

REPLY

 Assessing and Interpreting Interaction Effects: A Reply
to Vancouver, Carlson, Dhanani, and Colton (2021)

 Chad H. Van Iddekinge¹, Herman Aguinis², James M. LeBreton³,
Jeremy D. Mackey⁴, and Philip S. DeOrtentiis⁵
¹ Department of Management and Entrepreneurship, University of Iowa

² Department of Management, The George Washington University

³ Department of Psychology, The Pennsylvania State University

⁴ Department of Management, Auburn University

⁵ School of Human Resources and Labor Relations, Michigan State University

Van Iddekinge et al. (2018)'s meta-analysis revealed that ability and motivation have mostly an additive rather than an interactive effect on performance. One of the methods they used to assess the ability \times motivation interaction was moderated multiple regression (MMR). Vancouver et al. (2021) presented conceptual arguments that ability and motivation should interact to predict performance, as well as analytical and empirical arguments against the use of MMR to assess interaction effects. We describe problems with these arguments and show conceptually and empirically that MMR (and the ΔR and ΔR^2 it yields) is an appropriate and effective method for assessing both the statistical significance and magnitude of interaction effects. Nevertheless, we also applied the alternative approach Vancouver et al. recommended to test for interactions to primary data sets ($k = 69$) from Van Iddekinge et al. These new results showed that the ability \times motivation interaction was not significant in 90% of the analyses, which corroborated Van Iddekinge et al.'s original conclusion that the interaction rarely increments the prediction of performance beyond the additive effects of ability and motivation. In short, Van Iddekinge et al.'s conclusions remain unchanged and, given the conceptual and empirical problems we identified, we cannot endorse Vancouver et al.'s recommendation to change how researchers test interactions. We conclude by offering suggestions for how to assess and interpret interactions in future research.

Keywords: interactions, moderated multiple regression, ability, motivation, performance

Supplemental materials: <https://doi.org/10.1037/apl0000883.supp>

Various theories propose that ability (A) and motivation (M) interact to predict performance (P , e.g., Locke & Latham, 1990; Maier, 1955; Vroom, 1964). More specifically, these theories predict that as M increases, so does the strength of the positive relation between A and P . However, empirical evidence for this claim has been mixed because some studies have found support for $A \times M$ interactions (e.g., Fleishman, 1958; Perry et al., 2010), whereas others have not (e.g., Mount et al., 1999; Sackett et al.,

1998). Van Iddekinge et al. (2018) conducted a meta-analysis to resolve this issue. Using mainly original data or software output that primary study authors provided, Van Iddekinge et al. used moderated multiple regression (MMR) to compare the change in the multiple correlation coefficient (ΔR) from a model that included A and M (i.e., the additive model—first step) to a model that included A , M , and the product of A and M (i.e., the multiplicative model—second step). In addition to MMR, they also compared the relative weights for the two models and simple slopes for $A-P$ relations at different levels of M .

Results from all three types of analyses indicated that $A \times M$ interaction effects were small, which is consistent with research on the size of interactions in applied psychology (e.g., Aguinis et al., 2005). In addition, in line with previous meta-analyses (e.g., Schmidt & Hunter, 1998), the additive effects of A and M were moderate to large. Thus, Van Iddekinge et al. concluded that “including ability–motivation interactions in future theoretical explanations or empirical models will add complexity . . . but not necessarily increase understanding or prediction of performance. Instead, in most cases . . . researchers and practitioners can focus on the more parsimonious additive effects of ability and motivation” (p. 274).

Chad H. Van Iddekinge  <https://orcid.org/0000-0001-9813-0660>

Herman Aguinis  <https://orcid.org/0000-0002-3485-9484>

James M. LeBreton  <https://orcid.org/0000-0001-6748-380X>

Jeremy D. Mackey  <https://orcid.org/0000-0001-7790-5949>

Philip S. DeOrtentiis  <https://orcid.org/0000-0003-4299-4093>

The authors have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Chad H. Van Iddekinge, Department of Management and Entrepreneurship, University of Iowa, Tippie College of Business, 108 John Pappajohn Business Building, Iowa City, IA 52242-1994, United States. Email: chad-vaniddekinge@uiowa.edu

Vancouver et al. (2021) stated that Van Iddekinge et al. may “have drawn the wrong conclusion about the functional form of ability and motivation’s influence on performance” (p. 2). Specifically, Vancouver et al. challenged MMR as follows:

“... the standard practice of indexing the effect size for a moderator in terms of the incremental validity found when adding a product term, xz , to a regression equation that already includes the first-order terms (i.e., the predictor, x , and the moderator, z). The issue is that such a procedure assigns all the variance in the criteria, y , shared among x , z , and xz to the first-order terms (p. 2) ... if some joint effect is occurring (i.e., the product term makes a significant contribution), the procedure can lead to spurious conclusions that the first-order terms are also contributing to y independently of each other (i.e., nonjointly) as well as produce a substantially downwardly biased estimate of the size of the joint effect’s contribution to y ” (p. 3).

Given these concerns, Vancouver et al. suggested abandoning the two-step, hierarchical MMR procedure and instead basing all inferences regarding interactions on the statistical significance of the regression coefficient for the cross product term from the second step of MMR.

Understanding the issues Vancouver et al. raised is important for drawing correct conclusions not only from the study by Van Iddekinge et al. but also from any study that reports results based on MMR. Indeed, if Vancouver et al.’s concerns are justified, they would call into question the validity of conclusions from thousands of MMR tests of interactions (e.g., Aguinis et al., 2005, 2017; O’Boyle et al., 2019). Their article also could change how future research tests theories and interventions that involve interactions or moderator variables.

We address Vancouver et al.’s concerns with MMR and evaluate their proposed alternative to testing interactions. We first discuss Vancouver et al.’s interpretation of research that has tested the $A \times M$ interaction and research on best practices for testing interactions more generally. Then, we address their conceptual, analytic, and empirical arguments for the existence of $A \times M$ interactions and against the use of MMR to test this and other interactions. We also implement the approach Vancouver et al. recommended for assessing interactions using primary data from Van Iddekinge et al. We conclude by providing suggestions for assessing and interpreting interaction effects in the future.

Problems With Vancouver et al.’s Arguments Based on Prior Research

Vancouver et al. noted that “the notion that ability and motivation affect performance multiplicatively was considered settled theory prior to the Van Iddekinge et al. (2018) meta-analysis” (p. 6). However, there are problems with this characterization of the literature. First, the functional form of the effects of A and M on P was far from settled because primary studies regularly failed to find interactions (e.g., Dachler & Mobley, 1973; Mount et al., 1999; Sackett et al., 1998), and literature reviews noted ambiguous theorizing and inconsistent empirical evidence (e.g., Terborg, 1977). Second, even researchers whose theories proposed an $A \times M$ interaction effect sometimes expressed uncertainty about the specific form of the relationship. For example, Campbell (1990)—whom Vancouver et al. cited to support their statement that “scholars have long thought that

the functional form of those two factors is best described as multiplicative ...” (p. 20)—noted that “... the precise form [of the equation for predicting performance] is obviously not known and perhaps not even knowable” (Campbell, 1990, p. 706). In fact, Campbell and colleagues later concluded that “it is possible that the amount of improvement provided by a multiplicative functional form would not be substantial” (McCloy et al., 1994, p. 503).

There also are problems with Vancouver et al.’s description of research on testing interaction effects. For one, the fact that MMR is an appropriate method to assess interaction effects has been established for several decades (e.g., Arnold & Evans, 1979; Cohen, 1968, 1978; Cronbach, 1987). In the 1970s and 1980s, given concerns that it was difficult to detect hypothesized interactions (e.g., Zedeck, 1971), there was a vigorous debate of alternative procedures and interpretations (e.g., Althausen, 1971; Arnold, 1982; Cronbach, 1987; Darrow & Kahl, 1982; Friedrich, 1982; Sockloff, 1976; Stone & Hollenbeck, 1984). However, these debates have long been settled, and methodologists in applied psychology and other fields have widely endorsed MMR (e.g., Aguinis, 2004; Aiken & West, 1991; Cohen et al., 2003; Cortina, 1993; Dalal & Zickar, 2012; Evans, 1991; LeBreton et al., 2013; McClelland & Judd, 1993).

Vancouver et al. also referred to Arnold and Evans (1979) as a main source of support for their arguments against MMR. For example, Vancouver et al. stated that “if some joint effect is occurring, analytic methods like MMR cannot determine how much of the effects of the x and z are joint and how much, if any, are nonjoint, unless the interaction is disordinal and fully symmetric (Arnold & Evans, 1979, p. 6)”. However, nowhere in their article did Arnold and Evans raise this point. In contrast, one of the main goals of Arnold and Evans was to “demonstrate theoretically and algebraically how and why hierarchical multiple regression permits a valid test for the existence of interaction effects” (p. 42). In fact, Evans (1991) reiterated his support of MMR, noting that “an interaction term [$X \times Z$] ... involves variance explained over and above the variance explained by the main effects of [X and Z]. Cohen (1978) has demonstrated convincingly that it is the partialled product (partialing out the two component variable correlations with the dependent variable) that indexes the interaction effect ...” (p. 7).

Problems With Vancouver et al.’s Conceptual Arguments

As a conceptual argument for expecting large $A \times M$ interaction effects, Vancouver et al. presented a thought experiment in which two famous actors (i.e., Dwayne “the Rock” Johnson and Jim Parsons who played Sheldon Cooper in the television show “The Big Bang Theory”) had zero motivation to lift a barbell (because they were playing video games) or had zero ability to do so (because they were paralyzed). Vancouver et al. suggested that the multiplicative model would correctly predict that performance would be zero in both scenarios. In contrast, the additive model would be incorrect because ability would not have an independent, additive effect in the first situation and motivation would not have such an effect in the second situation.

Although the multiplicative model would hold in these hypothetical scenarios, these situations do not exist in real organizations because employees with zero physical ability are not hired to perform physically demanding jobs, and employees with

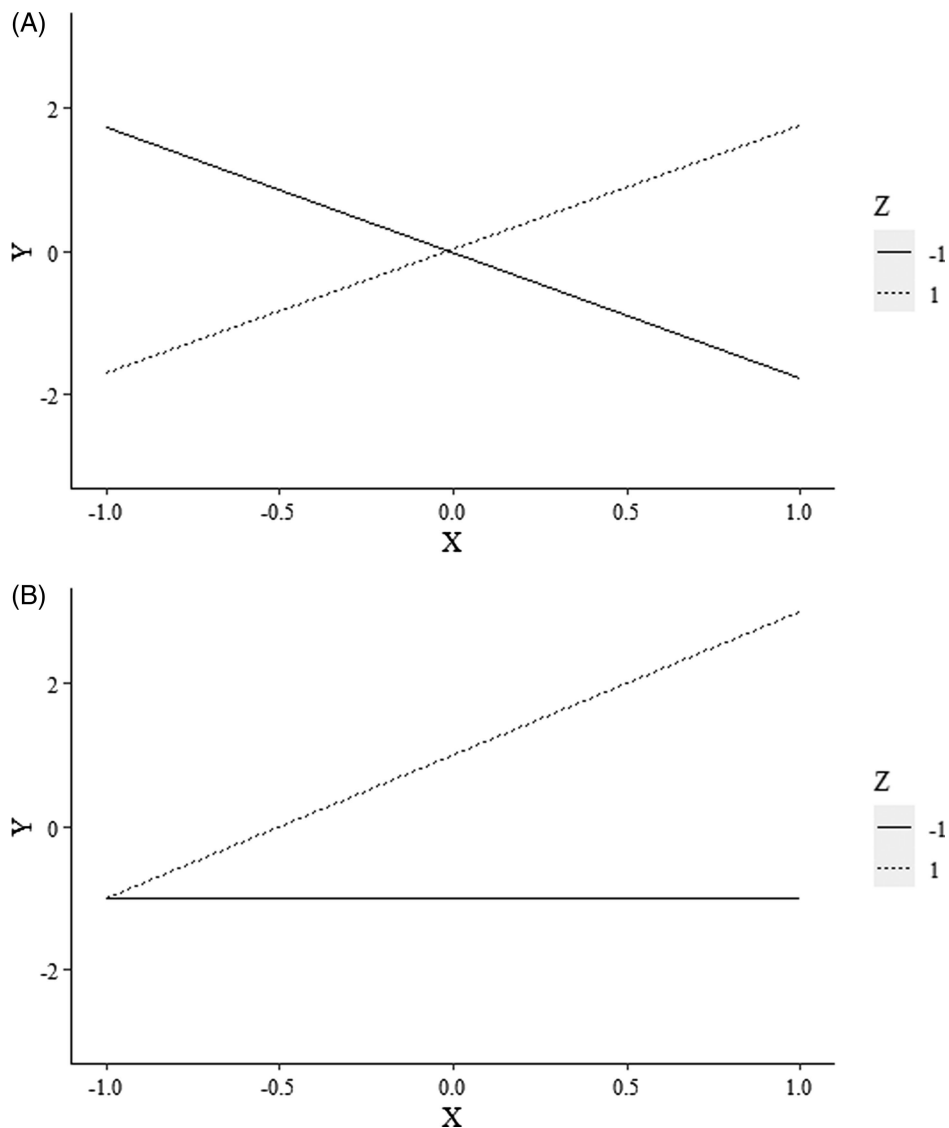
zero motivation tend to quit or get fired. Relatedly, Vancouver et al. stated that "... the more one samples closer to the conceptual zero value (e.g., no effort; ability negated by paralysis), the less variance in y is shared by x , z , and xz , putting the interaction in greater relief" (p. 18). Again, a scenario in which ability is zero does not exist in real organizations and represents an unrealistic instance of an extreme-groups design. Although such designs increase the chances of detecting an interaction (McClelland & Judd, 1993), the effect sizes they yield are upwardly biased and not generalizable (Cortina & DeShon, 1998).

Moreover, the only situation in which the effects of X and Z on Y would be purely multiplicative (i.e., $Y = X \times Z$) is when X and Z are bivariate unrelated to Y and only the interaction effect explains variance in Y . In this case, the form of the interaction must be

disordinal (Cohen et al., 2003). As we show in Figure 1, Panel A, a disordinal interaction takes the form of a complete crossover interaction. However, there is extensive evidence that both A and M have nontrivial (and most certainly nonzero) bivariate relationships with P (e.g., Hunter & Hunter, 1984; McHenry et al., 1990; Van Eerde & Thierry, 1996; Wright, 1990). Thus, any interaction between A and M would most likely take the form of an ordinal interaction as shown in Figure 1, Panel B. Furthermore, a disordinal interaction would suggest that A or M become detrimental to P at low values of the other, which is unrealistic given that both constructs tend to relate positively to performance. Thus, based on the existing evidence, it is unlikely that the effects of A and M on P would be consistent with a pure multiplying model.

Figure 1

Illustrations of Ordinal and Disordinal Interactions. Panel A: Disordinal Interactions, Data Generating Function for Y Based on Pure Multiplying Model. Panel B: Ordinal Interactions, Data Generating Function for Y Based on Combined Adding-Multiplying Model



Finally, in challenging Van Iddekinge et al.'s conclusion that an additive model is more parsimonious than a multiplicative model, Vancouver et al. argued that

A theory that states "Performance = Ability \times Motivation" does not seem any less parsimonious than a theory that states "Performance = Ability + Motivation Indeed, the multiplicative model is arguably more parsimonious because only one scaling value is needed (see Model 2), whereas with the additive model, two are needed in addition to the intercept term (see Model 1). Thus, a desire for parsimony cannot be used as a reason to reject the multiplicative hypothesis (pp. 13–14).

We disagree that multiplicative relations are more parsimonious than additive ones. A pure adding model is simpler than a pure multiplying model because the relationship between X and Y is not conditional on the specific values of Z (see also Murphy & Russell, 2017). Furthermore, if X and Z are continuous, a multiplicative model would result in an infinite number of conditional slopes when Y is regressed onto X . In contrast, even if the additive effects of both X and Z are significant, it would require only two lines to graph the relationships Y has with X and Z .

Problems With Vancouver et al.'s Analytical Arguments

Vancouver et al. raised several analytical concerns about MMR. We highlight problems with their arguments concerning how MMR partitions variance, differences between pure multiplying and combined adding–multiplying models, and bias in R^2 estimates.

Variance Partitioning in MMR

Vancouver et al. suggested that MMR is unable to provide a "clean, unpartitioned effect size estimate . . . unless the interaction is disordinal and fully symmetric (Arnold & Evans, 1979)" (p. 6). Moreover, they asserted that MMR is unable " . . . to determine the degree to which the variance in y shared by x , z , and xz 'belongs' to the additive or multiplicative conceptual models" (p. 6). Vancouver et al.'s concern appears to be based on the belief that the lower-order terms "steal" variance explained in Y that should be assigned to the interaction effect. Although Cohen (1978) and Cronbach (1987) proved this argument to be false, we address Vancouver et al.'s concern by considering how MMR partitions variance.

In the first step of MMR, the following equation is applied to the data:

$$Y = b_0 + b_1(X) + b_2(Z) + e. \quad (1)$$

The R^2 from this step represents the variance in Y predicted by the additive effects of X and Z . The R^2 from Equation 1 is visually shown in Figure 2, Panel A and is equal to $a + b + c$. These variance components reflect the variance in Y uniquely explained by X (i.e., a), uniquely explained by Z (i.e., b), and jointly explained by X and Z (i.e., c).

In the second step of MMR, the following equation is applied to the data:

$$Y = b_0 + b_1(X) + b_2(Z) + b_3(XZ) + e. \quad (2)$$

Figure 2, Panel B shows the R^2 for this model is equal to $a_1 + a_2 + b_1 + b_2 + c + d$ (or equivalently, $a + b + c + d$). This value

represents the proportion of variance in Y that is a function of both the additive ($a + b + c$) effects from Step 1 and the interactive effect (d). Some of the variance in Y is explained by X only (i.e., a_1), some is explained by Z only (i.e., b_1), and some is uniquely explained by the interaction of X and Z (i.e., d). In addition, some of the variance continues to be explained by the additive effects of X and Z , specifically a_2 , b_2 , and c . Because this variance also overlaps with the cross product term (XZ), Vancouver et al. concluded that there is no "clean" estimate of the interaction effect. This conclusion implies that the additive effects of X and Z contaminate the estimate of the interaction effect.

The fact that some of the variance in XZ overlaps with the variance in its components X and Z (i.e., they are "multicollinear") does not indicate that the components somehow steal variance that should be credited to the interaction effect (Cohen et al., 2003; Dalal & Zickar, 2012; LeBreton et al., 2013). The interaction effect is equal to area d , nothing more, nothing less. If there is any "theft" of variance, it is the cross product term that steals variance from the lower-order effects. As shown in Figure 2, variance XZ shares with X and Z in the second step of MMR (i.e., a_2 , b_2 , and c) is completely explained by the additive effects in the first step (i.e., a , b , and c). Thus, this variance must be assigned to the additive effects of X and Z and not to the interaction effect (Aiken & West, 1991; Arnold & Evans, 1979; Cohen et al., 2003; Cronbach, 1987; LeBreton et al., 2013).

Distinguishing Between Pure Multiplying and Combined Adding–Multiplying Models

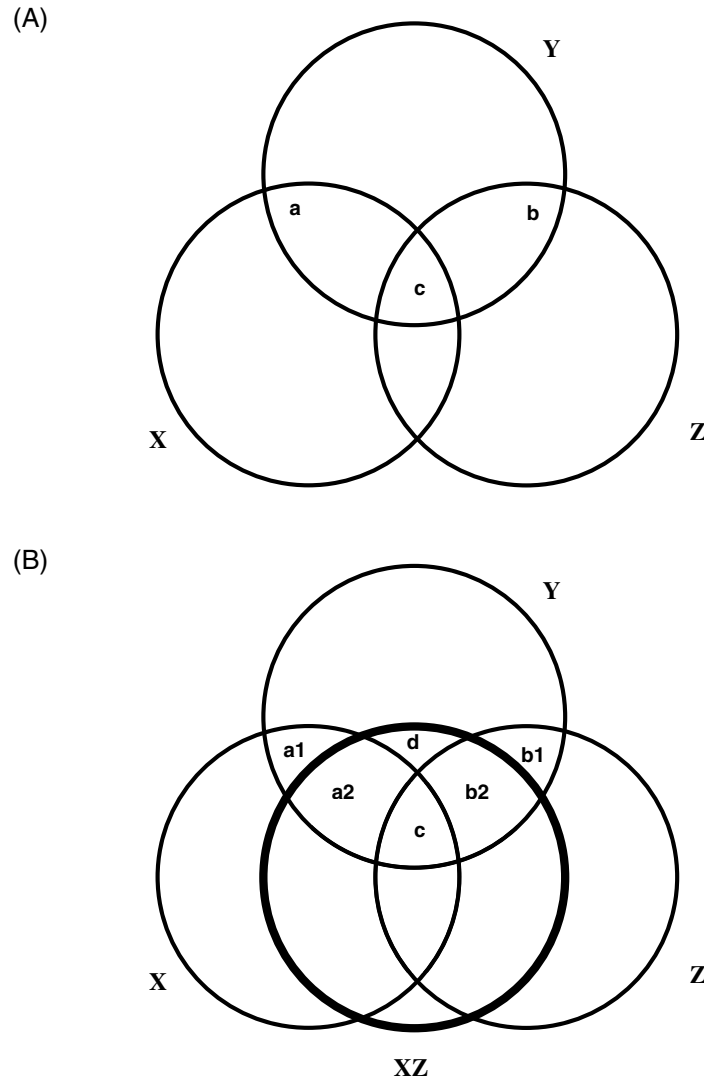
Vancouver et al. asserted that MMR cannot discern whether an interaction effect was generated by a pure multiplying conceptual model ($Y = XZ$) or by a combined adding–multiplying conceptual model ($Y = X + Z + XZ$). To support their assertion, they quoted Arnold and Evans (1979): "There is no empirical basis for distinguishing between a 'pure' multiplying model and a 'combined' adding–multiplying model" (emphasis in the original, p. 51). However, Arnold and Evans stated this in the context of cautioning researchers about interpreting the coefficients for X and Z in the presence or absence of an interaction.

To illustrate this point, Arnold and Evans (1979) transformed X and Z into new predictors, X^* and Z^* . Next, they compared the results of MMR applied to the original predictors (i.e., X , Z , and XZ) to results applied to the transformed predictors (i.e., X^* , Z^* , and X^*Z^*). When they applied Equation 2 to the original predictors, b_1 , b_2 , and b_3 were all significantly different from zero. However, when they applied Equation 2 to the transformed predictors, b_3 remained the same, whereas b_1 and b_2 were no longer different from zero. Thus, although the transformations changed the coefficients for X and Z , they did not alter the regression coefficient for XZ . Arnold and Evans summarized their findings as follows:

These results also serve to underscore the fact that the appropriate test of the multiplicative model is not the relative magnitude of the [regression weights], nor is it how much variance the [cross product] term explains. Rather, it is the difference in variance explained by the combined adding–multiplying model in comparison with the simple additive model . . . [the R and R^2 values from these models] are invariant across transformations, while both the [regression weights] and zero-order correlations between the dependent variable and the product . . . fluctuate with the transformation (pp. 52–53).

Figure 2

Venn Diagrams Illustrating Partitioning of Variance in Moderated Multiple Regression (MMR). Panel A: R^2 Obtained From the First Step of MMR: $Y = b_0 + b_1(X) + b_2(Z) + e$. Panel B: R^2 Obtained From the Second Step of MMR (i.e., $Y = b_0 + b_1(X) + b_2(Z) + b_3(XZ) + e$)



In summary, Arnold and Evans (1979) did not conclude that it is impossible to distinguish between a pure multiplying model and a combined adding–multiplying model, nor was their critique aimed at MMR or transformations. Rather, they showed how basing conclusions about additive versus interactive effects on the relative magnitude of the regression coefficients can be misleading. Specifically, although linear transformations of X and Z (e.g., centering) may impact the size and significance of b_1 and b_2 (in Equation 2), such transformations have no impact on b_3 , nor do they affect the size of the interaction effect (i.e., ΔR^2) or its functional form (see also Aguinis, 2004; Aiken & West, 1991; Cohen et al., 2003).

Bias in R^2

Vancouver et al. claimed that ΔR^2 is a downwardly biased estimate of the interaction effect (e.g., p. 3). Instead, it is well-known that the opposite is true as: R^2 (and by extension ΔR^2) is an upwardly biased estimate of the population parameter (e.g., Cohen et al., 2003; Shieh, 2008; Wherry, 1931). Indeed, this is why a best practice is to use adjusted R^2 by applying formulas that correct for shrinkage (e.g., Ployhart & Hakel, 1998; Raju et al., 1997; Yin & Fan, 2001). As Murphy and Russell (2017) noted, “small-sample estimates of ΔR^2 (which are sometimes offered as evidence of ‘substantive’ incremental prediction due

to XZ interactions) are likely to show substantial shrinkage when generalized to the population or other samples” (p. 556).

Problems With Vancouver et al.’s Empirical Arguments

Vancouver et al.’s Monte Carlo Simulation

Vancouver et al.’s empirical arguments against MMR were based on the results of a Monte Carlo simulation in which they “... created a data set with 1,000 cases composed of two constructs, X and Z , generated from a normal distribution with a mean of 0.5 and a standard deviation of 0.1 ... [and] then created Y by multiplying X and Z (i.e., $Y = XZ$)” (p. 8).

Although Vancouver et al. stated that Y was solely a function of the *interaction effect*, they set their Y scores equal to the *cross product term*. The problem with this approach is that the cross product term is not isomorphic with the interaction effect because, in addition to the interaction effect of X and Z , the cross product term also carries information about the main effects of X and Z (Cohen et al., 2003). In contrast, the interaction effect represents the portion of the cross product that is statistically independent of the main effects. To simulate data that were solely a function of the interaction effect, Vancouver et al. should have used the residuals obtained by regressing XZ onto $X + Z$ (LeBreton et al., 2013). Hereafter, we refer to the interaction effect as XZ_{res} to highlight the fact that the interaction reflects the residual (i.e., “res”) variance in the cross product term that remains after removing variance shared with the first-order effects of X and Z .

Because Vancouver et al.’s simulation used the cross product rather than the interaction effect, there was a misalignment between the conceptual model they intended to test (i.e., pure multiplying model; $Y = XZ_{\text{res}}$) and the model they actually tested (i.e., cross product model; $Y = XZ$); which, as we illustrate below, was nearly identical to the combined adding–multiplying model; $Y = X + Z + XZ$). The misalignment is evident in the observed correlations of .710 between Y and X and .707 between Y and Z (Vancouver et al., Table 1). If the data were a function of a pure multiplying model, these correlations would be exactly zero (see also Table 1 in Evans, 1985). The fact that additive effects of X and Z accounted for 98% of the variance in Y (see Vancouver et al.’s Table 2) confirms this misalignment between their theory and the data used to test that theory because if the effects of X and Z were solely multiplicative, the additive effects would be zero. Finally, as we noted, data simulated to mimic a pure multiplying model should result in a perfectly symmetrical disordinal interaction, whereas data simulated to mimic a combined adding–multiplying model should result in an ordinal interaction. We graphed the results of Vancouver et al.’s simulation (see our Figure 3), which verified two important points. First, interaction effects are invariant across linear transformations (in this case, raw vs. mean-centered predictors). Second, the form of the interaction reflects a strong ordinal effect, which would not be the case if the data were simulated to mimic a pure multiplying effect.

Correcting Vancouver et al.’s Simulation

Because Vancouver et al. did not simulate scores based on a pure interaction effect (XZ_{res}), some of their conclusions are inaccurate.

To demonstrate this, we replicated and extended their simulation to assess the accuracy of MMR. Specifically, we simulated four possible conceptual models (the [online supplement includes details of all procedures and results](#)) as follows:

1. *Adding model*: Y is only a function of the additive effects of X and Z ($Y = X + Z$).
2. *Cross product model*: Y is a function of the *unresidualized* cross product of X and Z ($Y = XZ$). This is the model Vancouver et al. used to simulate a pure multiplying model.
3. *Multiplying model*: Y is solely a function of the interaction effect of X and Z , which we simulated using the residualized cross product term ($Y = XZ_{\text{res}}$). This is the model Vancouver et al. *should have* used in their simulation.
4. *Combined adding–multiplying model*: Y is a combination of the additive and multiplicative effects of X and Z ($Y = X + Z + XZ$, which is equivalent to $Y = X + Z + XZ_{\text{res}}$).

We then applied MMR to the data generated from each of the four models and compared inferences regarding the presence, size, and form of the interaction effect. MMR assesses the statistical significance of the interaction using the F test for the ΔR^2 from the adding model to the combined adding–multiplying model in Equation 2. As Vancouver et al. correctly noted, the p value for this F -based test is identical to the p value for the t statistic for the test that $\beta_3 = 0$. In addition, MMR assesses the magnitude of the interaction effect using ΔR^2 , which reflects the variance in Y the interaction explains beyond the variance the first-order effects explain.

Table 1 shows the simulation results. In terms of statistical significance, results revealed that MMR correctly inferred the absence of an interaction effect for the adding model ($p = 1.00$) and correctly detected an interaction for the cross product, multiplying, and combined adding–multiplying models (all $p < .01$). In addition, MMR’s partitioning of the explained variance was consistent with the data-generating functions. Specifically, for the adding model, MMR correctly assigned all explained variance to the lower-order effects ($R^2 = 1.00$) and none to the interaction effect ($\Delta R^2 = .00$). For the cross product and combined adding–multiplying models, MMR estimated large lower-order effects ($R^2 = .980$ and $.998$) and very small interaction effects ($\Delta R^2 = .020$ and $.002$). Finally, for the multiplying model, MMR correctly detected no additive effects ($R^2 = .00$) and assigned all predicted variance to the interaction effect ($\Delta R^2 = 1.00$). Overall, these results verify that when the interaction effect is simulated correctly, MMR provides accurate information about the presence, size, and form of the interaction.

Reanalysis of Van Iddekinge et al.’s Data Using Vancouver et al.’s Approach

Vancouver et al. contended that “the [ΔR^2] procedure is merely meant to challenge the multiplicative form, not determine the relative contribution of the multiplicative effect relative to an

Figure 3

Graphs of Results From Vancouver et al.'s Simulation. Panel A: Interaction Effect Based on Construct-Level Data (Y, X, Z, XZ). Panel B: Interaction Effect Based on Observed Data (y,x,z,xz). Panel C: Interaction Effect Based on Centered (Observed) Data (y, x_c, z_c, x_cz_c)

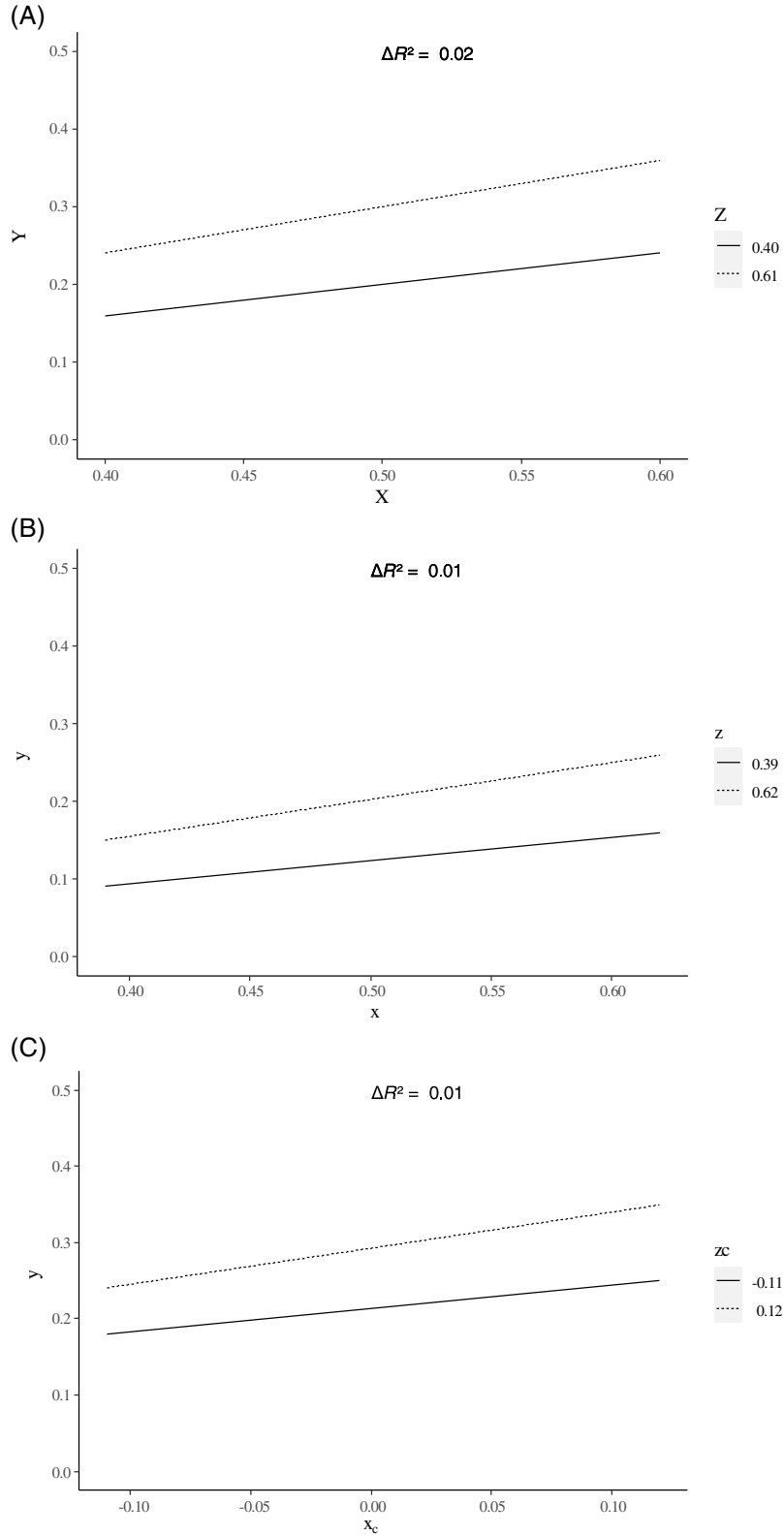


Table 1
Summary of Simulation Results Using Construct-Level Data

| Predictor | Parameter estimate | Y.ADD | | XZ _{res} | | Y.COMB | | Y.XZ | |
|-----------|--------------------|--------|--------|-------------------|--------|--------|--------|--------|--------|
| | | Step 1 | Step 2 | Step 1 | Step 2 | Step 1 | Step 2 | Step 1 | Step 2 |
| — | b_0 | 0.00 | 0.00 | 0.50 | 3.00 | -0.25 | 0.00 | -.25 | 0.00 |
| X | b_1 | 1.00 | 1.00 | 0.00 | -5.00 | 1.50 | 1.00 | .50 | 0.00 |
| Z | b_2 | 1.00 | 1.00 | 0.00 | -5.00 | 1.50 | 1.00 | .50 | 0.00 |
| XZ | b_3 | | 0.00 | | -10.00 | | 1.00 | | 1.00 |
| | R^2 | 1.00 | 1.00 | 0.00 | 1.00 | 0.998 | 1.00 | 0.98 | 1.00 |
| | ΔR^2 | | 0.00 | | 1.00 | | 0.002 | | 0.02 |

Note. Y.ADD = Y scores simulated using a pure adding model (i.e., $Y = X + Z$); $Y.XZ_{res} = Y$ scores simulated using a pure multiplying model (i.e., $Y = XZ_{res}$); Y.COMB = Y scores simulated using a combined adding–multiplying model (i.e., $Y = X + Z + XZ_{res}$); Y.XZ = Y scores simulated using a cross product model (i.e., $Y = XZ$).

additive effect. Critically, the test of the significance of $[b_3]$ when first-order terms are included will do this” (p. 21). Thus, instead of using the two-step MMR, which compares the ΔR^2 from Equations 1 and 2, Vancouver et al. recommended basing inferences about the interaction solely on the statistical significance of b_3 computed from Equation 2.

To illustrate this approach, we applied it to each of 69 unique sets of predictors and criteria from the primary study data in Van Iddekinge et al. ($k = 42$ studies, $N = 24,001$). Results in Table 2 indicate that the cross product term was significant ($p < .05$) in only 7 of the 69 analyses (10%).¹ Furthermore, we graphed the seven significant interactions (using IBM SPSS code from the study by O’Connor [1998]) and found that in five cases, the form of the interaction was consistent with theory (i.e., the $A-P$ relation increased as M increased), and in two cases, the form was the opposite (i.e., the $A-P$ relation decreased as M increased). Thus, theory-consistent $A \times M$ interaction effects were detected only 7% of the time (i.e., 5 out of 69 analyses).

Vancouver et al. suggested that “if no reliably detectable joint effect is occurring, the product term would not be significant and in that case the variance attributed to the first-order terms can be safely assigned to the set of them and the product term can be dropped” (p. 6). The results of these new analyses show that this was indeed the case for 90% of the interactions. Even when an interaction effect was detected, the effect sizes tended to suggest small, ordinal interactions, some of which were inconsistent with theory.²

Limitations of Vancouver et al.’s Recommended Approach

Although we conducted the analyses Vancouver et al. suggested, there are limitations to this approach. First, this test focuses only on the statistical significance of an interaction. As such, this approach is subject to the well-known limitations of null hypothesis significance testing, including reliance on a somewhat arbitrary, yes/no decision criterion (i.e., $p < \text{or } \geq .05$) and low statistical power (Aguinis, 1995; Cohen, 1994; Schmidt, 1996).

Second, unlike MMR, which provides an effect size estimate for the interaction (i.e., ΔR^2), the approach Vancouver et al. recommended does not provide a standardized effect size estimate for the interaction. Rather, it estimates the R^2 for a model containing both additive and interaction effects. Although Vancouver et al. did not provide recommendations for how to interpret R^2 when b_3 is

statistically significant, their critique of Van Iddekinge et al. may shed some light on their viewpoint as follows:

... the method Van Iddekinge et al. used to assess the relative contributions of additive as opposed to multiplicative models of ability and motivation on performance was destined to favor, substantially, the additive model. Indeed, we suspect that the 9.4% relative weight “effect,” which was obtained by dividing change in Multiple R by final Multiple R , underestimated the true effect by 90.6% (e.g., $100\% - 9.4\% = 90.6\%$). This is because, conceptually, it is difficult to imagine an additive, nonjoint effects of ability or motivation on performance (p. 12).³

This statement implies that Vancouver et al. believe *all* the predicted variance in Y should be attributed to the $A \times M$ interaction. Extending this logic to our simulation would render inaccurate inferences. For example, because b_3 was significant in both the combined adding–multiplying and the cross product models, Vancouver et al.’s recommendation would lead us to incorrectly

¹ A well-known potential limitation of statistical significance is low statistical power, which is a particular concern when testing interactions because their effects tend to be very small (Aguinis & Stone-Romero, 1997). Thus, we used the g^* power software (Faul et al., 2009) to compute power for the Van Iddekinge et al. data. Given the mean observed R^2 for the additive model was approximately .10, we computed the power needed to detect interactions that would increase R^2 by .05 and .10 beyond the additive effects. We selected these effect sizes because they seem to be the minimum amount of additional variance that would be useful in most situations (see also Murphy & Russell, 2017). We computed f^2 values corresponding to these R^2 changes, which were .04 and .13, respectively. In addition, we used a power of .80 and an alpha of .05. Results revealed a sample size of at least 159 would be needed to detect a .05 increase in R^2 and sample size of at least 81 would be needed to detect a .10 increase. The mean sample size across the primary studies in Table 2 was 215 ($SD = 157$). Thus, studies generally possessed sufficient power to detect what might be considered nontrivial interaction effects.

² An anonymous reviewer suggested that we also implement Vancouver et al.’s approach by using meta-analysis to estimate the fully corrected correlations (for A , M , $A \times M$, and P) and run a MMR on the meta-analytic correlations. We conducted these analyses (using data from the 69 analyses we used in the previous analyses) and found an R^2 of .202 for Equation 1 and a R^2 of .204 for Equation 2. Thus, the $A \times M$ interaction increased R^2 by only .002, or an additional 0.2% of explained variance in performance beyond the first-order effects of A and M . In short, substantive conclusions remained unchanged.

³ The 9.4% value from the study by Van Iddekinge et al. was not calculated as Vancouver et al. suggested. Rather, this value reflects the mean relative weight percentage for the ability–motivation interaction effect across all studies in their meta-analysis (see their Table 2, p. 263).

Table 2*Significance Tests of the Ability by Motivation Interaction Effect on Performance for Primary Studies From Van Iddekinge et al. (2018)*

| Study | Ability | Motivation | Performance | <i>N</i> | ΔR^2 | β | <i>t</i> (<i>df</i>) | <i>p</i> |
|----------------------------------|---------------------------------|----------------------------|---------------------------------|----------|--------------|---------|------------------------|----------|
| Barros et al. (2014) | GMA | Achievement | Task performance | 151 | .004 | .07 | .76 (147) | .45 |
| | | | Contextual performance | 115 | .001 | -.03 | -.28 (111) | .78 |
| Bell and Kozlowski (2002) | GMA | Effort (on task) | Lab performance | 277 | .004 | .07 | 1.23 (273) | .22 |
| | | Effort (mental) | | | .011 | .11 | 1.94 (273) | .06 |
| Bell and Kozlowski (2008) | GMA | Effort | Lab performance | 350 | .000 | -.01 | -.20 (346) | .84 |
| Bono (2001) | GMA | Effort | Lab performance | 143 | .006 | .08 | .98 (139) | .33 |
| Carretta et al. (2014) | GMA | Achievement | Training performance | 9,396 | .000 | .01 | 1.08 (9,392) | .28 |
| Corker (2012) | | | | | | | | |
| Study 1 | GMA | Effort | Lab performance | 205 | .000 | .03 | .41 (201) | .68 |
| | | Achievement | | 251 | .004 | -.07 | -1.04 (247) | .30 |
| Study 2 | GMA | Effort | Lab performance | 368 | .006 | .08 | 1.79 (364) | .07 |
| | | Achievement | | 424 | .001 | .04 | .77 (420) | .44 |
| Hausdorf and Risavy (2015) | GMA | Achievement | Training performance | 173 | .001 | -.04 | -.52 (169) | .60 |
| Huang (2012) | GMA | Achievement | Lab performance | 124 | .008 | -.09 | -1.06 (120) | .29 |
| | | Effort (persistence) | | | .004 | -.06 | -.87 (120) | .39 |
| | | Effort (time spent) | | | .000 | -.02 | -.28 (120) | .78 |
| | | Effort (attention) | | | .000 | .00 | -.00 (120) | .99 |
| Iliescu et al. (2012) | | | | | | | | |
| Sample 2 | GMA | Achievement | Job performance | 223 | .004 | -.06 | -1.06 (219) | .29 |
| Sample 3 | GMA | Achievement | Job performance | 61 | .000 | -.01 | -.04 (57) | .97 |
| Kluemper et al. (2013) | GMA | Achievement | Task performance | 101 | .013 | .12 | 1.14 (97) | .26 |
| | | | Contextual performance | | .000 | -.01 | -.10 (97) | .92 |
| | | | Counterproductive work behavior | | .013 | .12 | 1.12 (97) | .27 |
| Kozlowski and Bell (2006) | GMA | Effort (on task) | Lab performance | 539 | .007 | -.09 | -2.24 (535) | .03 |
| | | Effort (off-task thoughts) | | | .005 | -.07 | -1.86 (535) | .06 |
| | | Effort (time spent) | | | .000 | -.01 | -.32 (535) | .75 |
| Lee et al. (2011) | | | | | | | | |
| Sample 1 (German) | Fluid/crystallized intelligence | Achievement | Job performance | 57 | .042 | -.21 | -1.59 (53) | .12 |
| | | | Training performance | 141 | .019 | .14 | 1.64 (137) | .10 |
| Sample 2 (French) | GMA | Achievement | Job performance | 58 | .000 | -.04 | -.03 (54) | .97 |
| | | | Training performance | 68 | .001 | .04 | .29 (64) | .77 |
| LePine and Van Dyne (2001) | GMA | Achievement | Lab performance (task) | 276 | .001 | -.03 | -.50 (272) | .62 |
| | | | Lab performance (contextual) | | .000 | .01 | .20 (272) | .84 |
| Marcus et al. (2007) | GMA | Achievement | Job performance | 80 | .012 | .11 | 1.03 (76) | .31 |
| Motowidlo and Van Scotter (1994) | GMA | Work motivation | Task performance | 224 | .015 | .13 | 1.91 (220) | .06 |
| | | | Contextual performance | 232 | .030 | .18 | 2.88 (228) | |
| Mount et al. (2008) | GMA | Achievement | Task performance | 161 | .013 | -.12 | -1.49 (157) | .14 |
| | | | Contextual performance | | .002 | -.05 | -.56 (157) | .57 |
| Mussel (2013) | GMA | Achievement | Job performance | 224 | .005 | -.07 | -1.20 (220) | .23 |
| | | | Training performance | 190 | .004 | -.07 | 1.01 (186) | .31 |
| Mussel et al. (2011) | GMA | Achievement | Job performance | 218 | .022 | .15 | 2.26 (214) | .03 |
| | | | Training performance | 227 | .006 | .08 | 1.20 (223) | .23 |
| Ono et al. (2011) | GMA | Achievement | Job performance | 37 | .103 | -.33 | -2.23 (33) | .03 |
| | | | Training performance | 131 | .005 | -.08 | -.86 (127) | .39 |
| Perry et al. (2010) | | | | | | | | |
| Study 1 | Critical reasoning | Achievement | Job performance | 208 | .053 | .24 | 3.45 (204) | .00 |
| Study 2 | Quantitative ability | Achievement | Job performance | 218 | .019 | .14 | 2.10 (214) | .04 |
| Ployhart (1999) | GMA | Achievement | Job performance | 134 | .001 | .02 | .28 (130) | .78 |
| Robinson (2009) | GMA | Effort (on task) | Lab performance | 160 | .000 | -.02 | -.22 (156) | .83 |
| | | Effort (time spent) | | | .000 | -.01 | -.14 (156) | .89 |
| Robson et al. (2010) | GMA | Diligence | Job performance | 132 | .005 | .07 | .78 (128) | .44 |
| Savoy (2004) | GMA | Achievement | Job performance | 63 | .003 | .06 | .59 (5) | |
| Schmidt (2008) | GMA | Effort | Lab performance | 247 | .004 | .06 | 1.11 (243) | .27 |
| | | Goal commitment | | | .006 | -.08 | -1.38 (243) | .17 |
| Seijts and Crim (2009) | GMA | Goal commitment | Lab performance | 104 | .007 | .09 | .91 (100) | .36 |
| Stanhope et al. (2013) | GMA | Training motivation | Training performance | 339 | .007 | .09 | 1.57 (335) | .12 |
| Tolli (2009) | GMA | Effort | Lab performance | 166 | .000 | -.00 | -.02 (162) | .98 |
| | | Goals | | | .001 | -.03 | -.39 (162) | .70 |
| | | Goal commitment | | | .005 | .07 | .96 (162) | .34 |

Table 2 (continued)

| Study | Ability | Motivation | Performance | <i>N</i> | ΔR^2 | β | <i>t</i> (<i>df</i>) | <i>p</i> |
|-----------------------------|---------|---------------------|----------------------|----------|--------------|---------|------------------------|----------|
| Van Iddekinge et al. (2009) | GMA | Work motivation | Job performance | 454 | .004 | -.07 | -1.45 (450) | .15 |
| Vasilopoulos et al. (2007) | GMA | Achievement | Training performance | 1,008 | .000 | .01 | .41 (1,004) | .68 |
| Watson and Surface (2012) | GMA | Training motivation | Training performance | 550 | .000 | -.01 | -.25 (546) | .81 |
| Weilbaecher (2000) | GMA | Work drive | Job performance | 129 | .002 | .04 | .57 (125) | .57 |
| Wright et al. (1995) | GMA | Achievement | Job performance | 203 | .090 | n/a | n/a | .01 |
| Yeo and Neal (2004) | GMA | Achievement | Lab performance | 99 | .000 | .01 | .15 (95) | .88 |
| Ziegler and Bensch (2014) | | | | | | | | |
| Sample 1 | GMA | Achievement | Training performance | 148 | .000 | .00 | -.01 (144) | 1.00 |
| Sample 2 | GMA | | | 117 | .007 | .08 | .98 (113) | .33 |
| Sample 3 | GMA | | | 138 | .027 | .17 | 1.98 (134) | .05 |
| Sample 4 | GMA | | | 98 | .027 | -.17 | 1.62 (94) | .11 |

Note. ΔR^2 = change in the squared multiple correlation when the ability \times motivation interaction was added to a model that also included the first-order effects of ability and motivation. β = standardized regression coefficient for the interaction term. GMA = general mental ability.

attribute 100% of the explained variance to the interaction effect. However, MMR revealed that the interaction explained very little variance in both models ($\Delta R^2 = .002$ and $.02$, respectively).

Summary and Recommendations

Vancouver et al. challenged the MMR approach to testing interactions. They also questioned conclusions of Van Iddekinge et al. whose meta-analytic comparison of the additive and interactive effects of *A* and *M* on *P* was based, in part, on MMR. We thank Vancouver et al. for their efforts to ensure researchers use appropriate methods to test for interactions. However, we identified conceptual, analytical, and empirical problems with their arguments. First, $P = A \times M$ was not “settled theory” prior to Van Iddekinge et al. Rather, evidence for this interaction was mixed and ambiguous. Van Iddekinge et al.’s meta-analysis helped resolve this issue by showing that, in most cases, the interaction between *A* and *M* tends to provide very little additional prediction beyond their additive effects. Moreover, we applied the approach Vancouver et al. recommended to Van Iddekinge et al.’s data and found that the $A \times M$ interaction was statistically significant in only about 10% of cases. Thus, although some theories suggest that *A* and *M* interact to predict *P*, empirical evidence suggests this does not tend to be the case.

So why in most cases do *A* and *M* demonstrate additive effects on *P* but not interactive effects? Vancouver et al. suggested that the lack of support for $A \times M$ is because MMR is a flawed method for assessing interactions. However, we demonstrated conceptually and empirically that this is not the case. Another possibility is that theories that propose this interaction are incorrect or that the interaction occurs only in certain situations (e.g., see Van Iddekinge et al.’s moderator analyses for the few instances in which the interaction effect was somewhat stronger). Alternatively, perhaps theories that hypothesize $A \times M$ are correct, but empirical tests of the hypothesis are the problem. For example, as Van Iddekinge et al. noted, tests of $A \times M$ often have been based on questionable measures of *A* and/or *M*. And although they took considerable steps to screen the measures included in their meta-analysis, many primary studies used measures that may not fully assess the direction, intensity, or persistence of effort, which is the essence of *M* (e.g., Campbell, 1990; Kanfer, 1990). In addition, study design

factors such as scale coarseness, range restriction, and measurement error can reduce the magnitude of observed interaction effects (e.g., Aguinis & Stone-Romero, 1997). However, Van Iddekinge et al.’s results do not support this explanation regarding the effects of range restriction and measurement error because correcting for these artifacts did not increase support for $A \times M$.

Second, we also described problems in Vancouver et al.’s characterization of research on testing interaction effects more generally. The concerns Vancouver et al. raised about MMR are not new and have been successfully refuted many times (e.g., Arnold & Evans, 1979; Cohen, 1978; Cronbach, 1987). Moreover, Vancouver et al. frequently cited Arnold and Evans to support their arguments when these researchers were, in fact, strong advocates of MMR.

Third, Vancouver et al.’s conceptual arguments for $A \times M$ interactions are based on a thought experiment (involving “The Rock” and “Dr. Cooper”) in which ability or motivation are zero, a scenario that does not exist in real organizations staffed by real people. Furthermore, the only way a purely multiplicative model would be possible is if *X* and *Z* are uncorrelated with *Y* (i.e., a disordinal interaction). This would be a very uncommon scenario in applied psychology, including the case of *A* and *M*, which are known to have substantial, positive correlations with *P*.

Fourth, regarding Vancouver et al.’s analytical arguments, we showed how MMR correctly partitions variance between the lower-order effects and interaction effect. We also explained how Arnold and Evans’ (1979) work supports, rather than undermines, the appropriateness of using MMR to test interactions.

Finally, Vancouver et al.’s empirical arguments against MMR are based on a simulation in which they used the unresidualized cross product (*XZ*) to create data when they should have used the residualized cross product (XZ_{res}), which is independent of the lower-order effects and, by definition, represents the pure interaction effect. After correcting this error, we found that MMR accurately estimated both the statistical significance and magnitude of the interaction effect. We also noted limitations of Vancouver et al.’s suggested alternative to MMR, including a singular focus on the statistical significance of regression weights and the lack of an effect size estimate for the interaction.

We conclude by summarizing the approach we recommend to test and interpret interaction effects. First, address study design factors

identified as being critical for accurate tests of interactions (e.g., scale coarseness, range restriction, measurement error, sample size, and resulting statistical power). Of course, it also is critical to use valid measures of X , Z , and Y so that valid inferences can be made regarding the effects. Second, conduct an MMR analysis and use the p value for ΔR^2 (i.e., the F test) to assess the statistical significance of the interaction and the size of ΔR^2 to assess the effect size of the interaction. But, because MMR is known for its low statistical power (Aguinis, 1995), significance testing should only be conducted in the presence of sufficient power so that existing population effects are not incorrectly ignored in the sample. In addition, residualized relative weight analysis (LeBreton et al., 2013) can be used to understand the relative importance of interaction and first-order effects. Third, examine whether the form of the interaction is consistent with the hypothesized form (Gardner et al., 2017) using tools such as simple slopes analysis (e.g., O'Connor, 1998). Researchers also may assess the regions of significance of the interaction effect (e.g., Finsaas & Goldstein, 2020) to ensure it does not apply only to extreme cases. Finally, consider the significance of the interaction for theory and practice. For example, Aguinis et al. (2010) described qualitative methods that could be used to assess practical significance, such as focus groups with decision makers to understand the "bottom-line" impact of interaction effects.

Conclusion

We read the study by Vancouver et al. with an open mind, and we carefully considered their concerns about MMR and the alternative approach for testing interactions they suggested. However, we identified conceptual, analytical, and empirical problems with their arguments and evidence. Perhaps most importantly, when we addressed a key error in their empirical evidence (i.e., simulating the cross product rather than the interaction effect), we found that MMR (and the ΔR and ΔR^2 it yields) accurately assessed both the statistical significance and magnitude of interaction effects. Overall, our analysis of Vancouver et al. reaffirms Van Iddekinge et al.'s conclusion that the interaction between ability and motivation contributes little to the prediction of performance beyond the additive effects of the two variables. Our analysis also reaffirms that MMR is an appropriate tool for testing the presence and magnitude of interaction effects.

References

- Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, 21(6), 1141–1158. [https://doi.org/10.1016/0149-2063\(95\)90026-8](https://doi.org/10.1016/0149-2063(95)90026-8)
- Aguinis, H. (2004). *Regression analysis for categorical moderators*. 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(1), 94–107. <https://doi.org/10.1037/0021-9010.90.1.94>
- Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*, 20(4), 665–685. <https://doi.org/10.1177/1094428115627498>
- Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, 82(1), 192–206. <https://doi.org/10.1037/0021-9010.82.1.192>
- Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, 13(3), 515–539. <https://doi.org/10.1177/1094428109333339>
- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage Publications.
- Althaus, R. P. (1971). Multicollinearity and non-additive regression models. In H. M. Blalock (Ed.), *Causal models in the social sciences* (pp. 453–472). Atherton.
- Arnold, H. J. (1982). Moderator variables: A clarification of conceptual, analytic, and psychometric issues. *Organizational Behavior and Human Performance*, 29(2), 143–174. [https://doi.org/10.1016/0030-5073\(82\)90254-9](https://doi.org/10.1016/0030-5073(82)90254-9)
- Arnold, H. J., & Evans, M. G. (1979). Testing multiplicative models does not require ratio scales. *Organizational Behavior & Human Performance*, 24(1), 41–59. [https://doi.org/10.1016/0030-5073\(79\)90014-X](https://doi.org/10.1016/0030-5073(79)90014-X)
- Barros, E., Kausel, E. E., Cuadra, F., & Diaz, D. A. (2014). Using general mental ability and personality traits to predict job performance in three Chilean organizations. *International Journal of Selection and Assessment*, 22(4), 432–438. <https://doi.org/10.1111/ijsa.12089>
- Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, 55(2), 267–306. <https://doi.org/10.1111/j.1744-6570.2002.tb00111.x>
- Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, 93(2), 296–316. <https://doi.org/10.1037/0021-9010.93.2.296>
- Bono, J. E. (2001). *Self-determination at work: The motivational effects of transformational leadership* [Unpublished doctoral dissertation]. University of Iowa.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687–732). Consulting Psychologists Press.
- Carretta, T. R., Teachout, M. S., Ree, M. J., Barto, E. L., King, R. E., & Michaels, C. F. (2014). Consistency of the relations of cognitive ability and personality traits to pilot training performance. *The International Journal of Aviation Psychology*, 24(4), 247–264. <https://doi.org/10.1080/10508414.2014.949200>
- Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70(6), 426–443. <https://doi.org/10.1037/h0026714>
- Cohen, J. (1978). Partialled products are interactions: Partialled powers are curve components. *Psychological Bulletin*, 85(4), 858–866. <https://doi.org/10.1037/0033-2909.85.4.858>
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997–1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum.
- Corker, K. S. (2012). Enhancing achievement through intervention: How conscientiousness and cognitive ability impact responses to goal setting and implementation intentions interventions (Order No. AAI3512848). *Dissertation Abstracts International: Section B: The Sciences and Engineering*.
- Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management*, 19(4), 915–922. <https://doi.org/10.1177/014920639301900411>
- Cortina, J. M., & DeShon, R. P. (1998). Determining relative importance of predictors with the observational design. *Journal of Applied Psychology*, 83(5), 798–804. <https://doi.org/10.1037/0021-9010.83.5.798>
- Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin*, 102(3), 414–417. <https://doi.org/10.1037/0033-2909.102.3.414>

- Dachler, H. P., & Mobley, W. H. (1973). Construct validation of an instrumentality–expectancy–task-goal model of work motivation: Some theoretical boundary conditions. *Journal of Applied Psychology, 58*(3), 397–418. <https://doi.org/10.1037/h0036297>
- Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods, 15*(3), 339–362. <https://doi.org/10.1177/1094428111430540>
- Darrow, A. L., & Kahl, D. R. (1982). A comparison of moderated regression techniques considering strength of effect. *Journal of Management, 8*(2), 35–47. <https://doi.org/10.1177/014920638200800203>
- Evans, M. G. (1985). A monte carlo study of the effects of correlated method variance in moderated multiple regression. *Organizational Behavior and Human Decision Processes, 36*(3), 305–323.
- Evans, M. G. (1991). The problem of analyzing multiplicative composites: Interactions revisited. *American Psychologist, 46*(1), 6–15. <https://doi.org/10.1037/0003-066X.46.1.6>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Finsaas, M. C., & Goldstein, B. L. (2020). Do simple slopes follow-up tests lead us astray? Advancements in the visualization and reporting of interactions. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000266>
- Fleishman, E. A. (1958). A relationship between incentive motivation and ability level in psychomotor performance. *Journal of Experimental Psychology, 56*(1), 78–81. <https://doi.org/10.1037/h0044885>
- Friedrich, R. I. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science, 26*(4), 797–833.
- Gardner, R. G., Harris, T. B., Li, N., Kirkman, B. L., & Mathieu, J. E. (2017). Understanding “it depends” in organizational research: A theory-based taxonomy, review, and future research agenda concerning interactive and quadratic relationships. *Organizational Research Methods, 20*(4), 610–638. <https://doi.org/10.1177/1094428117708856>
- Hausdorf, P. A., & Risavy, S. D. (2015). Predicting training and job performance for transit operators. *International Journal of Selection and Assessment, 23*(2), 191–195. <https://doi.org/10.1111/ijssa.12107>
- Huang, L. (2012). *A multifocal framework of personality applied to training and transfer* [Unpublished doctoral dissertation]. Michigan State University.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72–98. <https://doi.org/10.1037/0033-2909.96.1.72>
- Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2012). Emotional intelligence in personnel selection: Applicant reactions, criterion, and incremental validity. *International Journal of Selection and Assessment, 20*(3), 347–358. <https://doi.org/10.1111/j.1468-2389.2012.00605.x>
- Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 75–170). Consulting Psychologists Press.
- Kluemper, D. H., DeGroot, T., & Choi, S. (2013). Emotion management ability: Predicting task performance, citizenship, and deviance. *Journal of Management, 39*(4), 878–905. <https://doi.org/10.1177/0149206311407326>
- Kozlowski, S. W. J., & Bell, B. S. (2006). Disentangling achievement orientation and goal setting: Effects on self-regulatory processes. *Journal of Applied Psychology, 91*(4), 900–916. <https://doi.org/10.1037/0021-9010.91.4.900>
- LeBreton, J. M., Tonidandel, S., & Krasikova, D. V. (2013). Residualized relative importance analysis: A technique for the comprehensive decomposition of variance in higher order regression models. *Organizational Research Methods, 16*(3), 449–473. <https://doi.org/10.1177/1094428113481065>
- Lee, Y.-T., Stettler, A., & Antonakis, J. (2011). Incremental validity and indirect effect of ethical development on work performance. *Personality and Individual Differences, 50*(7), 1110–1115. <https://doi.org/10.1016/j.paid.2011.01.036>
- LePine, J. A., & Van Dyne, L. V. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with Big Five personality characteristics and cognitive ability. *Journal of Applied Psychology, 86*(2), 326–336. <https://doi.org/10.1037/0021-9010.86.2.326>
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Prentice Hall.
- Maier, N. R. F. (1955). *Psychology in industry: A psychological approach to industrial problems*. Houghton-Mifflin.
- Marcus, B., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance, 20*(3), 275–285. <https://doi.org/10.1080/08959280701333362>
- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*(2), 376–390. <https://doi.org/10.1037/0033-2909.114.2.376>
- McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*(4), 493–505. <https://doi.org/10.1037/0021-9010.79.4.493>
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*(2), 335–354. <https://doi.org/10.1111/j.1744-6570.1990.tb01562.x>
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*(4), 475–480. <https://doi.org/10.1037/0021-9010.79.4.475>
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management, 25*(5), 707–721. <https://doi.org/10.1177/014920639902500505>
- Mount, M. K., Oh, I.-S., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology, 61*(1), 113–139. <https://doi.org/10.1111/j.1744-6570.2008.00107.x>
- Murphy, K. R., & Russell, C. J. (2017). Mend it or end it: Redirecting the search for interactions in the organizational sciences. *Organizational Research Methods, 20*(4), 549–573. <https://doi.org/10.1177/1094428115625322>
- Mussel, P. (2013). Introducing the construct curiosity for predicting job performance. *Journal of Organizational Behavior, 34*(4), 453–472. <https://doi.org/10.1002/job.1809>
- Mussel, P., Spengler, M., Litman, J., & Schuler, H. (2011). Development and validation of the German work-related curiosity scale. *European Journal of Psychological Assessment, 28*(2), 109–117. <https://doi.org/10.1027/1015-5759/a000098>
- O’Boyle, E., Banks, G. C., Carter, K., Walter, S., & Yuan, Z. (2019). A 20-year review of outcome reporting bias in moderated multiple regression. *Journal of Business and Psychology, 34*, 19–37. <https://doi.org/10.1007/s10869-018-9539-8>
- O’Connor, B. P. (1998). SIMPLE: All-in-one programs for exploring interactions in moderated multiple regression. *Educational and Psychological Measurement, 58*(5), 836–840. <https://doi.org/10.1177/0013164498058005009>
- Ono, M., Sachau, D. A., Deal, W. P., Englert, D. R., & Taylor, M. D. (2011). Cognitive ability, emotional intelligence, and the Big Five personality dimensions as predictors of criminal investigator performance. *Criminal Justice and Behavior, 38*(5), 471–491. <https://doi.org/10.1177/0093854811399406>

- Perry, S. J., Hunter, E. M., Witt, L. A., & Harris, K. J. (2010). $P = f$ (conscientiousness \times ability): Examining the facets of conscientiousness. *Human Performance, 23*(4), 343–360. <https://doi.org/10.1080/08959285.2010.501045>
- Ployhart, R. E. (1999). An interactionist approach to assessing personality in work contexts: Construct validation of a predictor of customer service performance (Order No. AAI9948157). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 5246.
- Ployhart, R. E., & Hakel, M. D. (1998). The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance. *Personnel Psychology, 51*(4), 859–901. <https://doi.org/10.1111/j.1744-6570.1998.tb00744.x>
- Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement, 21*(4), 291–305. <https://doi.org/10.1177/01466216970214001>
- Robinson, R. P. (2009). The effect of individual differences on training process variables in a multistage computer-based training context (Order No. AAI3361738). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 3816.
- Robson, S. M., Abraham, J. D., & Weiner, J. (2010). Characteristics of successful direct support professionals: An examination of personality and cognitive ability requirements. *International Journal of Selection and Assessment, 18*(2), 215–219. <https://doi.org/10.1111/j.1468-2389.2010.00503.x>
- Sackett, P. R., Gruys, M. L., & Ellingson, J. E. (1998). Ability–personality interactions when predicting job performance. *Journal of Applied Psychology, 83*(4), 545–556. <https://doi.org/10.1037/0021-9010.83.4.545>
- Savoy, P. J. (2004). *Development and validation of a measure of self-directed learning competency* [Unpublished doctoral dissertation]. Kent State University.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*(2), 115–129. <https://doi.org/10.1037/1082-989X.1.2.115>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>
- Schmidt, G. B. (2008). *Learning and performance goals: Disentangling the effects of goal specificity* [Unpublished master's thesis]. Michigan State University.
- Seijts, G., & Crim, D. (2009). The combined effects of goal type and cognitive ability on performance. *Motivation and Emotion, 33*, 343–352. <https://doi.org/10.1007/s11031-009-9143-3>
- Shieh, G. (2008). Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient. *Organizational Research Methods, 11*(2), 387–407. <https://doi.org/10.1177/1094428106292901>
- Sockloff, A. L. (1976). Spurious product correlation. *Educational and Psychological Measurement, 36*(1), 33–44. <https://doi.org/10.1177/001316447603600104>
- Stanhope, D. S., Pond, S. B., & Surface, E. A. (2013). Core self-evaluations and training effectiveness: Prediction through motivational intervening mechanisms. *Journal of Applied Psychology, 98*(5), 820–831. <https://doi.org/10.1037/a0032599>
- Stone, E. F., & Hollenbeck, J. R. (1984). Some issues associated with the use of moderated regression. *Organizational Behavior & Human Performance, 34*(2), 195–213. [https://doi.org/10.1016/0030-5073\(84\)90003-5](https://doi.org/10.1016/0030-5073(84)90003-5)
- Terborg, J. R. (1977). Validation and extension of an individual differences model of work performance. *Organizational Behavior & Human Performance, 18*(1), 188–216. [https://doi.org/10.1016/0030-5073\(77\)90028-9](https://doi.org/10.1016/0030-5073(77)90028-9)
- Tolli, A. P. (2009). *Motivational and self-regulatory responses to interruptions* [Unpublished doctoral dissertation]. University of Akron.
- Van Eerde, W., & Thierry, H. (1996). Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of Applied Psychology, 81*(5), 575–586. <https://doi.org/10.1037/0021-9010.81.5.575>
- Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management, 44*(1), 249–279. <https://doi.org/10.1177/0149206317702220>
- Van Iddekinge, C. H., Ferris, G. R., & Heffner, T. S. (2009). Test of a multistage model of distal and proximal antecedents of leader performance. *Personnel Psychology, 62*(3), 463–495. <https://doi.org/10.1111/j.1744-6570.2009.01145.x>
- Vancouver, J., Carlson, B. W., Dhanani, L. Y., & Colton, C. E. (2021). Interpreting moderated multiple regression: A comment on Van Iddekinge, Aguinis, Mackey, and DeOrtentiis (2018). *Journal of Applied Psychology, 106*(3), 467–475. <https://doi.org/10.1037/apl0000522>
- Vasilopoulos, N. L., Cucina, J. M., & Hunter, A. E. (2007). Personality and training proficiency: Issues of bandwidth-fidelity and curvilinearity. *Journal of Occupational and Organizational Psychology, 80*(1), 109–131. <https://doi.org/10.1348/096317906X102114>
- Vroom, V. H. (1964). *Work and motivation*. Wiley.
- Watson, A. M., & Surface, E. A. (2012). *Trainee characteristics and achievement during special operations forces initial acquisition foreign language training*. SWA Consulting.
- Weilbaecher, A. D. (2000). Self-report personality measures of fake good in employment selection (Order No. AAI9985678). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 5039.
- Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annals of Mathematical Statistics, 2*(4), 440–457.
- Wright, P. M. (1990). Operationalization of goal difficulty as a moderator of the goal difficulty–performance relationship. *Journal of Applied Psychology, 75*(3), 227–234. <https://doi.org/10.1037/0021-9010.75.3.227>
- Wright, P. M., Kacmar, K. M., McMahan, G. C., & Deleeuw, K. (1995). $P=f(M \times A)$: Cognitive ability as a moderator of the relationship between personality and job performance. *Journal of Management, 21*(6), 1129–1139. [https://doi.org/10.1016/0149-2063\(95\)90025-X](https://doi.org/10.1016/0149-2063(95)90025-X)
- Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology, 89*(2), 231–247. <https://doi.org/10.1037/0021-9010.89.2.231>
- Yin, P., & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytical methods. *Journal of Experimental Education, 69*(2), 203–224. <https://doi.org/10.1080/00220970109600656>
- Zedeck, S. (1971). Problems with the use of “moderator” variables. *Psychological Bulletin, 76*(4), 295–310. <https://doi.org/10.1037/h0031543>
- Ziegler, M., Bensch, D., Maaß, U., Schult, V., Vogel, M., & Bühner, M. (2014). Big Five facets as predictor of job training performance: The role of specific job demands. *Learning and Individual Differences, 29*, 1–7. <https://doi.org/10.1016/j.lindif.2013.10.008>

Received June 15, 2020

Revision received December 7, 2020

Accepted December 8, 2020 ■