrees of freedom.[1]

One can then use the test either to test a priori hypotheses or to explore detected interactions in a post hoc manner. Although in the former case a specific difference between two slopes has been hypothesized (e.g., Kwong & Leung, 2002), the researcher aims to interpret a three-way interaction effect, which is more exploratory in nature in the latter (e.g., Smither & Walker, 2004). These two cases require different conditions for acceptable levels of test power: If the researcher has hypothesized a difference (or more than one difference), then he or she can use the test as described in this article without necessary adjustment. If, conversely, the researcher discovers a three-way interaction effect and wishes to explore which differences are significant, a correction for multiple post hoc testing (analogous to adjustment in multiple tests for analysis of variance) is necessary. A common such adjustment is the Bonferroni correction, which approximately divides the accepted significance level of a test (α) by the number of simultaneous tests carried out (e.g., Miller, 1981).

Monte Carlo Simulation

In the following section, we aim to describe under which conditions the significance test will yield reliable results and when significant slope differences may remain undetected because of sample limitations. A computer program was written in S-Plus (Insightful Corporation, 2001) to generate multiple data sets, under different conditions, with underlying three-way interactions. Aguinis (1995, 2004) reviewed four areas that have been discussed to affect the power of MMR—variable distribution, operationalization of predictor and dependent variables, sample size, and characteristics of the predictor variables—as well as interactive effects of these four. We have selected one key criterion from each of these categories and examine their individual and combined effect on test power.

Method

Variable distribution. One characteristic of variable distribution known to affect power of MMR with categorical moderators is range restriction (reduction in the variance of the predictor; e.g., Aguinis, 1995; Aguinis & Stone-Romero, 1997). However, range restriction per se is unlikely to cause a problem for continuous moderators. A three-way interaction effect, if it is linear in nature (as the regression equation for continuous moderators assumes it is), should be equally detectable even when there is a lack of variance in a predictor, because the estimation of the equation relies only on the variance within the sample. If the three-way interaction is indeed linear, it should be as observable within this sample variance as within any larger population variance (unless the sample variance is negligible). However, an important side effect of range restriction is a change in distribution. Departure from normality is known to cause a loss of power for many statistical techniques, including multiple regression (e.g., Wilcox, 1998). For these reasons, we examine four different distributions: the normal distribution, a skewed normal distribution (calculated by the natural logarithm of the normal distribution), the uniform distribution, and a triangular distribution (calculated by the square root of a uniform distribution—in particular, this is not dissimilar from a truncated normal distribution, as may be caused by range restriction).

Operationalization of predictor and dependent variables. Aguinis (2004) highlighted three variable operationalizations that can affect the power of MMR: measurement error, scale coarseness, and polychotomization of truly continuous variables. Both Stone-Romero and Anderson (1994) and Aguinis and Stone-Romero (1997) found measurement error to be a highly important factor in the power of MMR, so we chose to vary this. We used reliabilities of 1.00, .80, and .60 (in both predictor and moderator variables) to represent values of perfect, acceptable, and barely acceptable reliability, respectively.

Sample size. Subgroup sample size is not an issue with continuous moderators, so we varied only the overall sample size, choosing values of 50, 100, 200, and 500 cases to represent sample sizes typically found in empirical studies.

Characteristics of the predictor variables. Research has shown that the intercorrelation between predictor variables has relatively little effect on the power of MMR (e.g., Aguinis, 1995). Rather, the relation between predictor and criterion is a far more important factor (Rogers, 2002). However, we are not interested so much in the power of finding a three-way interaction itself as in detecting the difference between a pair of slopes. Therefore, the correlations we are interested in varying are those of the relation between X and Y at particular values of Z and W. We chose differences between the correlations of the slopes of 0.50, 0.30, and 0.10 to tie in with Cohen's (1988) large, medium, and small effect sizes.[2] Additionally, we generated data with no underlying effect (i.e., zero correlation differences between the slopes) to test the level of Type I error of the test under the different conditions.

Consistent with this body of literature examining factors that reduce the power of MMR, we expect that restrictions in variable distribution, reliability, sample size, and magnitude of slope difference (which we refer to for the remainder of this article as *effect size*) will result in reduced test power. Additionally, we expect the four factors to interact such that test power will decline as a function of combinations of the four variables.

For each of the 108 ($4 \times 3 \times 3 \times 3$) conditions, we generated 10,000 data sets on the basis of populations with the specified effect size. We generated effects of a given size for one selected pair of slopes (slopes representing the relation between X and Y): for low values of W, with differences between high and low values of Z (Situations b and d described earlier; see Difference c, Figure 1). For each data set, we performed the test of slope differences associated with this difference (Test c). We used the percentage of occasions when a significant result (with $p < .05$) was detected (i.e., the test correctly rejected the null hypothesis of no difference

---

[1] A Microsoft Excel worksheet for conducting these slope difference tests is available online at www.jeremydawson.co.uk/slopes.htm

[2] Note that the size of the slope difference is not necessarily indicative of the traditional effect size (such as ΔR^2 or f^2; Aguinis, 2004), which refers to the three-way interaction term and does not represent the difference between only one pair of slopes. It is possible that a small value of f^2 (e.g., .002, the value revealed by Aguinis, Beaty, Boik, & Pierce, 2005, as the median effect size in 261 MMR studies) is associated with a large slope difference or vice versa.

Table 3
Percentage of Rejection of the Null Hypothesis as a Function of Size of Slope Difference, Reliability, Sample Size, and Distribution at $p < .05$

Distribution and sample size	Reliability of data											
	Large effect size			Medium effect size			Small effect size			Zero effect size		
	1.00	.80	.60	1.00	.80	.60	1.00	.80	.60	1.00	.80	.60
Normal												
n = 50	25.2	12.7	8.9	0.3	2.5	4.5	0.0	0.6	2.8	4.8	4.9	5.0
n = 100	98.5	39.0	16.7	35.6	10.3	7.5	0.0	2.0	3.7	4.7	4.7	5.1
n = 200	100.0	80.6	30.6	99.8	34.4	13.2	10.5	7.2	5.6	5.1	4.9	4.8
n = 500	100.0	99.9	66.1	100.0	89.9	33.7	100.0	37.7	12.8	5.0	5.1	4.8
Skewed normal												
n = 50	25.1	12.1	9.5	0.0	2.5	4.7	0.0	0.5	2.9	4.9	5.2	4.7
n = 100	99.9	40.0	17.1	27.0	10.4	7.6	0.0	1.8	4.3	4.9	5.3	5.1
n = 200	100.0	83.4	32.1	100.0	37.2	13.4	5.0	7.3	5.9	4.8	5.1	5.0
n = 500	100.0	99.9	69.2	100.0	92.9	34.5	100.0	39.7	13.4	5.3	4.7	4.9
Uniform												
n = 50	21.0	8.9	6.9	0.0	1.8	3.8	0.0	0.3	2.0	5.4	5.0	5.1
n = 100	99.8	34.6	13.2	22.3	8.4	6.5	0.0	1.3	3.1	5.0	5.2	5.1
n = 200	100.0	80.2	26.0	100.0	32.3	11.7	6.8	5.4	5.0	5.0	4.9	5.1
n = 500	100.0	99.9	62.7	100.0	91.9	31.8	100.0	37.7	12.4	4.4	5.0	4.8
Triangular												
n = 50	19.7	9.0	7.4	0.1	5.1	5.3	0.0	0.3	1.5	4.9	5.2	5.0
n = 100	100.0	37.9	16.1	22.5	9.4	6.9	0.0	1.3	2.5	4.6	5.1	4.8
n = 200	100.0	85.4	29.4	100.0	32.3	13.1	1.6	5.0	4.4	5.1	5.1	4.9
n = 500	100.0	100.0	66.6	100.0	93.5	29.5	100.0	38.7	12.3	5.0	5.2	5.0

between slopes) as an estimate of test power (e.g., Stone-Romero & Anderson, 1994).

Results and Discussion

Table 3 illustrates that test power varied as a function of the conditions, whereas the level of Type I error was consistently around the .05 level, as it should be. To test more accurately how the different conditions affected test power, we used generalized linear models[3] to assess the significance of each factor individually and interactions between factors. To increase the power of this analysis, we ran further Monte Carlo simulations, with sample sizes of 150, 250, 300, 350, 400, and 450; reliabilities of .90 and .70; and effect sizes of 0.40 and 0.20, giving a total of 1,000 cases. The results of these simulations are not reported separately here but are available on request from Jeremy F. Dawson.

Main effects. Four separate models showing the effect of each predictor variable on test power were used to examine main effects. Results revealed no significant effect of sample distribution on test power (deviance = 0.13, p = .98). Each of the other three predictor variables had a significant main effect: sample size (deviance = 193.57, $p < .001$), reliability (deviance = 184.03, $p < .001$), and effect size (deviance = 117.07, $p < .001$). When entered together into one model, all remained significant, which indicates that the three effects were independent of each other. As suggested by Table 3, the effects of all three variables were positive: The greater sample size, reliability, and effect size were, the higher the power was. To test for nonlinear effects, we added the squared term of each predictor to the model. The (sample size)^2 term was significant (deviance = 21.83, $p < .001$), indicating the presence of a nonlinear effect. In particular, the coefficient was negative, indicating that the positive effect of sample size decreased as sample size increased. The other effects were not significant, indicating that these main effects were (approximately) linear in nature.

Interaction effects. Each pair of predictors was tested for an interaction effect on the power of the slope test. Two significant interaction effects emerged: between sample size and reliability (deviance = 30.58, $p < .001$), and between effect size and reliability (deviance = 16.03, $p < .001$). The final model predicting test power was represented by the ensuing equation: Power = Φ [$-0.65 - 0.020 \times$ sample size + .000017 $\times$ (sample size)^2 $-$ 4.33 $\times$ reliability $-$ 2.75 $\times$ effect size + 0.039 $\times$ (sample size $\times$ reliability) $-$ 0.000029 $\times$ (sample size)^2 $\times$ reliability + 11.18 $\times$ reliability $\times$ effect size], where Φ represents the cumulative normal distribution function. Interpretation of the interaction effects is not straightforward because of this transformation, but the multiplicative effect of sample size and reliability suggests that the higher the sample size was, the less was the advantage of having more reliable data. However, the multiplicative effect of reliability and effect size suggests that the larger the underlying effect size was, the greater was the advantage of having more reliable data (until power got close to 1). As is often the case with interaction effects, the negative coefficients associated with the main effects are misleading: Without the interaction terms present, these were all positive. The model fitted the data well: The residual deviance was only 51.54 on 992 degrees of freedom, compared with 747.87 on 999 degrees of freedom for the null model.

[3] Generalized linear models were needed as the dependent variable (power) was a percentage (a probit link function was used).

Table 3 illustrates that to achieve power of at least 80% of detecting a large effect (80% being a commonly used figure for minimum acceptable power in experimental design), it is necessary to have either (a) perfectly reliable data and around 100 cases or (b) data with a reliability of .80 and 200 cases. Using the model equation above, we can show that either a situation of 125 cases and .90 reliability or a situation of 200 cases and .79 reliability would result in approximately 80% test power.

To detect a moderate effect with at least 80% power, Table 3 shows, it is necessary to have either (a) perfectly reliable data and approximately 200 cases or (b) data with a reliability of .80 and 500 cases. From the model equation, we can show that either a situation of 249 cases and .90 reliability or a situation of 500 cases and .72 reliability would yield about 80% test power. This points to the danger of failing to detect moderate slope differences due to small sample size. To detect a small effect size with power of at least 80%, Table 3 suggests, it is necessary to have both perfectly reliable data and 500 cases. The model equation above shows that, actually, .90 reliability and 410 cases would suffice.

An interesting feature of the results is where power dipped below 5%, suggesting that, in this case, it is more likely that a significant result was due to chance than an underlying effect. This is particularly the case for small sample sizes. Note also the apparent increase of power as reliability decreased in cases in which power was below 5%, which was due to the data becoming more random; conversely, where reliability was good, the (genuinely) small effect was too small to be picked up by the test, resulting in power well below 5%.

Application of the Slope Difference Test to Published Data

To test the utility of the slope difference test, we reanalyzed three published articles that used alternative ways to probe significant three-way interactions in MMR. We aim to illustrate that (a) the test provides the possibility to accurately test hypotheses that are not testable with alternative methods and (b) the use of the test may change results and conclusions relative to other probing methods. Table 4 summarizes key features of this reanalysis.

Study 1

The first article (Kwong & Leung, 2002) hypothesized slope differences and used subgroup analysis to probe the interaction. In a sample of 199 undergraduate students, Kwong and Leung (2002, Study 1) investigated the contingent effect of outcome favorableness and closeness of interpersonal relationships on the impact of interactional justice on happiness. The authors proposed that the tendency for interactional justice to have a stronger and more positive effect on happiness when outcome is unfavorable versus favorable should be

Table 4
Reanalysis of Data from Kwong and Leung (2002), van Yperen and Janssen (2002), and Smither and Walker (2004)

Study and reanalysis thereof	Hypothesis	Method used for probing significant three-way interaction	Results	Implications
Kwong & Leung (2002)	Two slope difference hypotheses	Subgroup analysis; significance test of two-way interactions per subgroup	Two-way interactions are significant for both subgroups; effect of both two-way interactions is reversed in subgroups	Hypotheses supported
Reanalysis of Kwong & Leung (2002)	Two slope difference hypotheses; further exploration of slope differences	Slope difference tests (a priori and post hoc)	The two pairs of slopes differ significantly as expected; additional revelation of two significant slope differences	Hypotheses supported; possible refinement of hypothesis implied
van Yperen & Janssen (2002)	Slope difference hypothesis (implied)	Pick-a-point approach (Aiken & West, 1991; Rogosa, 1980)	Only one of four simple slopes predicted significant variance in the outcome	Hypothesis supported
Reanalysis of van Yperen & Janssen (2002)	Slope difference hypothesis	Slope difference test (a priori)	No slope differences detected	Hypothesis rejected; the relation between job demands and job satisfaction varies across the range of both moderators
Smither & Walker (2004)	No hypothesis formulated	Subgroup analysis; use of effect size measure mean standardized difference *(d)*	Large difference between two cells; small to medium differences between others	The joint effect of unfavorable, behavior/task-focused comments on performance ratings varies depending on number of comments
Reanalysis of Smither & Walker (2004)	Exploration of slope differences	Slope difference test (post hoc)	No slopes differences detected	Performance ratings vary across the range of the three predictor variables

more pronounced when the relationship between the disputants is close rather than remote. In our terms, the authors proposed two slope difference hypotheses: First, in the prediction of effects of interactional justice on happiness, the slope of low favorableness and close relationships should be greater than the slope of low favorableness and remote relationships; second, the slope of high favorableness and close relationships should be less positive than the slope of high favorableness and remote relationships. To further examine the significant three-way interaction term, the authors conducted subgroup analyses by means of a median split on closeness and then ran separate two-way interactions for both subgroups of close and remote relationships. Both two-way interactions were significant, with t values showing reversed signs. The authors then plotted the interactions by subgroup and interpreted the plots. Using the slope difference test in its a priori form, we arrive at partially the same conclusion as the authors: Indeed, in the prediction of happiness by interactional justice, the slope for unfavorable outcomes and close relationships differed significantly from the slope for unfavorable outcomes and remote relationships, $t(191) = -3.64$, $p < .001$; similarly, the slopes for favorable outcomes differed significantly, $t(191) = -3.65$, $p < .001$. We then explored whether we could find additional slope differences by applying the slope difference test as a post hoc test (we used a Bonferroni-adjusted alpha value of .0125 for a significance level of .05, as we made a total of four exploratory comparisons). Two of the four remaining slope comparisons were significant; so did, for example, the slope for high favorableness and close relationships differ from the slope for low favorableness and remote relationships, $t(191) = -3.40$, $p = .001$. Because of the use of subgroup analysis, Kwong and Leung were unable to reveal such slope differences across subgroups. Even though it was not hypothesized, this information would have further clarified the nature of the three-way interaction and might have led to a refinement of the hypothesis in the authors' replication study (Study 2). Thus, even though we arrived at the same conclusion as Kwong and Leung regarding the support for the hypothesis, use of the slope difference test revealed additional slope differences that were undetected by the authors.

Study 2

The second article (van Yperen & Janssen, 2002) implies a slope difference hypothesis, and the authors used the pick-a-point approach to probe the interaction. They investigated the contingent effect of both performance orientation and mastery orientation on the relation between job demands and job satisfaction, using a sample of 322 university employees. The authors hypothesized that job demands and job satisfaction are negatively related only when an employee's performance orientation is stronger and his or her mastery orientation is weaker. As a test of their hypothesis, the authors examined whether simple slopes were significant predictors of the dependent variable, irrespective of differences among slopes. Because only one of the four slopes, the slope for high performance orientation and low mastery orientation, reached statistical significance, the authors concluded that their hypothesis was supported. We argue that the hypothesis—that the relation between job demands and satisfaction is negative only when an employee's performance orientation is stronger and his or her mastery orientation is weaker—implies the use of a relational test of slope differences. The hypothesis would receive stronger support if one could show that the slope for high performance orientation and weak mastery orientation is more strongly negative than any of the other slopes. Application of the slope difference test to the data in its a priori form revealed that this slope did not differ significantly from any other slope. There was nearly a significant difference between the slopes for strong mastery orientation and weak performance orientation versus strong mastery orientation and strong performance orientation, $t(284) = 1.91$, $p = .06$; however, the lack of difference between the slopes for strong mastery orientation and weak performance orientation versus weak mastery orientation and weak performance orientation, $t(284) = -0.56$, $p = .58$, means that the level of mastery orientation when performance orientation was weak was not shown to be important. This result suggests that there is little support for the authors' hypothesis. In sum, use of the slope difference test yielded different conclusions compared with use of the pick-a-point approach.

Study 3

Smither and Walker (2004) did not formulate a hypothesis for the found significant three-way interaction and probed the interaction in an exploratory manner. The authors investigated the combined effect of behavior- and task-focused feedback, number of comments, and favorableness of feedback comments on improvement in the behavior of feedback recipients (as reflected in subsequent ratings), using a sample of 176 managers. The authors further examined a significant three-way interaction by conducting a median split on each of the three predictors and then comparing the mean change scores for the resulting eight cells by using the effect size measure standardized mean difference. The authors then compared the four cells with low numbers of comments with the four cells with large numbers of comments. Comparisons revealed one difference with a large effect size, whereas the other three differences resulted in small to medium effect sizes. Smither and Walker (2004) interpreted the results by stating,

> When managers received a small number of unfavorable, behavior/task-focused comments, their subsequent performance improved more than that of managers *in other conditions* [italics added]. However, when managers received a large number of unfavorable, behavior/task-focused comments, their subsequent performance declined more than that of managers *in other conditions* [italics added]. (p. 578)

The conclusions the authors drew, however, would have required a test that compared effect size among all possible combinations of predictor variables or, alternatively, a slope difference test. We applied the slope difference test to the data as a post hoc test, as no hypothesis was formulated (a total of six comparisons yielded a Bonferroni-adjusted alpha of .0083 for a significance level of .05). We did not find any significant differences among slopes, which suggests the conclusion that performance ratings varied across the range of the three predictor variables. The one large effect size difference reported by the authors, however, showed the strongest slope difference, $t(144) = 1.95$, $p = .052$. In sum, application of the slope difference test suggests that a large effect size revealed with effect size measures was actually nonsignificant—and would not be significant even with a nonadjusted alpha value.

Discussion

Summary of Key Results

In the course of this article, we first developed a test for slope differences for three-way interaction effects, building on the work of Aiken and West (1991). We proceeded to show how effect size of slope differences, sample size, reliability, and data distribution affect the power of this test and illustrated that test power is especially restricted when samples are small and scales unreliable. Finally, we demonstrated, by reanalyzing published data, that the slope difference test can provide an accurate and useful tool that allows researchers to test and explore slope differences that may remain undetected by use of alternative probing techniques. Use of the test can even result in changes to results and conclusions.

Applicability and Usefulness of the Test

Monte Carlo simulations and analysis of the results from these showed that detection of moderate or large slope differences should be possible with data conditions that are not at all unusual for researchers who wish to test for three-way interactions: A combination of highly reliable data and around 250 cases or marginally reliable data and around 500 cases yields about 80% power for a slope (correlation) difference of 0.30. To detect smaller effect sizes, such as a slope difference of only 0.10, one needs more stringent conditions, such as highly reliable data and over 400 cases. Even these, however, are not beyond the realm of possibility, and many researchers achieve such conditions. Of course, some researchers may be interested in detecting even smaller slope differences, and, in these cases, yet more stringent conditions are needed. However, slope differences smaller than this may be of little practical significance, and even when they are of a meaningful nature, large sample sizes and highly reliable variables are needed to detect the three-way interaction at all.

A strong advantage of this test is that it is applicable whatever the form of the interaction. Because the test is based on the calculated difference between a pair of slopes, it does not matter whether both slopes are positive, both are negative, or there is a crossover effect: Only the magnitude of the difference is important. Additional Monte Carlo simulations confirmed that such a difference in form did not affect the power of the test (details are available from Jeremy F. Dawson on request). We stress that, irrespective of statistical conditions, a strong theoretical basis for proposing slope differences, in combination with attempts to replicate the findings, is strongly advised to make this test a powerful tool.

Limitations and Future Perspectives

It is important to note that a researcher may well have a situation in which a three-way interaction is significant but significant slope differences are not detected. We offer two potential explanations for this: First, one or more of the three study characteristics described in this article has restricted test power. Second, the significance of the three-way interaction is a function of a complex interaction of more than one pair of slopes rather than a single pair of slopes (see the reanalysis of the Smither & Walker, 2004 data). Given either of these cases, we suggest that the researcher revert to explaining those three-way interactions by restricting the interpretation that the effect of a variable *X* on a dependent variable *Y* varies as a function of the combined effect of the two moderators *W* and *Z*. Additional research would be necessary how to further probe these interactions.

Finally, although we studied the effects of four key factors on the power of the test, in the future, researchers should examine several other factors that may affect test power (see Aguinis, 1995, 2004, for reviews). Also, the test could be extended to probe the nature of three-way interaction effects including categorical moderator variables.

References

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

Aguinis, H. (2004). *Regression analysis for categorical moderators.* New York: Guilford Press.

Aguinis, H., Beaty, 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.

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.

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

Baker, D. D., & Cullen, J. B. (1993). Administrative reorganization and configurational context: The contingent effect of age, size, and change in size. *Academy of Management Journal, 36,* 1251–1277.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40,* 373–400.

Brown, S. P., Ganesan, S., & Challagalla, G. (2001). Self-efficacy as a moderator of information-seeking effectiveness. *Journal of Applied Psychology, 86,* 1043–1053.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: 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: Erlbaum.

Dickson, P. H., & Weaver, K. M. (1997). Environmental determinants and individual-level moderators of alliance use. *Academy of Management Journal, 40,* 404–425.

Erez, M., & Earley, P. C. (1987). Comparative analysis of goal-setting strategies across cultures. *Journal of Applied Psychology, 72,* 658–665.

Insightful Corporation. (2001). S-plus [Computer software]. Seattle, WA: Author.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69,* 307–321.

Janssen, O., van de Vliert, E., & Veenstra, C. (1999). How task and person conflict shape the role of positive interdependence in management groups. *Journal of Management, 25,* 117–142.

Kwong, J. Y. Y., & Leung, K. (2002). A moderator of the interaction effect of procedural justice and outcome favorability: Importance of the relationship. *Organizational Behavior and Human Decision Processes, 87,* 278–299.

Miller, R. G. (1981). *Simultaneous statistical inference* (2nd ed.). Berlin: Springer-Verlag.

Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal, 34,* 607–634.

Oldham, G. R., & Fried, Y. (1987). Employee reactions to workspace characteristics. *Journal of Applied Psychology, 72,* 75–80.

Peters, L. H., O'Connor, E., & Wise, S. L. (1984). The specification and testing of useful moderator variable hypotheses. In T. S. Bateman & G. R. Forris (Eds.), *Method and analysis in organizational research* (pp. 128–139). Reston, VA: Reston.

Rogers, W. M. (2002). Theoretical and mathematical constraints of interaction regression models. *Organizational Research Methods, 5,* 212–230.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88,* 307–321.

Schaubroek, J., Ganster, D. C., & Kemmerer, B. E. (1994). Job complexity, "type A" behavior, and cardiovascular disorder: A prospective study. *Academy of Management Journal, 37,* 426–439.

Schaubroek, J., & Merritt, D. E. (1997). Divergent effects of job control on coping with work stressors: The key role of self-efficacy. *Academy of Management Journal, 40,* 738–754.

Smither, J. W., & Walker, A. G. (2004). Are the characteristics of narrative comments related to improvement in multirater feedback ratings over time? *Journal of Applied Psychology, 89,* 575–581.

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

van Yperen, N. W., & Janssen, O. (2002). Fatigued and dissatisfied or fatigued but satisfied? Goal orientations and responses to high job demands. *Academy of Management Journal, 45,* 1161–1171.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53,* 300–314.

Xie, J. L. (1996). Karasek's model in the People's Republic of China: Effects of job demands, control, and individual differences. *Academy of Management Journal, 39,* 1594–1618.

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

Appendix

Derivation of Standard Errors of Slope Differences

To determine whether a difference in slopes is significant, it is necessary to compare the difference with its standard error. As the standard error of an estimate is the square root of its variance, to determine the standard error of these slope differences, we need to know the variances and covariances of the regression estimates b_4, b_5, and b_7. (Note that as b_0, b_2, b_3, and b_6 only form part of the intercept section of the equation and b_1 cancels out when we calculate slope differences, it is not necessary to consider these parameters here.)

For Slope Difference a, we can calculate the standard error—using the statistical identity $\text{var}(aX + bY) = a^2\,\text{var}(X) + b^2\,\text{var}(Y) + 2ab\,\text{cov}(X, Y)$—for any two variables X and Y and constants a and b as

$$SE_{\text{a}} = \sqrt{\text{var}[b_5(W_{\text{H}} - W_{\text{L}}) + b_7 Z_{\text{H}}(W_{\text{H}} - W_{\text{L}})]}$$

$$= \sqrt{\begin{array}{l}(W_{\text{H}} - W_{\text{L}})^2\text{var}(b_5) + (W_{\text{H}} - W_{\text{L}})^2 Z_{\text{H}}^2\text{var}(b_7)\\ + 2(W_{\text{H}} - W_{\text{L}})^2 Z_{\text{H}}\,\text{cov}(b_5,b_7)\end{array}}$$

$$= (W_{\text{H}} - W_{\text{L}})\sqrt{\text{var}(b_5) + Z_{\text{H}}^2\text{var}(b_7) + 2Z_{\text{H}}\text{cov}(b_5,b_7)}\,.$$

If we use the notation proposed by Aiken and West (1991), where the variance–covariance matrix of regression parameter estimates is given by

$$S = \begin{pmatrix} s_{00} & \cdots & s_{07} \\ \vdots & \ddots & \vdots \\ s_{70} & \cdots & s_{77} \end{pmatrix}$$

—where, for example, s_{11} is the variance of parameter estimate b_1 (the coefficient of X) and s_{45} is the covariance of parameter estimates b_4 and b_5—this gives us the following formula for the standard error of Slope Difference a:

$$SE_{\text{a}} = (W_{\text{H}} - W_{\text{L}})\sqrt{s_{55} + Z_{\text{H}}^2 s_{77} + 2Z_{\text{H}} s_{57}}\,.$$

In a similar way, the standard errors of Slope Differences b, c, and d can be calculated by

$$SE_{\text{b}} = (Z_{\text{H}} - Z_{\text{L}})\sqrt{s_{44} + W_{\text{H}}^2 s_{77} + 2W_{\text{H}} s_{47}}\,,$$

$$SE_{\text{c}} = (Z_{\text{H}} - Z_{\text{L}})\sqrt{s_{44} + W_{\text{L}}^2 s_{77} + 2W_{\text{L}} s_{47}}\,,$$

and

$$SE_{\text{d}} = (W_{\text{H}} - W_{\text{L}})\sqrt{s_{55} + Z_{\text{L}}^2 s_{77} + 2Z_{\text{L}} s_{57}}\,.$$

Note that if variables Z and W had been standardized prior to the analysis, these (simplified) formulas would be

$$SE_{\text{a}} = 2\sqrt{s_{55} + s_{77} + 2s_{57}}\,,$$

$$SE_{\text{b}} = 2\sqrt{s_{44} + s_{77} + 2s_{47}}\,,$$

$$SE_{\text{c}} = 2\sqrt{s_{44} + s_{77} - 2s_{47}}\,,$$

and

$$SE_{\text{d}} = 2\sqrt{s_{55} + s_{77} - 2s_{57}}\,.$$

In the case of Slope Differences e and f, the formula is more complicated. Using the same principle, we get

$$SE_{\text{e}} = \sqrt{\text{var}[b_4(Z_{\text{H}} - Z_{\text{L}}) + b_5(W_{\text{H}} - W_{\text{L}}) + b_7(Z_{\text{H}}W_{\text{H}} - Z_{\text{L}}W_{\text{L}})]}$$

$$= \sqrt{\begin{array}{l}(Z_{\text{H}} - Z_{\text{L}})^2 s_{44} + (W_{\text{H}} - W_{\text{L}})^2 s_{55} + (Z_{\text{H}}W_{\text{H}} - Z_{\text{L}}W_{\text{L}})^2 s_{77}\\ +2[(Z_{\text{H}} - Z_{\text{L}})(W_{\text{H}} - W_{\text{L}})s_{45} + (Z_{\text{H}} - Z_{\text{L}})(Z_{\text{H}}W_{\text{H}} - Z_{\text{L}}W_{\text{L}})s_{47}\\ + (W_{\text{H}} - W_{\text{L}})(Z_{\text{H}}W_{\text{H}} - Z_{\text{L}}W_{\text{L}})s_{57}]\end{array}}\,.$$

and, similarly,

$$SE_{\text{f}} = \sqrt{\text{var}[b_4(Z_{\text{H}} - Z_{\text{L}}) + b_5(W_{\text{L}} - W_{\text{H}}) + b_7(Z_{\text{H}}W_{\text{L}} - Z_{\text{L}}W_{\text{H}})]}$$

$$= \sqrt{\begin{array}{l}(Z_{\text{H}} - Z_{\text{L}})^2 s_{44} + (W_{\text{L}} - W_{\text{H}})^2 s_{55} + (Z_{\text{H}}W_{\text{L}} - Z_{\text{L}}W_{\text{H}})^2 s_{77}\\ +2[(Z_{\text{H}} - Z_{\text{L}})(W_{\text{L}} - W_{\text{H}})s_{45} + (Z_{\text{H}} - Z_{\text{L}})(Z_{\text{H}}W_{\text{L}} - Z_{\text{L}}W_{\text{H}})s_{47}\\ + (W_{\text{L}} - W_{\text{H}})(Z_{\text{H}}W_{\text{L}} - Z_{\text{L}}W_{\text{H}})s_{57}]\end{array}}\,.$$

Again, note that for standardized Z and W, these become

$$SE_{\text{e}} = 2\sqrt{s_{44} + s_{55} + 2s_{45}}$$

and

$$SE_{\text{f}} = 2\sqrt{s_{44} + s_{55} - 2s_{45}}\,,$$

which makes standardization of the variables highly desirable!

Received October 12, 2004
Revision received February 11, 2005
Accepted March 14, 2005 ■