Journal of Organizational Behavior
J. Organiz. Behav. **31**, 776–786 (2010)
Published online 30 April 2010 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/job.686

Researcher's Notebook

Best-practice recommendations for estimating interaction effects using moderated multiple regression

HERMAN AGUINIS* AND RYAN K. GOTTFREDSON

Department of Management & Entrepreneurship, Kelley School of Business, Indiana University, Bloomington, Indiana, U.S.A.

Summary An interaction effect indicates that a relationship is contingent upon the values of another (moderator) variable. Thus, interaction effects describe conditions under which relationships change in strength and/or direction. Understanding interaction effects is essential for the advancement of the organizational sciences because they highlight a theory's boundary conditions. We describe procedures for estimating and interpreting interaction effects using moderated multiple regression (MMR). We distill the technical literature for a general readership of organizational science researchers and include specific best-practice recommendations regarding actions researchers can take before and after data collection to improve the accuracy of MMR-based conclusions regarding interaction effects. Copyright © 2010 John Wiley & Sons, Ltd.

Introduction

Interaction, also labeled moderating, effects provide information on whether the relationship between two variables is contingent upon the value of a third variable. In other words, an interaction effect hypothesis states that the relationship between two variables, or the effect of one variable on a second one, depends on the value of a third (moderator) variable.

It is no exaggeration to assert that virtually every major theory in the organizational sciences includes interaction effects (Aguinis, Beaty, Boik, & Pierce, 2005). This is because most theories, particularly those which are more mature, include a consideration of their boundary conditions. So, when we state that the direction or strength of a relationship (i.e., first-order relationship) depends upon or is contingent upon other factors, these other factors are labeled *moderator variables* because the first-order relationship is moderated (or, in general terms, it changes) as the moderator variable changes. Consider two illustrations that have made it into the pages of most textbooks in organizational behavior. The first one is about leadership theory. For decades, leadership research focused on leader characteristics as antecedents of outcomes such as group performance. However, Fiedler (1967) discovered that a more fruitful approach was to assess interaction effects between leadership style and

* Correspondence to: Herman Aguinis, Department of Management & Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Bloomington, IN 47405-1701, U.S.A. E-mail: haguinis@indiana.edu

Received 20 January 2010
Accepted 23 January 2010

situational factors. The work by Fiedler and others led to the conclusion that situational factors serve as moderators of the effects of leadership style on group performance. A second illustration is in the area of work design. Similar to leadership theory, for many years researchers had focused on first-order relationships only; specifically, most research efforts were directed at trying to understand the direct effects of job characteristics (e.g., job autonomy) on work outcomes such as job satisfaction. However, Oldham, Hackman, and Pearce (1976) and others, produced a conceptual improvement and found evidence for an interaction effect: The relationship between core job dimensions and work outcomes is moderated by the degree of an individual's growth-need strength. The presence of this interaction effect led to important conceptual as well as practical implications. For example, implementing job enrichment interventions will not lead to uniformly positive outcomes: Positive outcomes will be observed mainly for individuals who are high on growth-need strength.

Given the pervasiveness, prominence, and central role of interaction effects in the organizational sciences, the goal of this article is to provide an overview, together with best-practice recommendations, regarding how to test hypotheses, assess the possible presence and magnitude, interpret, and report interaction effects. Note, however, that researchers may wish to assess possible interaction effects in various contexts and implement various types of designs resulting in various types of data structures such as involving one or more indicators per latent construct, multi-level data structures, and quantitative literature reviews (i.e., meta-analysis). However, this article focuses only on estimating interaction effect utilizing one particularly common approach only: Moderated multiple regression (MMR). There is an extensive body of technical literature regarding the use of MMR. Much of this literature is quite specialized and includes analytic work that is mathematically sophisticated as well as Monte Carlo simulations involving lengthy and complex procedures and results. Due to the nature of this research, much of this work is not easily accessible to researchers with the usual methodological and statistical background resulting from doctoral-level training in the organizational sciences. Accordingly, this article distills the technical literature for a general readership and includes specific recommendations that researchers will be able to use in their quest for interaction effects.

Estimating Interaction Effects Using Multiple Regression: Moderated Multiple Regression

Moderated multiple regression (MMR) seems to be the statistical tool of choice for estimating interaction effects in the organizational sciences. For example, Aguinis et al. (2005) reviewed all issues of *Journal of Applied Psychology* (JAP), *Personnel Psychology* (PPsych), and *Academy of Management Journal* (AMJ) from 1969 to 1998 and identified all articles reporting a categorical moderator test using MMR. Although the review included only three journals and the review focused on one type of moderator variable only (i.e., categorical moderators and not continuous moderators), results showed that the total number of reported MMR analyses was 636. This review found that the first article using MMR to assess effects of categorical moderator variables in these three journals was published in 1977 and there is an upward trend in the use of MMR overtime. Overall, the frequency of MMR to estimate categorical moderators in JAP, PPsych, and AMJ has remained at a level of approximately 20–40 analyses per year since the mid-1980s. Given that the general multiple regression approach has remained one of the top five most popular data-analysis techniques over the past three decades (Aguinis, Pierce, Bosco, & Muslin, 2009), it is not surprising that MMR is one of the most popular, if not the most popular, approach for testing hypothesis about interaction effects.

DOI: 10.1002/job

Estimating interaction effects using MMR usually consists of creating two ordinary least squares (OLS) regression equations involving scores for a continuous predictor variable Y, scores for a predictor variable X, and scores for a second predictor variable Z hypothesized to be a moderator. Stated differently, there is an interest in investigating whether there is an interaction effect between x and z observed scores in predicting or causing observed criterion scores y. Equation (1) shows the ordinary least squares (OLS) regression equation for a model predicting y scores from the first-order effects of x and z observed scores:

$$y = a + b_1 x + b_2 z + e \tag{1}$$

where a is the least-squares estimate of the intercept, b_1 is the least-squares estimate of the population regression coefficient for x observed scores, b_2 is the least-squares estimate of the population regression coefficient for z observed scores, and e is an estimated residual. Note that the observed criterion scores y are quantitative; other procedures such as logistic regression can be used in situations in which y scores are categorical (Jaccard, 2001). The moderator scores z can be continuous or binary (i.e., categorical moderator with two levels). MMR can accommodate categorical moderators with more than two levels as well (see Aguinis, 2004, Chapter 8, for a more detailed description).

The multiple regression model shown in Equation (1) assumes that all variables identified by the theory are included in the model and that the variables are measured reliably (i.e., without error). In addition, it is assumed that the population data have the following characteristics: (a) The relationship between each of the predictors and the criterion is linear, (b) residuals (i.e., difference between predicted and actual y scores) exhibit homoscedasticity (i.e., constant variance across values of each predictor; that is, residuals are evenly distributed throughout the regression line), (c) residuals are independent (i.e., there is no relationship among residuals for any subset of cases in the sample), (d) residuals are normally distributed, and (e) there is less than complete multicollinearity (i.e., less than perfect correlation between the predictors). These are the usual assumptions of all OLS multiple regression models and are described in detail in regression textbooks (e.g., Aguinis, 2004; Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003).

The second equation, called the MMR model, is formed by creating a new set of scores, the product of the observed scores for the two predictors (i.e., $x \cdot z$), and including it as a third term in the equation. The addition of the product term yields the following model:

$$y = a + b_1 x + b_2 z + b_3 x \cdot z + e \tag{2}$$

where b_3 is the least squares estimate of the population regression coefficient for the product term scores.

To formally test for the presence of a hypothesized moderating effect (i.e., interaction between x and z scores in predicting y scores), a t-statistic can be computed to test the null hypothesis H_0: $\beta_3 = 0$. The term β_3 is used to symbolize the regression coefficient for the product term in the population (i.e., b_3's parameter). Conceptually, if Z is a continuous variable, a test of this null hypothesis indicates whether the amount of change in the slope of the regression of y on x scores that results from a unit-change in z scores is greater than would be expected by chance alone. Note that interaction effects are symmetrical because, in Equation (2), $x \cdot z = z \cdot x$. So, a rejection of the null hypothesis H_0: $\beta_3 = 0$ can also be interpreted as indicating that the amount of change in the slope of the regression of y on z scores that results from a unit-change in x is greater than would be expected by chance alone. The former interpretation treats Z as the moderator variable, whereas the latter treats X as the moderator variable. Choosing to label one or the other variable as "predictor" or "moderator" is mathematically equivalent and, thus, it is based on conceptual considerations. In other words, from a theory perspective, does it make more sense and is there an interest in understanding whether the relationship

DOI: 10.1002/job

between X and Y changes as Z changes, or whether the relationship between Z and Y changes as X changes?

Alternatively, and equivalently, the coefficients of determination (i.e., sample-based squared multiple correlation coefficients, R^2) are compared for Equations (1) (i.e., R_1^2) and (2) (i.e., R_2^2). The null hypothesis tested is H$_0$: $\psi_2^2 - \psi_1^2 = 0$, where ψ_2^2 and ψ_1^2 are the population parameters for R_2^2 and R_1^2, respectively. Conceptually, this null hypothesis tests whether the addition of the product term to the regression equation improves the proportion of explained variance in y scores above and beyond the proportion of variance explained by x and z scores alone. In other words, this hypothesis answers the question of whether the addition of the moderating effect in the model improves the prediction of variable Y above and beyond the first-order effects of variables X and Z. To test H$_0$: $\psi_2^2 - \psi_1^2 = 0$, an F statistic (distributed with $k_2 - k_1$ and $N - k_2 - 1$ degrees of freedom) is computed using Equation (3)

$$F = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1 - R_2^2)/(N - k_2 - 1)} \tag{3}$$

where k_2 is the number of predictors in Equation (2), k_1 is the number of predictors in Equation (1), and N is the total sample size. Note that the statistical significance (i.e., p value) associated with the t statistic for the null hypothesis H$_0$: $\beta_3 = 0$ and the F statistic for the null hypothesis H$_0$: $\psi_2^2 - \psi_1^2 = 0$ are identical.

In terms of implementing an MMR analysis using commonly available software packages, the first step consists of creating a new variable, which is the product term between X and Z. Most packages' output include F and t statistics, and the associated statistical significance level (i.e., p), which allows researchers to answer the question of whether there is empirical support for the presence of an interaction effect.

Although testing hypotheses about interaction effects using MMR is a fairly straightforward procedure, results of such analysis are suspect in many, if not most, research studies. The reason is that MMR suffers from low statistical power (Aguinis, Boik, & Pierce, 2001; Aguinis, Culpepper, & Pierce, in press). Low statistical power means that, when they exist, the probability that population effects will be detected in the sample is low. In other words, researchers may conclude that their hypothesized interaction effect does not exist, whereas in fact it does. That is, researchers are likely to make a Type II error in their inferences regarding population interaction effects and fail to discover an existing interaction effect.

Best-Practice Design, Measurement, Analysis, and Reporting Recommendations for Estimating Moderating Effects Using MMR

Due to the typical low statistical power of MMR, it is necessary to consider design, measurement, and analysis issues to improve the chances that existing moderators will be detected if they exist. If there is a sole focus on data-analysis issues, it is likely that the low power problem will not be overcome and researchers are likely to conclude that there is no interaction effect, although the effect may actually be quite strong in the population (Aguinis et al., 2001). By the time the data have been collected, not having paid attention to design and measurement issues may mean that the existing moderating effect will not be detected, even if we conduct the analysis following best practices.

Regarding design issues, statistical power is enhanced when the variance of predictor variables is not negatively biased. In many situations in organizational science research, the variance of scores in the sample under study is negatively biased usually due to selection. For example, if a study investigates

DOI: 10.1002/job

work motivation as a predictor of job performance, it may be that only those individuals more highly motivated participated in the study (i.e., there was a positive motivation spill-over effect from work to completing the study's questionnaire). The power of MMR is reduced markedly when the observed variance in *X* is negatively biased due to selection (Aguinis & Stone-Romero, 1997). Aguinis and Stone-Romero's (1997) Monte Carlo study revealed that a surprisingly mild ratio of sample to population variance can have substantial detrimental effect on power. For example, in a situation with a total sample size of 300 and no truncation on *X* scores, the power to detect what Aguinis and Stone-Romero operationalized as a medium-size moderating effect was .81. However, when the scores were sampled from only the top 80% of the distribution of the population scores, power decreased to an unacceptable .51. Thus, even in the presence of a relatively mild degree of selection (i.e., the bottom 20% of the distribution is truncated), power loss poses a serious threat to the validity of MMR-based conclusions.

Two other important design-related considerations are related to the size of the sample included in the study. When the moderator is a continuous variable, total sample size is the focal issue. The smaller the sample size, the lower the power (other things equal). When the moderator is a categorical variable (e.g., gender, ethnicity), the focal issue is not only total sample size but also the subgroup sample sizes (e.g., number of men vs. number of women, number of ethnic minority vs. ethnic minority individuals). Statistical power is enhanced when total sample size increases and when the subgroup proportions approach .50 (i.e., similar number of individuals in each of the moderator-based subgroups).

Regarding measurement issues, statistical power is enhanced as we improve the reliability (i.e., decrease measurement error) of the criteria or dependent variables. In organizational science research, constructs such as job performance, job satisfaction, trust, leadership effectiveness, and organizational commitment are never measured with perfect reliability. In addition, measurement error in the predictors tends to bias the regression coefficients. In other words, in situations of less than perfect reliability in the predictor *X* and moderator *Z*, the reliability of $x \cdot z$ scores is affected adversely, and the sample-based coefficient b_3 may underestimate the population coefficient β_3 (*cf.* Equation (2)). Busemeyer and Jones (1983) provided the following expression which estimates the reliability for the product term based on the reliabilities of the predictor *X* and moderator *Z* variables when they are both standardized:

$$\rho_{xz,xz} = \frac{\rho_{xz}^2 + \rho_{xx}\rho_{zz}}{\rho_{xz}^2 + 1} \tag{4}$$

Equation (4) indicates that when the predictor *X* and the moderator *Z* are uncorrelated (i.e., $\rho_{xz}^2 = 0$), the reliability of the product term is reduced to the product of the reliabilities of the predictors. In other words, if the reliability of *X* is .70 and the reliability of *Z* is also .70, the resulting reliability of the product term is only .49! Not many organizational science researchers would find it acceptable to collect data knowing that 50% of the variance in scores is random error and is completely unrelated to the underlying construct of interest. However, unbeknownst to the researcher, this situation is not infrequent for the product term scores when using MMR. When *Y*, *X*, or *Z* are single-rater measures, reliability can be improved by using multiple raters and averaging their scores. In addition, reliability can also be improved by increasing the number of items (assuming the new items measure the same underlying construct as the existing items). These are two ways to improve reliability in general that also apply to the specific context of MMR.

A second important measurement issue to consider is scale coarseness. Statistical power will be enhanced to the extent that measurement scales are less coarse (Aguinis, Pierce, & Culpepper, 2009). A measurement scale is coarse when a construct that is continuous in nature is measured using items such that different true scores are collapsed into the same category. This practice leads to the introduction of

DOI: 10.1002/job

errors because continuous constructs are collapsed. Although this fact is seldom acknowledged, organizational science researchers use coarse scales every time continuous constructs are measured using Likert-type scales (Aguinis, Pierce, & Culpepper, 2009). As noted eloquently by Russell and Bobko (1992), ''summing responses to multiple Likert-type items on a dependent scale (as is often done in between-subjects survey designs) is not the same as providing subjects with a continuous response scale... if an individual responds to coarse Likert scales in a similar manner across items, the problem of reduced power to detect interaction remains... information loss that causes systematic error to occur at the item level would have the same effect on moderated regression effect size regardless of whether the dependent-response items were analyzed separately (as was done here in a within-subject design) or cumulated into a scale score'' (p. 339). Thus, using scales that are as continuous as possible when measuring underlying continuous constructs will enhance statistical power.

Regarding analysis issues, statistical power will be enhanced if predictor variables are not artificially dichotomized via a median split or similar procedures leading to ''high'' versus ''low'' subgroups. The issue of artificial dichotomization is an issue related to data analysis and is different from the issue of scale coarseness, which is a design problem. Specific to the context of MMR, Stone-Romero and Anderson (1994) demonstrated that artificial dichotomization of a quantitative predictor leads to substantial power loss in tests of moderators.

A second important issue in terms of analysis refers to the violation of the (within-group) homogeneity of error variance assumption in the context of estimating interaction effects of categorical moderators (Aguinis, Petersen, & Pierce, 1999; Aguinis & Pierce, 1998). Within-group homogeneity of variance means that the variance in Y that remains after predicting Y from X is equal across moderator-based subgroups (i.e., values of Z). In other words, the predicted scores for Y should be similarly distributed about the regression line for each of the moderator-based subpopulations. Violating the assumption can affect MMR results and conclusions substantially. Alexander and DeShon (1994) conducted a Monte Carlo study to investigate the impact of violating the homogeneity of error assumption on statistical power. They found that when sample sizes are unequal across subgroups and the subgroup with the larger sample size is associated with the larger error variance, violating the assumption causes statistical power to decrease substantially. For example, in the case with two moderator-based subgroups and subgroup ns (rs) of 20 (.20) and 40 (.50), the statistical power of the MMR test was 1.00 (i.e., MMR results based on each sample resulted in the correct conclusion that there is a moderator in every case). Yet, when the subgroup with the larger sample size was paired with the smaller correlation (i.e., when ns [rs] were 20 [.50] and 40 [.20]), power dropped to approximately .79. In short, making sure that the homogeneity of error variance assumption is not violated enhances statistical power and, overall, the validity of MMR-based conclusions.

Table 1 includes a summary of 12 best-practice recommendations for estimating interaction effects using MMR. The first nine recommendations summarize each of the design, measurement, and analysis issues we have discussed thus far. The last three recommendations included in Table 1 refer to the interpretation and reporting of interaction effects. Recommendation #10 refers to centering or standardizing predictor scores. A situation that would warrant centering or standardizing predictor scores is that there may be an interest in interpreting the coefficients b_1 and b_2 (and the intercept a) in Equation (2) when there is a non-zero interaction effect. For example, assume that $b_2 = 2.0$ and a continuous moderator Z. The interpretation of this regression coefficient is that Y scores are predicted to increase by 2 points when Z scores increase by 1 point at a value of $X = 0$. However, if X was measured using a scale that does not include a true zero point meaning the complete absence of an attribute (e.g., Likert-type scale ranging from 1 to 7), then the interpretation of b_2 becomes meaningless because a value of $x = 0$ may not exist. For example, if we are predicting job performance (Y) based on general cognitive abilities (X) and conscientiousness (Z), then b_2 is the predicted change for job performance based on conscientiousness for an individual with a score of precisely zero on general cognitive abilities, which is

DOI: 10.1002/job

Table 1. Summary of best-practice recommendations for estimating and interpreting interaction effects using moderated multiple regression (MMR)

Pre data collection recommendations
(1) Have a good rationale for why the interaction effect should exist
(2) Understand needed design features (e.g., total sample size, reliability, moderator-based subgroup sample sizes for categorical moderators) using computer programs available in the public domain (Aguinis, 2004 for categorical moderators; Shieh, 2009 for continuous moderators)
(3) Make design choices considering practical constraints but also based on the statistical power analysis results
(4) Minimize scale coarseness (Aguinis, Pierce, & Culpepper, 2009)
(5) Use the most reliable measurement instruments available
(6) Minimize truncation of predictor scores
Post data collection recommendations
(7) Do not dichotomize or polichotomize continuous variables
(8) If estimating effects of categorical moderators, check for compliance with (within-group) homogeneity of error variance assumption using computer programs in the public domain (Aguinis, 2004; http://mypage.iu.edu/~haguinis/)
(9) If estimating effects of categorical moderators and the homogeneity of error variance assumption has been violated, use alternatives to MMR using computer programs in the public domain (Aguinis, 2004; http://mypage.iu.edu/~haguinis/)
(10) Standardize or mean-center predictor scores if there is an interest in interpreting first-order effects in the presence of an interaction effect
(11) Create graphs to illustrate the nature of the interaction effect (O'Connor, 1998; https://people.ok.ubc.ca/brioconn/simple/simple.html, Preacher et al., 2006; http://people.ku.edu/~preacher/interact/mlr2.htms)
(12) Estimate effect size and practical significance

an impossibility. Mean-centering achieves the goal of making the interpretation of the first-order coefficients meaningful by the process of re-scaling. If Equation (2) is computed after mean-centering the quantitative predictors, then the interpretation of $b_2 = 2.0$ is as follows: Y scores are predicted to increase 2 points when Z scores increase by 1 point for $x = \overline{X}$. Centered scores are obtained by simply subtracting the mean from each score, which can also be achieved by computing standardized scores (i.e., each score is subtracted from the mean and divided by the variable's standard deviation). Note that centering or standardizing predictors does not affect the statistical significance of the test of the null hypothesis H_0: $\beta_3 = 0$ or H_0: $\psi_2^2 - \psi_1^2 = 0$ and it does not have any effect on the observed values of the difference between R_2^2 and R_1^2 (see Aguinis, 2004, Chapter 3, for a more detailed discussion of centering). However, centering can have an important effect on the interpretation as well as values for the intercept as well as coefficients for the first-order effects (i.e., b_1 and b_2 in Equation (2)).

Recommendation #11 refers to creating graphs to illustrate the nature of the interaction effect. Because the interpretation of b_3 can be confusing, it is typically useful to create a graph displaying the X–Y relationship for various levels or values of Z. When Z is a categorical variable, producing a graph involves first creating regression equation for each of the groups. Assuming a binary variable Z and the use of dummy coding (i.e., 0s for members of one group and 1s for members of the second group), reworking Equation (2) for Group 1 (i.e., $z = 0$ for members of this group) yields the following predicted scores for y:

$$\hat{y} = a + b_1x + b_2z + b_3x \cdot z$$

$$\hat{y} = a + b_1x + b_20 + b_3x \cdot 0$$

$$\text{Group 1 equation} : \hat{y} = a + b_1x$$

DOI: 10.1002/job

Similarly, reworking Equation (2) for Group 2 assuming that $z = 1$ for members of this group yields the following predicted scores for y:

$$\hat{y} = a + b_1 x + b_2 z + b_3 x \cdot z$$

$$\hat{y} = a + b_1 x + b_2 1 + b_3 x \cdot 1$$

$$\hat{y} = a + b_1 x + b_2 + b_3 x$$

$$\text{Group 2 equation}: \hat{y} = a + b_2 + x(b_1 + b_3)$$

Now, the X–Y relationship can be plotted for each group. To do so, we can use values of one standard deviation below the mean, the mean, and one standard deviation above the mean for X. Aguinis (2004, Chapter 3) includes specific examples.

If Z is a quantitative variable, Equation (2) can be reworked as follows:

$$\hat{y} = (b_2 z + a) + (b_1 + b_3 z)x \tag{5}$$

Now, the X–Y relationship can be plotted by selecting any value for Z.

The equation describing the relationship between X and Y for a specific value of Z is called a *simple regression equation* and the slope of Y on X at a single value of Z is called a *simple slope*. For example, in Equation (5), we can use values of one standard deviation below the mean, the mean, and one standard deviation above the mean for Z. O'Connor (1998) and Preacher, Curran, and Bauer (2006) include examples as well as a description of computer programs in SAS, SPSS, and *R* that allow for the creation of plots to more easily understand the nature of the interaction effect, including the plotting of simple slopes. *R* is an open-source and free statistical package that is becoming increasingly popular because it has several advantages over some of the more established commercially available packages (Culpepper & Aguinis, 2010). The Preacher et al. (2006) programs also allow for plotting *regions of significance*, which are values of Z between which the simple slope of Y on X is statistically significant. Note that centering does not affect simple slopes or regions of significance in any way and the interaction plot will be exactly the same when using uncentered and centered scores with the only difference that it will include different values on the X-axis (Aiken & West, 1991).

The last recommendation in Table 1, which also addresses issues of interpreting and reporting MMR results, involves an understanding of the size of the interaction effect. Obviously, it is first important to actually be able to detect an existing effect by having sufficient statistical power. However, if we do have sufficient statistical power and do find an interaction effect that is statistically significant, how do we know about the size of this interaction effect? Two types of metrics are readily available: Measures of fit and measures of prediction. One measure of fit is $R_2^2 - R_1^2 = \Delta R^2$, which indicates the sample-based proportion of variance in Y explained by the interaction effect above and beyond the variance explained by the first-order effects of X and Z. In the case of categorical moderators and when the homogeneity of error variance assumption is violated, a more appropriate estimate of effect size is a modified f^2 estimate (Aguinis, 2004; program available at http://mypage.iu.edu/~haguinis/). In contrast to ΔR^2, f^2 is the ratio of systematic variance accounted for by the moderator variable relative to unexplained variance in Y in Equation (2) (i.e., $1 - R_2^2$). So, for example, if $f^2 = .025$, the interpretation is that the interaction effect between X and Z explains 2.5% of the variance in Y that was unexplained by the first-order (i.e., X and Z) and the interactive effects of X and Z (i.e., $X \cdot Z$). In terms of measures of prediction, an effect size estimate is b_3 (*cf*. Equation (2)). In the case of continuous moderators, b_3 indicates how the slope of Y on X changes per a 1-unit change in the moderator Z. So, for example, if $b_3 = -2$, this means that when the moderator Z increases by 1 point, the slope of Y on X decreases 2 points. In the case of a binary categorical moderator, and when dummy-coding is used (i.e., 1s for

DOI: 10.1002/job

members of one subgroup and 0s for members of the other subgroup), b_3 indicates the difference in slopes of Y on X between the subgroup coded as 1 and the subgroup coded as 0. So, for example, if $b_3 = 5$, this means that there is a 5-point difference in the slope of Y on X in favor of the subgroup coded as 1 compared to the subgroup coded as 0.

Conclusions

Interaction effects play a pivotal role in the theoretical development of the organizational sciences and all sciences in general because they allow us to understand the conditions under which relationships between variables change in strength and direction. Hence, interaction effects provide information on the boundary conditions of theories and their predictions. As such, the better our understanding of how to estimate interaction effects, the better our theories will be because they allow us to know conditions under which theories have better or worse explanatory and predictive power. As noted by Aguinis, Pierce, Bosco et al. (2009), ‘‘Although the improvement of methodological tools in the absence of good theory is not likely to produce important advances, theory cannot advance in the absence of good empirical methods either’’ (p. 109). Given this recursive relationship between theory and methods, many of the best-practice recommendations offered in the present article address issues that pertain to the interplay between theory and methods. For example, the better the rationale for the interaction effect (i.e., recommendation #1) and the better the understanding of the underlying constructs and how to measure them (i.e., recommendation #4), the more likely it is that existing interaction effects will be detected. It is precisely this interplay between theory and methods that has the greatest potential in terms of allowing us to improve the scientific value of organizational research.

In addition to the theory and methods interplay, we must also consider the research and practice interplay. Many authors have lamented the divide between research and practice in organizational behavior and related fields (e.g., Cascio & Aguinis, 2008). Reporting interaction effect size estimates such as f^2 and ΔR^2 does not necessarily provide information on the practical importance of a given effect. In many contexts, small effect sizes are very meaningful for practice (Aguinis, Werner, Abbott, Angert, Park, & Kohlhausen, 2009). Conversely, in other contexts a large effect size may not be very meaningful and impactful. Thus, Aguinis, Werner et al. (2009) proposed a ‘‘customer-centric’’ approach to reporting research results, which involves conducting a qualitative study that describes the importance of the results for specific stakeholder groups in specific contexts. If we are concerned with practical significance from the perspective of practitioners, this qualitative study would involve determining the meaning of the interaction effect from the perspective of practitioners (i.e., using their own language and context). In short, it is common practice to report the statistical significance or statistical non-significance of an interaction effect. However, it is important to also accurately describe the nature and magnitude of the interaction effect in practical terms.

Acknowledgements

The authors thank Thomas A. Wright and Douglas G. Bonett for comments on previous drafts.

Author biographies

Herman Aguinis is the Dean's Research Professor and Professor of Organizational Behavior and Human Resources at Indiana University's Kelley School of Business and has been a visiting scholar at universities in China, Malaysia, Singapore, Argentina, France, Spain, Puerto Rico, Australia, and South Africa. He has written more than 70 refereed journal articles and delivered more than 160 presentations at professional conferences. He is the author of *Applied Psychology in Human Resource Management* (2011, with W. Cascio), *Performance Management* (2009), and *Regression Analysis for Categorical Moderators* (2004); and the editor of *Opening the Black Box of Editorship* (2008, with Y. Baruch, A. Konrad, & W. Starbuck) and *Test-score Banding in Human Resource Selection* (2004).

Ryan K. Gottfredson is a doctoral student in organizational behavior at Indiana University's Kelley School of Business. He received his Bachelor's degree from Brigham Young University in Business Administration with an emphasis in Finance. His current research includes research methods and analysis, trust violations, and personnel selection.

References

Aguinis, H. (2004). *Regression analysis for categorical moderators*. NY: Guilford.

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., Boik, R. J., & Pierce, C. A. (2001). A generalized solution for approximating the power to detect effects of categorical moderator variables using multiple regression. *Organizational Research Methods*, *4*, 291–323.

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (in press). Revival of test bias research in reemployment testing. *Journal of Applied Psychology.*

Aguinis, H., Petersen, 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.

Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S., (2009). First decade of Organizational Research Methods: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, *12*, 69–112.

Aguinis, H., Pierce, C. A., & Culpepper, S. A. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods*, *12*, 623–652.

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

Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2009). Customer-centric science: Reporting research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods.* Advance online publication. doi:10.1177/1094428109333339

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

Alexander, R. A., & DeShon, R. P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin*, *115*, 308–314.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, *93*, 549–562.

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, *93*, 1062–1081.

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.

DOI: 10.1002/job

Culpepper, S. A., & Aguinis, H. (2010). R is for revolution: A cutting-edge, free, open source statistical package. *Organizational Research Methods.* Advance online publication. doi: 10.1177/1094428109355485

Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.

Jaccard, J. J. (2001). *Interaction effects in logistic regression*. Newbury Park, CA: Sage.

O'Connor, B. P. (1998). All-in-one programs for exploring interactions in moderated multiple regression. *Educational and Psychological Measurement*, *58*, 833–837.

Oldham, G. R., Hackman, J. R., & Pearce, J. L. (1976). Conditions under which employees respond positively to enriched work. *Journal of Applied Psychology*, *61*, 395–403.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448.

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

Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous variables: Power and sample size considerations. *Organizational Research Methods*, *12*, 510–528.

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

DOI: 10.1002/job