CAMBRIDGE
UNIVERSITY PRESS

COMMENTARY

Three cheers for descriptive statistics—and five more reasons why they matter

Marcus Credé[1] and P. D. Harms[2]*

[1]Iowa State University and [2]University of Alabama
*Corresponding author. Email: pdharms@cba.ua.edu

We would like to echo the sentiments of Murphy (2021) in his provocative reflections on the importance of descriptive statistics. The reported values that are typically found in Table 1 of empirical studies are not only important and informative but also often convey far more interesting information than what is conveyed by the "sophisticated" statistical methods and tests that often follow and that most researchers decide to focus on when interpreting their work. To add further weight to this point, we offer five other reasons for how and why descriptive statistics and descriptive information are so critical and why researchers, editors, reviewers, and readers should pay more attention to them when evaluating research findings. We discuss these in turn.

Readers often wish to understand the context in which research occurred: What country was the data gathered in, what jobs did participants hold, what industry did they work in, what was the response rate? This and other similar type of information reflecting substantive situational, methodological, or sample characteristics are also often of primary interest to meta-analytic researchers who wish to examine potential moderators of relationships (i.e., wishing to account for effect size heterogeneity). Is a relationship consistent across industries (e.g., financial versus manufacturing), types of organization (e.g., private versus public, small versus large), or countries (economically developing versus economically developed)? Remarkably, this information is often missing from articles that are published in our top journals. One of us recently started a review project in which articles from four top journals in the discipline are coded for this type of information. Although we are far from complete with this review, we have been struck with how often such basic descriptive information about samples and methods is simply absent. Among the articles coded to date, less than half report the ethnicity of the participants or the types of jobs held by the participants and only 56% report data on the industry in which the data were collected. Other interesting—and to meta-analysts potentially important—information is also remarkably often unreported. We have conducted many meta-analytic reviews of the industrial-organizational (I-O) and I-O-adjacent literatures, and repeatedly we have been stymied in our attempt to model effect-size heterogeneity by the widespread failure to report what we consider relatively basic information about the sample, the research setting, and the research methodology. Uniform adherence to even a basic I-O-specific modification of the APA reporting guidelines (Appelbaum et al., 2018) would go a long way toward allowing meta-analyses to better explore the reasons for effect-size heterogeneity.

Second, we hold the view that minimum and maximum scores on variables should be routinely reported and attended to. This not only provides information about possible outliers but also can draw attention to other potentially problematic data characteristics such as low-effort responding. For example, in one very influential article (Woolley et al., 2010) that forms the foundation of an entire subdiscipline of research on team performance, the authors reported that at least one of the teams whose data were included in the analysis scored zero on a task of team creativity. The task in question involved a team of four individuals being given 10 minutes to come up with different uses

for a brick; each generated use being given a point irrespective of quality. A score of zero indicates that a team could not think of a single use—strongly suggestive of either confusion about the task, a coding error, or extremely low effort on the part of the team. Similar problematic characteristics in the descriptive statistics that are reported in that paper led Credé and Howardson (2015) to question the validity of the highly influential inferences that were drawn by Woolley et al.

Third, we would strongly urge scholars to consider reporting percentage of maximum possible scores (POMP; Cohen et al., 1999). This simple linear transformation of descriptive statistics means that effect sizes can be more readily compared across scales with different response options and renders the size of effects to a more readily interpretable scale (out of 100). In our own experience in working with practitioners, we find that it allows for much more immediate understanding and understanding in terms of return on investment for interventions and the advantages to be gained from using various predictors in employment decisions. POMP scores are a simple and elegant solution for data presentation and interpretation, and yet we almost never see organizational scholars using them in the published literature. Rather than the exception, we believe it should be the norm for reporting in descriptive statistics.

Fourth, we are of the view that standard deviations of scores are not sufficiently interpreted by researchers or attended to. We suspect that this is primarily because population standard deviations are almost entirely unknown for our field's focal variables such that the degree of range restriction in a sample typically remains unknown. But the absence of normative information should not lead us to ignore the information in standard deviations. One, perhaps very crude, way of aiding the interpretation of standard deviations absent population norms might be to indicate the ratio of the observed standard deviation relative to the standard deviation that would occur in the presence of a uniform distribution. The standard deviation of a uniform distribution is easy to calculate as $\sqrt{[(b - a)^2 \div 12]}$, where b and a are the maximum and minimum possible scores on the distribution.

So, a researcher who reports a standard deviation of 0.52 on a measure that is based on a 1–5 Likert-type scale might note that this standard deviation is approximately 45% of the standard deviation of a uniform distribution on the same 1–5 Likert scale. Another researcher who measures the same variable using a 1–7 Likert-type scale and finds a standard deviation of 0.62 might note that this standard deviation is approximately 36% of the standard deviation of a uniform distribution on the same 1–7 Likert-type scale. This might allow both a more intuitive understanding of how much range restriction is present (relative to the uniform distribution, not the population) and might also aid comparison across studies that rely on different response options.

Finally, it is worth noting that means, standard deviations, and correlations can be used to check for the reproducibility of many popular analytic techniques, including regression, path analysis, structural equation modeling, and factor analysis. Indeed, the inability to reproduce the report path analysis results of one paper from the reported correlation matrix contributed to the ultimate retraction of one of six articles from the journal *Leadership Quarterly* as described by Atwater et al. (2014). Similarly, one of us used the correlational data that were reported in four separate papers to show that the factor-analysis-based inferences regarding a supposed "collective intelligence" factor were incorrect (Credé & Howardson, 2015). In another paper (Credé et al., 2016), one of us was able to use a published correlation matrix to demonstrate that a reported regression result was due to a negative suppression effect that occurred only because of the specific pattern of a series of highly collinear predictor variables. Even simple tests for examining whether the reported means are even possible given the scale response options and sample sizes have been demonstrated to be effective indicators of whether inappropriate data manipulation has taken place (Brown & Heathers, 2016). At a time when issues of the reproducibility of research findings are of increasing concern (see Goodman et al., 2016; Grand et al., 2018), we believe that readers should not assume that data analyses have been correctly conducted or interpreted, should

routinely consider whether the reported findings are consistent with the data that are often reported in Table 1, and should take the initiative toward correcting the record when they are not (Harms & Credé, 2020; Harms et al., 2018).

References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, **73**, 3–25.

Brown, N., & Heathers, J. (2016). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science*, **8**(4), 363–369.

Cohen, P., Cohen, J., Aiken, L., & West, S. (1999). The problem of units and the circumstances for POMP. *Multivariate Behavioral Research*, **34**, 315–346.

Credé, M., Gelman, A., & Nickerson, C. (2016). Questionable association between front boarding and air rage. *Proceedings of the National Academy of Sciences*, **113**, E7348.

Credé, M., & Howardson, G. (2018). The structure of group task performance: A second look at "Collective Intelligence." *Journal of Applied Psychology*, **102**(10), 1483–1492.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine, 8*(341), Article 341ps12. http://stm.sciencemag.org/content/8/341/341ps12.full

Grand, J. A., Rogelberg, S. G., Allen, T. D., Landis, R. S., Reynolds, D. H., Scott, J. C., Tonidandel, S., & Truxillo, D. M. (2018). A systems-based approach to fostering robust science in industrial-organizational psychology. *Industrial and Organizational Psychology*, **11**, 4–42.

Harms, P. D. & Credé, M. (2020). Bringing the review process into the 21st century: Post-publication peer review. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **13**(1), 51–53.

Harms, P. D., Credé, M., & DeSimone, J. A. (2018). The last line of defense: Corrigenda and retractions. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **11**(1), 61–65.

Murphy, K. (2021). In praise of Table 1: The importance of making better use of descriptive statistics. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **14**(4), 461–477.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, **330**(6004), 686–688. DOI: 10.1126/science.1193147