Journal of Management
Vol. 48 No. 8, November 2022 2159–2166
DOI: 10.1177/01492063221088516


# Reporting Interaction Effects: Visualization, Effect Size, and Interpretation

Kevin R. Murphy
*The University of Limerick*
Herman Aguinis
*The George Washington University*

*Most management theories include hypotheses about interaction effects (i.e., the relation between two variables depends on values of another), but it is common for articles to present results that make it difficult to evaluate the nature, strength, and importance of the effect. We offer recommendations for improving the reporting of interaction effects by focusing on (a) visualizations, (b) effect size estimates, and (c) assessments of the nature, meaning, and importance of interactions for theory and practice.*

**Keywords:** *moderator; moderating effect; contingency effects; boundary conditions*

One of the most common hypotheses in management research is that the relation between some pair of variables ($x$ and $y$) is conditioned by, contingent upon, or influenced by some third variable ($z$) (Aguinis, Edwards, & Bradley, 2017)—in other words, that $x$ and $z$ have an interactive effect on $y$. Some examples include the interaction effect of different human resources (HR) practices on HR system effectiveness (Boon, Den Hartog, & Lepak, 2019), interaction effects of individual-level and organizational-level factors on crises and the crisis management process (Bundy, Pfarrer, Short, & Coombs, 2017), interaction effect of discrepancy of social performance and foreign exposure on ratio of visuals to donation events (Wang, Jia, Xiang, & Lan, 2021), and the interaction effect of macro- and microlevel antecedents on social entrepreneurship (Saebi, Foss, & Linder, 2019).

We provide evidence that studies are frequently reported in ways that make it difficult for researchers, evaluators of research (e.g., journal editors and reviewers), and consumers of research (e.g., other researchers, organizational leaders, policy makers) to draw sensible

*Corresponding Author: Herman Aguinis, Department of Management, The George Washington University, Funger Hall, Suite 311, 2201 G Street NW, Washington, DC 20052, USA.*
*E-mail: haguinis@gwu.edu*

conclusions about the nature, strength, and importance of interaction effects. The problems we identified have a clear and direct detrimental impact on theory and practice, and we offer simple and practical solutions. The studies we discuss generally use moderated multiple regression, tests of differences in simple slopes, or some variant to test for interaction effects.

## Reports of Interaction Effects Are Often Inconsistent and Incomplete

We reviewed every study published in *Journal of Management* (*JOM*) in 2020 and examined the ways interaction effects were presented and described. In addition, as a baseline and for comparison purposes, we examined all 2020 articles in *Academy of Management Journal* (*AMJ*), *Journal of Applied Psychology* (*JAP*), and *Strategic Management Journal* (*SMJ*). Taken together, these four journals published 96 papers in 2020 that analyzed single-level and primary data to test interaction hypotheses (15 in *JOM*, 24 in *AMJ*, 30 in *JAP*, and 27 in *SMJ*).

We found that interaction effects were rarely reported in a way that would make their nature and strength fully interpretable. For example, it was common to omit graphs or to present graphs that were potentially confusing because they focused on only a small portion of the outcome variable scale. In addition, it was common to omit interpretable information about the strength of interaction effects. We reviewed four prestigious and highly visible journals, so it is likely that the problems we identified are also present in many other journals as well.

### *Visualizations Often Exaggerate the Size of Interaction Effects*

In principle, visualizations should help readers understand and interpret the findings that they illustrate. But, the visualizations used to illustrate interactions did not provide this sort of help. Over two thirds of the papers that presented graphs describing interactions (67.5%) provided incomplete visualizations of the interaction, and this approach often inadvertently magnified the apparent size and importance of the interaction effect by truncating the *y*-axis. For example, this truncation from a 7-point range for the *y*-scale to a 2-point range for the *y*-axis in a figure causes weak interaction effects to look stronger than they are.

Truncation of the *y*-axis is common, but the extent to which truncation occurs varies widely. In some papers, authors used a *y*-axis that covered less than 10% of the range of *y* values, and in others, authors used a *y*-axis that covered over 80% of this range. Another 16.6% of the papers that tested interaction hypotheses did not present an interaction figure at all. Overall, about only 25% of the papers we reviewed included a figure that would make it reasonably easy for readers to accurately assess the strength of the interaction effect.

To demonstrate problems associated with the practice of using a truncated *y*-axis, consider one of the articles included in our review, which examined the interaction effect between perspective taking and self-efficacy on feedback seeking (Sherf & Morrison, 2020). This exemplary study included easily interpretable figures. In Study 1, these authors reported a Self-Efficacy × Perspective Taking interaction, plotting the relation between self-efficacy and feedback seeking at −1 standard deviation (low) and +1 standard deviation (high) from the perspective-taking mean, using a *y*-axis that spanned the full range of possible values (i.e., 7 points) for their dependent variable, feedback seeking.

We use the data from Sherf and Morrison (2020) to illustrate what would have happened if instead these authors had truncated the range of values for the dependent variable in their graph as done in the vast majority of published articles as documented by our review. The top panel of Figure 1 uses the entire range of values of $y$, closely mirroring the figure presented by Sherf and Morrison (2020). In Figure 1's center panel, we plotted this same interaction using a shorter range of points for the $y$-axis of 1.5 to 4.5 instead of the full range of 1 to 7, resulting in an apparent larger difference between the two slopes. In the bottom panel of Figure 1, we again plotted this same interaction but using an even smaller range of 2 points for the $y$-axis (i.e., 1.8 to 3.8), resulting in an apparent even more impressive difference between the slopes. The examples of $y$-axis truncation in Figure 1's center and bottom panels are realistic and representative based on results of our review and show how truncation of the $y$-axis can dramatically magnify the apparent strength of the interaction effect.

In sum, routinely presenting graphs in a way that magnifies their apparent strength and importance results in theory and practice derailment; researchers as well as managers and policy makers can devote time and effort to the pursuit of dead-end interaction hypotheses, believing that they are chasing after interesting and exciting interaction possibilities. An incautious reader who was presented with the center or bottom panels of Figure 1 could easily come away with the impression that the interaction effect is a strong one.

### *Effect Size Estimates Are Often Missing*

Although every article included in our review reported the outcome of a significance test (i.e., $p$ values), 69.7% failed to report an interpretable measure of the strength of the interaction effect. For example, $\Delta R^2$ represents the proportion of variance explained by the interaction effect (i.e., $x*z$) above the variance explained by first-order effects (i.e., $x$ and $z$). There have been some discussions in the literature regarding the best effect size measure for interaction effects (e.g., Liu & Yuan, 2021; Murphy & Russell, 2017; Van Iddekinge, Aguinis, LeBreton, Mackey, & DeOrtentiis, 2021), and different research teams might make different choices regarding effect size measures. Our concern here is not that some authors choose the wrong effect size measure but rather that they rarely report any measure of the strength of interaction effects.

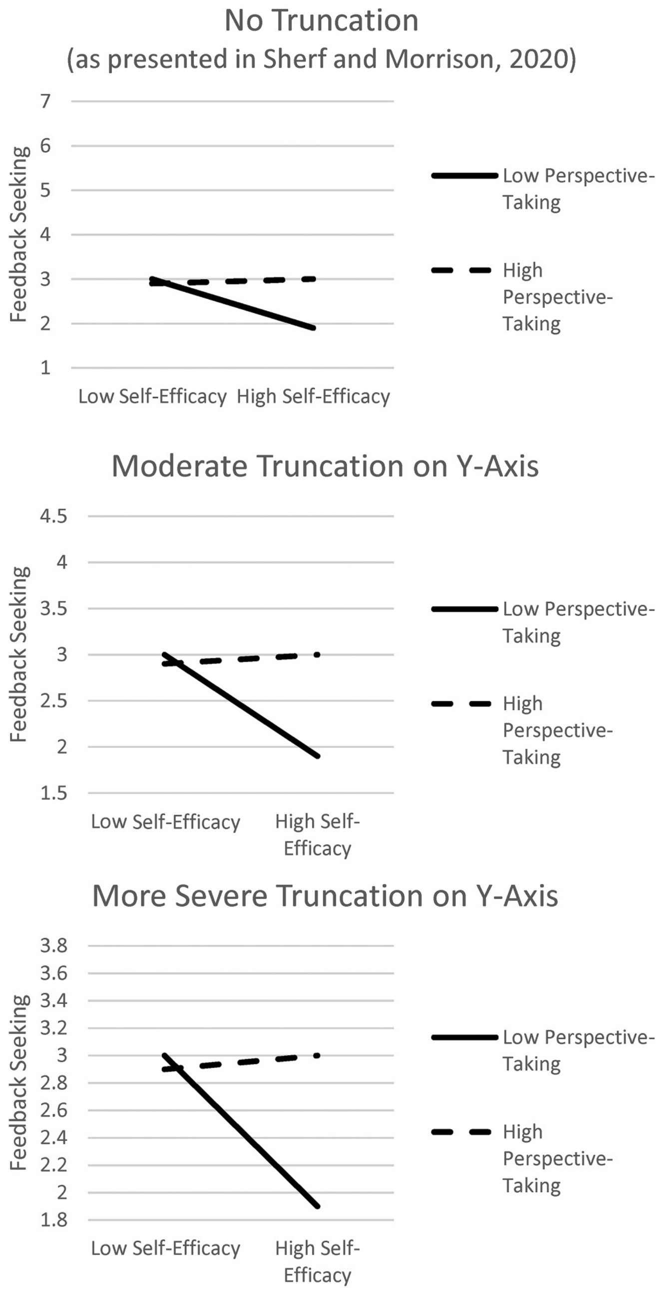
The failure to include an interpretable measure of the strength of interaction effects might not be a serious problem if articles routinely included figures that clearly showed the strength and nature of the interaction effect. But, as documented by our review, this is not the case. In fact, only five of the 96 articles we reviewed (5.2%) paired interpretable effect size measures, such as $\Delta R^2$, with $y$-axes that span the full range of possible values for $y$. Given the centrality of interaction effects for so many theories, these results show that we are not doing a good job of conveying the nature and strength of interaction effects in our research. Accordingly, we lay out three concrete recommendations for improving the way interactions are presented, described, and interpreted.

## Recommendations for Reporting Interactions

### *Recommendation 1: Visualizations*

We recommend that all articles that include interaction hypothesis tests present figures spanning the entire range of possible values for the $y$ scale. This common frame of reference

**Figure 1**
**Three Visualizations of the Same Interaction Effect**

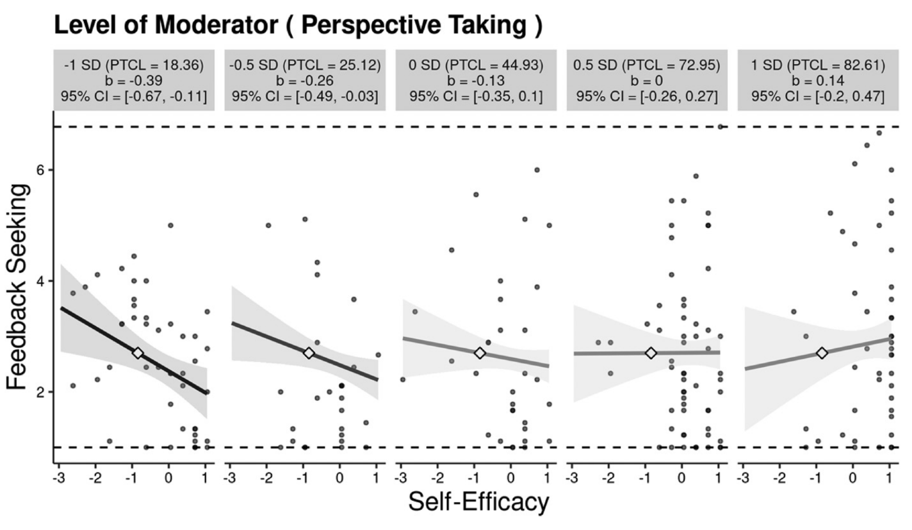


will not exaggerate the size of interaction effects and make it easier to accurately compare plots across studies even if they use different scales of measurement for *y*.

In addition, McCabe, Kim, and King (2018) developed a useful R Shiny app, available at https://connorjmccabe.shinyapps.io/interactive, that makes it easy to display interaction effects across a range of values of *z* rather than focusing solely on extreme values, such as the usual practice of plotting only one standard deviation below and one standard deviation above the mean on *z* (i.e., roughly the 15th and 85th percentiles when the distribution is approximately normal). To illustrate the usefulness of examining more than just two values of *z* when examining interaction effects, we used the R Shiny app to reexamine the results presented by Sherf and Morrison (2020), whose simple-slope graph is reproduced in the top panel of Figure 1. Figure 2 presents graphs of the same data, plotting the *x*-*y* slope when *z* is one standard deviation below the mean, one half standard deviation below the mean, at the mean, one half standard deviation above the mean, and one standard deviation above the mean. The panels in Figure 2 also include precise percentile (i.e., PCTL) information for each of these values as well as information about the relative size of errors in prediction. For example, error is considerably larger when *z* is one standard deviation above the mean than when it is one standard deviation below the mean.

Figure 2 tells a more informative story about the way perspective taking influences the relation between self-efficacy and feedback seeking than the graphs shown in Figure 1 solely based on one standard deviation below and one standard deviation above the mean of *z*. For example, for below-average levels of self-efficacy, perspective taking clearly has some effect. But, the regression lines at the mean, one half of an standard deviation above the mean, and one standard deviation above the mean are virtually flat and also virtually indistinguishable (especially given the growing levels of error as perspective taking increases). In

**Figure 2**
**Five-Panel Display of Sherf and Morrison (2020) Interaction**



other words, the visualizations show that the interaction is not equally important across the entire range of values of self-efficacy.

### *Recommendation 2: Effect Size Estimates*

Many scientific fields are rapidly moving away from sole reliance on null hypothesis significance testing (Amrhein, Greenland, & McShane, 2019; Wasserstein, Schirm, & Lazar, 2019). Unfortunately, most of the articles we reviewed reported *p* values and regression coefficients or partial eta squares, but these measures do not provide a clear and standardized indication of the strength of the interaction effect. We therefore recommend that all studies that report an interaction provide clear, interpretable, and comparable (i.e., across studies) effect size estimates.

Kelley and Preacher (2012) and Liu and Yuan (2021) reviewed the literature on desirable properties for effect size measures and concluded that $R^2$-based measures have many of these properties. Specifically, the $\Delta R^2$ that is commonly reported in moderated multiple regression has two desirable properties (Van Iddekinge et al., 2021). First, it represents the incremental contribution of considering $x$ and $z$ jointly in a context where they have already been considered separately. Second, it can be interpreted in terms of the proportion of variance explained, a statistic that is widely used in all variations of the general linear model (Murphy, Myors, & Wolach, 2014). We recognize that other effect size measures have been proposed to index the strength of interactions (Liu & Yuan, 2021), and our intent is not to mandate that all authors employ the index we prefer. The choice of effect size measures is less important than the principle that every interaction report should present measures of the strength of that interaction.

### *Recommendation 3: Nature and Meaning of the Interaction*

Regarding the nature of interactions, these effects can be classified into one of three types (Gardner, Harris, Li, Kirkman, & Mathieu, 2017). First, an interaction has a *strengthening* effect when the relation between $x$ and $y$ becomes stronger (i.e., the slope becomes steeper) as $z$ increases. Second, an interaction has a *weakening* effect when the relation between $x$-$y$ dissipates (i.e., the slope becomes flatter) as $z$ increases. Third, an interaction has a *reversing* effect when the relation between $x$ and $y$ changes from positive (negative) to negative (positive) through the range of $z$.

It is common practice to examine the size and sign of coefficients included in a moderated regression equation (i.e., coefficients for $x$, $z$, and $x$*$z$) in an effort to understand the nature of interactions. We do not recommend this approach because a particular type of interaction effect can be present when there is the same or opposite directionality of the $z$-$y$ and $x$*$z$-$y$ relations. For example, a strengthening interaction is present when the coefficients for $x$, $z$, and $x$*$z$ are positive; when all three are negative; when the coefficient for $x$ and $x$*$z$ are positive and the coefficient for $z$ is negative; and when the coefficient for $z$ is positive and the coefficients for $z$ and $x$*$z$ are negative (Gardner et al., 2017). Accordingly, implementing our first recommendation regarding visualizations is also critical because it allows for the reporting of the specific nature of the interaction effect (strengthening, weakening, or reversing) without having to rely on ambiguous interpretations based on the signs of the coefficients. On a related note, when values of the predictors are dichotomous or limited to just a few, bar graphs are usually preferable to line graphs, but if these are appropriately and consistently scaled, they can provide the same sort of information as the more common line graphs.

Aguinis et al. (2010) argued that in addition to providing information about statistical significance and variance explained (i.e., $\Delta R^2$), authors should also provide information about the meaning of interactions for theory and practice. For example, one important concern in research on Aptitude × Treatment interactions is the need to determine the circumstances under which different educational interventions might be appropriate, depending on the student's aptitude level. This call to consider the theoretical and practical consequences of interactions seems to have had little impact; none of the articles we reviewed provided a clear argument for the meaningfulness of the interaction effects they examined. This type of practical argument might, for example, involve specifying for whom interactions matter (Murphy & Russell, 2017).

Techniques developed in the 1950s have proven useful for understanding the meaning of interaction effects by identifying the range of $z$ for which the $x$-$y$ slope is significantly different from zero (see Hayes & Matthes, 2009, for useful computational tools). Even when less formal methods are used to probe interactions (i.e., visualizations), there is often considerable value in asking whether interactions have meaningful effects for certain ranges of the data points in a sample—be it individuals, teams, or firms (Murphy & Russell, 2017). For example, Figure 2 suggests that perspective taking is relevant to feedback seeking only for that small subset of study participants who are half a standard deviation or more below the mean on self-efficacy. For virtually everyone else in this study, regression slopes are so similar and so flat that the predicted value of feedback seeking will be near the mean regardless of the perspective-taking score. Generally, when interaction effects are small, as is the norm in management and related fields (Aguinis, Beaty, Boik, & Pierce, 2005), their theoretical and practical significance will often be felt by only that subset of the sample that is at or near the extreme values for the moderator.

## Conclusions

Interactions are at the heart of most theories in management in both micro and macro domains (e.g., organizational behavior, HR management, strategy, entrepreneurship). Unfortunately, interactions are often presented in ways that make it difficult to accurately understand their nature and meaning. Worse yet, there appears to be bias in the way these effects are displayed, which could lead to the impression that weak interactions are strong ones. We recommend that future research testing interaction hypotheses include complete and consistent visualizations and understandable effect size measures, and attention be given to the nature and meaning of interactions. We hope our recommendations will result in more credible and useful research addressing interaction hypotheses in the future.

## ORCID iDs

Kevin R. Murphy https://orcid.org/0000-0002-4886-9499
Herman Aguinis https://orcid.org/0000-0002-3485-9484

## References

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., Edwards, J. R., & Bradley, K. J. 2017. Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods,* 20: 665-685.

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: 515-539.

Amrhein, V., Greenland, S., & McShane, B. 2019. Scientists rise up against statistical significance. *Nature,* 567: 305-307.

Boon, C., Den Hartog, D. N., & Lepak, D. P. 2019. A systematic review of human resource management systems and their measurement. *Journal of Management,* 45: 2498-2537.

Bundy, J., Pfarrer, M. D., Short, C. E., & Coombs, W. T. 2017. Crises and crisis management: Integration, interpretation, and research development. *Journal of Management,* 43: 1661-1692.

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: 610-638.

Hayes, A. F., & Matthes, J. 2009. Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods,* 41: 924-936.

Kelley, K., & Preacher, K. J. 2012. On effect size. *Psychological Methods,* 17: 137-152.

Liu, H., & Yuan, K. 2021. New measures of effect size in moderation analysis. *Psychological Methods,* 26: 680-700.

McCabe, C. J., Kim, D. S., & King, K. M. 2018. Improving present practices in the visual display of interactions. *Advances in Methods and Practices in Psychological Science,* 1: 147-165.

Murphy, K. R., Myors, B., & Wolach, A. 2014. *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (4th ed.). New York: Routledge.

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: 549-573.

Saebi, T., Foss, N. J., & Linder, S. 2019. Social entrepreneurship research: Past achievements and future promises. *Journal of Management,* 45: 70-95.

Sherf, E. N., & Morrison, E. W. 2020. I do not need feedback! Or do I? Self-efficacy, perspective taking, and feedback seeking. *Journal of Applied Psychology,* 105: 146-165.

Van Iddekinge, C. H., Aguinis, H., LeBreton, J. M., Mackey, J. D., & DeOrtentiis, P. S. 2021. Assessing and interpreting interaction effects: A reply to Vancouver, Carlson, Dhanani, and Colton (2021). *Journal of Applied Psychology,* 106: 476-488.

Wang, H., Jia, M., Xiang, Y., & Lan, Y. 2021. Social performance feedback and firm communication strategy. *Journal of Management*. https://doi.org/10.1177/01492063211042266.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. 2019. Moving to a world beyond "$p < 0.05$." *American Statistician,* 73(Suppl.1): 1-19.