© *Academy of Management Annals*
2018, Vol. 12, No. 1, 1–28.
https://doi.org/10.5465/annals.2016.0011

WHAT YOU SEE IS WHAT YOU GET? ENHANCING METHODOLOGICAL TRANSPARENCY IN MANAGEMENT RESEARCH

HERMAN AGUINIS[1]
George Washington University

RAVI S. RAMANI
George Washington University

NAWAF ALABDULJADER
George Washington University

We review the literature on evidence-based best practices on how to enhance methodological transparency, which is the degree of detail and disclosure about the specific steps, decisions, and judgment calls made during a scientific study. We conceptualize lack of transparency as a "research performance problem" because it masks fraudulent acts, serious errors, and questionable research practices, and therefore precludes inferential and results reproducibility. Our recommendations for authors provide guidance on how to increase transparency at each stage of the research process: (1) theory, (2) design, (3) measurement, (4) analysis, and (5) reporting of results. We also offer recommendations for journal editors, reviewers, and publishers on how to motivate authors to be more transparent. We group these recommendations into the following categories: (1) manuscript submission forms requiring authors to certify they have taken actions to enhance transparency, (2) manuscript evaluation forms including additional items to encourage reviewers to assess the degree of transparency, and (3) review process improvements to enhance transparency. Taken together, our recommendations provide a resource for doctoral education and training; researchers conducting empirical studies; journal editors and reviewers evaluating submissions; and journals, publishers, and professional organizations interested in enhancing the credibility and trustworthiness of research.

The field of management and many others are currently debating the credibility, trustworthiness, and usefulness of the scholarly knowledge that is produced (Davis, 2015; George, 2014; Grand et al., in press). It is worrisome that from 2005 through 2015, 125 articles have been retracted from business and management journals, and from 2005–2007 to 2012–2015, the number of retractions has increased by a factor of ten (Karabag & Berggren, 2016). In addition, 25 to 50 percent of published articles in management and other fields have inconsistencies or errors (Goldfarb & King, 2016; Nuijten, Hartgerink, Assen, Epskamp, & Wicherts 2016; Wicherts, Bakker, & Molenaar, 2011). Overall, there is a proliferation of evidence indicating substantial reasons to doubt the veracity and, justifiably, the conclusions and implications of scholarly work (Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016b; Schwab & Starbuck, 2017) because researchers are often unable to reproduce published results (Bakker, van Dijk, & Wicherts, 2012; Bergh, Sharp, Aguinis, & Li, 2017a; Bergh, Sharp, & Li, 2017b; Cortina, Green, Keeler, & Vandenberg, 2017b). Regardless of whether this lack of reproducibility is a more recent phenomenon, or one that has existed for a long time but has only recently gained prominence, it seems that we have

Ravi S. Ramani and Nawaf Alabduljader contributed equally to this work.

We thank Daan van Knippenberg, Sharon K. Parker, and two anonymous *Academy of Management Annals* reviewers for highly constructive feedback on previous drafts. Also, we thank P. Knight Campbell for his assistance with the literature review and data collection during the initial stages of our project. A previous version of this article was presented at the meetings of the Academy of Management, Atlanta, GA, August 2017.

[1] Corresponding author.

reached a tipping point such that there is an urgency to understand this phenomenon and find solutions to address it.

Concerns about lack of reproducibility are not entirely surprising considering the relative lack of methodological transparency about the process of conducting empirical research that eventually leads to a published article (Banks et al., 2016a; Bedeian, Taylor, & Miller, 2010; John, Loewenstein, & Prelec, 2012; O'Boyle, Banks, & Gonzalez-Mulé, 2017; Schwab & Starbuck, 2017; Simmons, Nelson, & Simonsohn, 2011; Wicherts et al., 2011; Wigboldus & Dotsch, 2016). We define *methodological transparency* as the *degree of detail and disclosure about the specific steps, decisions, and judgment calls made during a scientific study*. Based on this definition, we conceptualize transparency as a continuum—a matter of degree—and not as a dichotomous variable (i.e., transparency is present or absent). Clearly, researchers make numerous choices, judgment calls, and decisions during the process of conceptualizing and designing studies, as well as collecting data, analyzing them, and reporting results. The more explicit, open, and thorough researchers are about disclosing each of these choices, judgment calls, and decisions, the greater the degree of methodological transparency.

Low methodological transparency has a detrimental impact on the credibility and trustworthiness of research results because it precludes inferential reproducibility. *Inferential reproducibility is the ability of others to draw similar conclusions to those reached by the original authors regarding a study's results* (Goodman, Fanelli, & Ioannidis, 2016). Note that this is different from *results reproducibility*, which is the *ability of others to obtain the same results using the same data as in the original study*. From a measurement perspective, results reproducibility is conceptually analogous to reliability because it is about consistency. Specifically, do researchers other than those who authored a study find the same (i.e., consistent) results as reported in the original paper? On the other hand, inferential reproducibility is conceptually analogous to validity because it is about making similar inferences based on the results. Specifically, do researchers other than those who authored a study reach similar conclusions about relations between variables as described in the original study? Results reproducibility (i.e., reliability) is a necessary but insufficient precondition for inferential reproducibility (i.e., validity). In other words, if we cannot obtain the same results as in the published study using the same data, inferences are clearly going to be different. But, it is possible to reproduce results (i.e., high reliability) but not inferences (i.e., low validity). Inferential reproducibility (i.e., validity or relations between variables) is the critical issue in terms of building and testing theories and the credibility of the knowledge that is produced, whereas results reproducibility (i.e., reliability or consistency) is a means to an end.

For example, assume that a team of researchers uses archival data and publishes an article reporting a test of a model including five variables with satisfactory fit statistics. Then, a separate team of researchers uses the same dataset with the same five variables and is able to reproduce the exact same results (i.e., high reliability). This is a situation with a high degree of results reproducibility. Now, assume that, unbeknownst to the second team, the first team of researchers had tested 50 different configurations of variables and, in the end, they found and reported the one configuration of the five variables that resulted in the best possible fit statistics. Obviously, testing so many configurations maximized capitalization on chance, and the good fit of the final model is more likely due to chance rather than substantive relations (Aguinis, Cascio, & Ramani, 2017). Enhancing transparency by disclosing that 50 different configurations of variables were tested until the final set was found would not affect results reproducibility, but it would certainly change inferential reproducibility. That is, the second team of researchers would reach very different inferences from the same results because the good fit of the model would be attributed to sampling error and chance rather than the existence of substantive relations between variables.

Many articles published in management journals represent situations similar to the example described previously: We simply do not know whether what we see is what we get. Most things seem just right: measures are valid and have good psychometric qualities, hypotheses described in the Introduction section are mostly supported by results, statistical assumptions are not violated (or not mentioned), the "storyline" is usually neat and straightforward, and everything seems to be in place. But, unbeknownst to readers, many researchers have engaged in various trial-and-error practices (e.g., revising, dropping, and adding scale items), opaque choices (e.g., including or excluding different sets of control variables), and other decisions (e.g., removing outliers, retroactively creating hypotheses after the data were analyzed) that are not disclosed fully. Researchers in management and other fields have considerable latitude in terms of the choices, judgment calls, and trial-and-error decisions they make in every step of the research process—from theory, to design, measurement, analysis, and reporting of results (Bakker et al., 2012; Simmons et al., 2011). Consequently, other researchers are unable to reach similar conclusions due to insufficient information (i.e., low transparency) of what happened in what we label the "research kitchen" (e.g., Bakker et al., 2012; Bergh et al., 2017a, 2017b; Cortina et al. 2017b).

As just one of many examples of low methodological transparency and its negative impact on inferential reproducibility, consider practices regarding the elimination of outliers—data points that are far from the rest of the distribution (Aguinis & O'Boyle, 2014; Joo, Aguinis, & Bradley, 2017). A review of 46 journals and book chapters on methodology and 232 substantive articles by Aguinis, Gottfredson, and Joo (2013) documented the relative lack of transparency in how outliers were defined, identified, and handled. Such decisions affected substantive conclusions regarding the presence, absence, direction, and size of effects. Yet, as Aguinis et al. (2013) found, many authors made generic statements, such as "outliers were eliminated from the sample," without offering details on how and why they made such a decision. This lack of transparency makes it harder for a healthily skeptical scientific readership to evaluate the credibility and trustworthiness of the inferences drawn from the study's findings. Again, without adequate disclosure about the processes that take place in the "research kitchen," it is difficult, if not impossible, to evaluate the veracity of the conclusions described in the article.

We pause here to make an important clarification. Our discussion of transparency, or lack thereof, does not mean that we wish to discourage discovery- and trial-and-error-oriented research. To the contrary, epistemological approaches other than the pervasive hypothetico-deductive model, which has dominated management and related fields since before World War II (Cortina, Aguinis, & DeShon, 2017a), are indeed useful and even necessary. For example, inductive and abductive approaches can lead to important theory advancements and discoveries (Fisher & Aguinis, 2017; Hollenbeck & Wright, 2017; Murphy & Aguinis, 2017). Sharing our perspective, Hollenbeck and Wright (2017) defined "tharking" as "clearly and transparently presenting new hypotheses that were derived from post hoc results in the Discussion section of an article. The emphasis here is on how (transparently) and where (in the Discussion section) these actions took place" (p. 7). So, we are not advocating a rigid adherence to a hypothetico-deductive approach but, rather, epistemological and methodological plurality that has high methodological transparency.

THE PRESENT REVIEW

The overall goal of our review is to improve the credibility and trustworthiness of management research by providing evidence-based best practices on how to enhance methodological transparency. Our recommendations provide a resource for doctoral education and training; researchers conducting empirical studies; journal editors and reviewers evaluating submissions; and journals, publishers, and professional organizations interested in enhancing the credibility and trustworthiness of research. Although we focus on the impact that enhanced transparency will have on inferential reproducibility, many of our recommendations will also help improve results reproducibility. Returning to the reliability–validity analogy, improving validity will, in many cases, also improve reliability.

A unique point of view of our review is that we focus on enhancing methodological transparency rather than on the quality or appropriateness of methodological practices as already addressed by others (Aguinis & Edwards, 2014; Aguinis & Vandenberg, 2014; Williams, Vandenberg, & Edwards, 2009). We focus on the relative lack of methodological transparency because it masks outright fraudulent acts (as committed by, for example, Hunton & Rose, 2011 and Stapel & Semin, 2007), serious errors (as committed by, for example, Min & Mitsuhashi, 2012; Walumbwa, Luthans, Avey, & Oke, 2011), and questionable research practices (as described by Banks, et al., 2016a). Moreover, because of low methodological transparency, many of these errors are either never identified or identified several years after publication. For example, it took at least four years to retract articles published in the *Academy of Management Journal*, *Strategic Management Journal*, and *Organization Science* by disgraced former University of Mannheim professor Ulrich Lichtenthaler. Greater methodological transparency could have substantially aided earlier discovery and possibly even prevented these and many other articles from being published in the first place by making clear that the data used were part of a larger dataset (Min & Mitsuhashi, 2012), providing information regarding decisions to include certain variables (Lichtenthaler, 2008), and being explicit about the levels of inquiry and analysis (e.g., Walumbwa et al., 2011). Although enhanced transparency is likely to help improve research quality because substandard practices are more likely to be discovered early in the manuscript review process, our recommendations are not about the appropriateness of methodological choices, but rather on making those methodological choices explicit.

The remainder of our article is structured as follows. First, we offer a theoretical framework that helps us understand the reasons for the relatively low degree of methodological transparency and how to address this problem. Second, we describe the procedures involved in our literature review. Third, based on results from the literature review, we offer evidence-based best-practice recommendations for how researchers can enhance methodological transparency regarding theory, design, measurement, analysis, and the reporting of results. Finally, we provide recommendations that can be used by editors, reviewers, journals, and publishers to enhance transparency in the manuscript submission and evaluation process. Taken together, our review, analysis, and recommendations aim at enhancing methodological transparency, which will result in improved reproducibility and increase the credibility and trustworthiness of research.

REASONS FOR LOW TRANSPARENCY: THE PERFECT STORM

Why do so many published articles have low methodological transparency (Aytug, Rothstein, Zhou, & Kern, 2012; Cortina et al. 2017b)? To answer this question, we use a theoretical framework from human resource management and organizational behavior and conceptualize the low degree of transparency as a performance problem or, more specifically, a *research performance problem*. Within this framework, excellent research performance is complete transparency, and poor performance is low transparency—resulting in low inferential reproducibility.

Overall, decades of performance management research suggest that performance is determined to a large extent by two major factors: (a) motivation and (b) knowledge, skills, and abilities (KSAs) (Aguinis, 2013; Maier, 1955; Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2017; Vroom, 1964). In other words, individuals must *want* (i.e., have the necessary motivation) to perform well and *know* how to perform well (i.e., have the necessary KSAs).

Recent meta-analytic evidence suggests that ability and motivation have similarly strong effects on performance (Van Iddekinge et al., 2017). For example, when considering job performance (as opposed to training performance and laboratory performance), the mean range restriction and measurement error corrected ability-performance correlation is .31, whereas the motivation-performance correlation is .33. Ability has a stronger effect on objective (i.e., results such as sales) performance (i.e., .51 vs. .26 for motivation), but the effects on subjective performance (i.e., supervisory ratings) are similar (i.e., .32 vs. .31 for motivation). Also, particularly relevant regarding our recommendations for solving the "research performance problem" is that the overall meta-analytically derived corrected correlation between ability and motivation is only .07 (based on 55 studies). In other words, ability and motivation share only half of 1 percent of common variance. Moreover, as described by Van Iddekinge et al. (2017), contrary to what seems to be a common belief, the effects of ability and motivation on performance are not multiplicative but, rather, additive. This means that they do not interact in affecting performance and, therefore, it is necessary to implement interventions that address *both* KSAs and motivation. Accordingly, we include a combination of recommendations targeting these two independent antecedents of research performance.

Regarding motivation as an antecedent for research performance, at present, there is tremendous pressure to publish in "A-journals" because faculty performance evaluations and rewards, such as promotion and tenure decisions, are to a large extent a consequence of the number of articles published in these select few journals (Aguinis, Shapiro, Antonacopoulou, & Cummings, 2014; Ashkanasy, 2010; Butler, Delaney, & Spoelstra, 2017; Byington & Felps, 2017; Nosek, Spies, & Motyl, 2012). Because researchers are rewarded based on the number of publications, they are motivated to be less transparent when transparency might adversely affect the goal of publishing in those journals. As an example, consider the following question: Would researchers fully report the weaknesses and limitations of their study if it jeopardized the possibility of publication? In most cases, the answer is no (Brutus, Aguinis, & Wassmer, 2013; Brutus, Gill, & Duniewicz, 2010).

Interestingly, transparency is not related to the number of citations received by individual articles (Bluhm, Harman, Lee & Mitchell, 2011) or reviewer evaluations regarding methodological aspects of submitted articles (Green, Tonidandel, & Cortina, 2016). Specifically, Bluhm et al. (2011) measured transparency by using two coders who assessed "whether the article reported sufficient information in both data collection and analysis for the study to be replicated to a reasonable extent" (p. 1874) and "statistical analysis revealed no significant relationship between transparency of analysis and the number of cites received by articles ($F_{4,190} = 1.392$, $p = 25$)" (p. 1881). In addition, Green et al. (2016) used a constant comparative method to code reviewers' and editors' decision letters to "build conceptual categories, general themes, and overarching dimensions about research methods and statistics in the peer review process" (p. 406). They generated 267 codes from 1,751 statements in 304 decision letters regarding 69 articles. Green et al. (2016: 426) concluded their article with the following statement: "In conclusion, the present study provides prospective authors with detailed information regarding what the gatekeepers say about research methods and analysis in the peer review process." Transparency was not mentioned once in the entire article. These results provide evidence that greater transparency is not necessarily rewarded and many of the issues described in our article may be "under the radar screen" in the review process. In short, the focus on publishing in "A-journals" as the arbiter of rewards is compounded by the lack of obvious benefits associated with methodological transparency and the lack of negative consequences for those who are not transparent, thus further reducing the motivation to provide full and honest methodological disclosure.

Our article addresses motivation as an antecedent for research performance by providing actionable recommendations for editors, reviewers, and journals and publishers on how to make methodological transparency a more salient requirement for publication. For example, consider the possible requirement that authors state whether they tested for outliers, how outliers were handled, and implications of these decisions for a study's results (Aguinis et al., 2013). This actionable and rather easy to implement manuscript submission requirement can switch an author's expected outcome from "dropping outliers without mentioning it will make my results look better, which likely enhances my chances of publishing" to "explaining how I dealt with outliers is required if I am to publish my paper—not doing so will result in my paper being desk-rejected." In other words, our article offers suggestions on how to influence motivation such that being transparent becomes instrumental in terms of obtaining desired outcomes (i.e., publishing), whereas a low degree of transparency will negatively affect chances of such success.

Insufficient KSAs is the second factor that results in low transparency—our "research performance problem." Because of the financial constraints placed on business and other schools (e.g., psychology, industrial, and labor relations), many researchers and doctoral students are not receiving state-of-the-science methodological training. Because doctoral students receive tuition waivers and stipends, many schools view doctoral programs as cost centers when compared with undergraduate and master's programs. The financial pressures faced by schools often result in less resources being allocated to training doctoral students, particularly in the methods domain (Byington & Felps, 2017; Schwab & Starbuck, 2017; Wright, 2016). For example, a study by Aiken, West, and Millsap (2008) involving graduate training in statistics, research design, and measurement in 222 psychology departments across North America concluded that "statistical and methodological curriculum has advanced little [since the 1960s]" (p. 721). Similarly, a 2013 benchmarking study conducted within the United States and involving 115 industrial and organizational psychology programs found that although most of them offer basic research methods and entry-level statistics course (e.g., Analysis of Variance (ANOVA), regression), the median number of universities offering courses on measurement/test development, meta-analysis, hierarchical linear modeling, nonparametric statistics, and qualitative/mixed methods is zero (Tett, Walser, Brown, Simonet, & Tonidandel, 2013).

This situation is clearly not restricted only to universities in the United States. For example, in many universities in the United Kingdom and Australia, there is minimal methodological training beyond that offered by the supervisor. In fact, doctoral students in Australia are expected to graduate in 3.5 years at the most. Combined with the paucity of methodological courses offered, this abbreviated timeline makes it very difficult for doctoral students, who are the researchers of the future, to develop sufficient KSAs. Lack of sufficient methodological KSAs gives authors even more "degrees of freedom" when faced with openly disclosing choices and judgment calls because the negative consequences associated with certain choices are simply unknown.

The lack of sufficient methodological training and KSAs is also an issue for editors, reviewers, and publishers/professional organizations (e.g., Academy of Management). As documented by Bedeian, Van Fleet, and Hyman (2009), the sheer volume of submissions requires expanding editorial boards to include junior researchers, even at "A-journals." Unfortunately, these junior researchers themselves may not have received rigorous and comprehensive methodological training because of the financial constraints on schools and departments. The lack of broad and state-of-the-science methodological training, the rapid developments in research methodology (Aguinis, Pierce, Bosco, & Muslin, 2009; Cortina et al., 2017a), and the sheer volume and variety of types of manuscript submissions mean that even the gatekeepers can be considered novices and, by their own admission, often do not have the requisite KSAs to adequately and thoroughly evaluate all the papers they review (Corley & Schinoff, 2017).

To address the issue of KSAs, our review identifies choices and judgment calls made by researchers during the theory, design, measurement, analysis, and reporting of results that should be described transparently. By distilling the large, fragmented, and often-technical literature into evidence-based best-practice recommendations, our article can be used as a valuable KSA resource and blueprint for enhancing methodological transparency.

While we focus on individual-level factors, such as motivation and KSAs, context clearly plays an important role in creating the research performance problem. That is, researchers do not work in a vacuum. In fact, many of the factors we mentioned as influencing motivation (e.g., pressure to publish in "A-journals") and KSAs (e.g., fewer opportunities for methodological training and re-tooling) are contextual in nature. In describing the importance of context, Blumberg and Pringle (1982) offered the example that researchers are faced with environments that differ in terms of research-related resources (resulting in different KSAs), which in turn affect their research performance. Another contextual factor related to the pressure to publish in "A-journals" is the increase in the number of manuscript submissions, causing an ever-growing workload on editors and reviewers. Many journals receive more than 1,000 submissions a year, making it necessary for many action editors to produce a decision letter every three days or so—365 days a year (Cortina et al., 2017a). But, the research performance of editors and reviewers is still contingent on their own publications in "A-journals" (Aguinis, de Bruin, Cunningham, Hall, Culpepper, & Gottfredson, 2010a). So, the increased workload associated with the large number of submissions, along with other obligations (e.g., teaching, administrative duties), suggests that our current system places enormous, and arguably unrealistic, pressure on editors and reviewers to scrutinize manuscripts closely and identify areas where researchers need to be more transparent (Butler et al., 2017).

In short, we conceptualize low transparency as a research performance problem. Decades of research on performance management suggest that addressing performance problems at the individual-level requires that we focus on its two major antecedents: motivation and KSAs. Our evidence-based recommendations on how to enhance transparency can be used by publishers to update journal submission and review policies and procedures, thereby positively influencing authors' motivation to be more transparent. In addition, editors and reviewers can use our recommendations as checklists to evaluate the degree of transparency in the manuscripts they review. Finally, our recommendations are a source of information that can be used to improve doctoral student training and the KSAs of authors. Next, we offer a description of the literature review process that we used as the basis to generate these evidence-based recommendations.

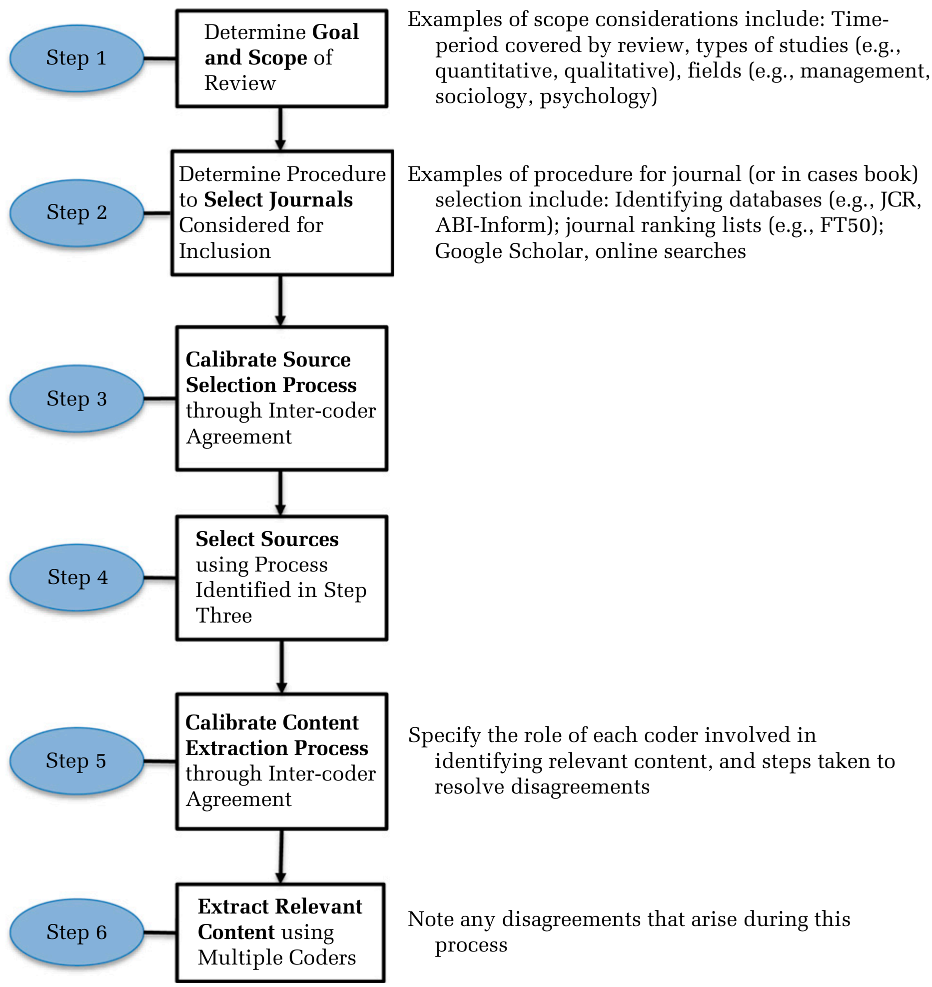
LITERATURE REVIEW

Overview of Journal, Article, and Recommendation Selection Process

We followed a systematic and comprehensive process to identify articles providing evidence-based recommendations for enhancing methodological transparency. Figure 1 includes a general overview of the six steps we implemented in the process of identifying journals, articles, and recommendations. Table 1 offers more detailed information on each of these six steps. As described in Table 1, our process began with 62 journals and the final list upon which we based our recommendations includes 28 journals and 96 articles.

An additional contribution of our article is that the description of our systematic literature review procedures presented generally in Figure 1 and detailed in Table 1 can be used by authors of review articles to appear in the *Academy of Management Annals* (AMA) and other journals. In fact, our detailed description that follows is in part motivated by our observation that the many reviews published in AMA and elsewhere are not sufficiently explicit about criteria for study inclusion and, thus it may be difficult to reproduce the body of work that is included in any particular review article.

FIGURE 1
Overview of Systematic and Reproducible Process for Identifying Journals, Articles, and Content to Include in a Literature Review Study



Note: JCR = Web of Science Journal Citation Reports; FT50 = Financial Times journal ranking list

Step 1: Goal and Scope of Review

We adopted an inclusive approach in terms of substantive and methodological journals from management, business, sociology, and psychology (i.e., general psychology, applied psychology, organizational psychology, and mathematical psychology). Because methods evolve rapidly, we only considered articles published between January 2000 and August 2016. Furthermore, we only focused on literature reviews providing evidence-based recommendations regarding transparency in the form of analytical work, empirical data (e.g., simulations), or both. In a nutshell, we included literature reviews that integrated and synthesized the available evidence. As such, our article is a "review of reviews."

Step 2: Journal Selection Procedures

As shown in Table 1, we used three sources to select journals. First, we used the Web of Science Journal Citation Reports (JCR) database and included journals from several categories, such as business, management, psychology, sociology, and others. Second, we used the list of journals created by the Chartered Association of Business Schools (ABS) to increase the representation of non-US journals in our journal selection process. Third, we included journals from the Financial Times 50 (FT50) list. Several journals were included in more than one of the sources, so subsequent sources added increasingly fewer new journals to our list. We excluded journals not directly relevant to the field of management, such

TABLE 1
Detailed Description of Steps Used in Literature Review to Identify Evidence-Based Best-Practice Recommendations to Enhance Transparency

Step	Procedure
1	**Goal and Scope of Review:** • Provide recommendations on how to enhance transparency using published, evidence-based reviews encompassing a wide variety of epistemologies and methodological practices. We only focused on literature reviews providing evidence-based recommendations regarding transparency in the form of analytical work, empirical data (e.g., simulations), or both • Due to rapid development of methods, we only considered articles published between January 2000 and August 2016 (including in press articles) • To capture the breadth of interests of Academy of Management (AOM) members, we considered journals from the fields of management, business, psychology, and sociology
2	**Selection of Journals Considered for Inclusion:** • Used a combination of databases and journal ranking lists to minimize a U.S.-centric bias and to identify outlets that covered a wide range of topics of interest to AOM members, covering both substantive and methodological (technical) journals • Identified 62 potential journals for inclusion: ∘ Web of Science Journal Citation Reports (JCR) Database: –51 unique journals; excluded duplicates –Business / Management / Applied Psychology: Top-25 journals in each category –Mathematical Psychology: All 13 journals –Social Sciences / Mathematical Methods / Sociology / Multidisciplinary Psychology: 4 journals ∘ Chartered Association of Business Schools (ABS) Journal Ranking List: –4 unique journals; excluded duplicates –General Management, Ethics, and Social Responsibility / Organization Studies categories –4* and 4 rated journals only ∘ Financial Times 50 (FT50): –7 unique journals; excluded duplicates –Management journals only
3	**Calibrate Source (i.e., Article) Selection Process through Intercoder Agreement (3 calibration rounds):** • Identified 81 articles from 6 journals: ∘ Adopted a manual search process to identify articles to increase comprehensiveness of our review in case a relevant article did not contain a specific keyword ∘ In each round of calibration, the coders (coder 1 = NA, coder 2 = RSR, coder 3 = PKC) independently coded articles in five year increments from six of the 62 journals. Articles were coded as "In/Out/Maybe" based on whether they met the inclusion criteria outlined in Step 1 ∘ Coders combined results after the first two rounds and met to discuss reasons for coding and status of articles which had been assigned a "Maybe" rating –After two rounds, coders agreed on 67% of their ratings ∘ HA reviewed results after the second round and narrowed differing perceptions of inclusion criteria. Coders recoded article selections from the first two rounds based on feedback from HA ∘ Coders began the third round of calibration and followed procedure outlined above regarding rating articles and discussion –After three rounds, coders agreed on 90% of their ratings ∘ HA reviewed results after the third round and provided further feedback
4	**Select Source (i.e., Articles) using Process Identified in Step Three:** • Identified an additional 84 articles from 33 journals: ∘ Coders reviewed different time periods of the same journals to reduce chances of confounding coder with journal ∘ Time Period Reviewed: –Coder 1 (NA): 2000–2005 –Coder 2 (RSR): 2006–2010 –Coder 3 (PKC): 2011–2016 ∘ Excluded 23 journals from the original list of journals considered because we did not find any articles meeting our article inclusion criteria • Initial selection of 165 articles from 39 journals for consideration
5	**Calibrate Content Extraction Process through Intercoder Agreement:** • All authors read the same five articles (out of the 165 identified in steps 3 and 4) and derived recommendations regarding methodological transparency • Authors met to compare notes from selected articles and to confirm they addressed evidence-based recommendations to enhance transparency • Process was repeated with another five articles to ensure intercoder agreement

TABLE 1
(Continued)

Step	Procedure
6	**Extract Relevant Content using Multiple Coders:** • Each coder read the full text of the remaining 155 articles and made notes on evidence-based best practice recommendations provided, both in terms of how to make research more transparent and the rationale for those recommendations • Coders compiled and categorized recommendations into theory, design, measurement, analysis, and reporting of results categories • All the authors reviewed results together to fine-tune recommendations and rationale • Excluded 69 articles from 11 journals due to a lack of relevant recommendations on how to enhance transparency • Final list of recommendations extracted from 96 articles out of 28 journals (see Table 2)

Notes: HA = Herman Aguinis, NA = Nawaf Alabduljader, RSR = Ravi S. Ramani, PKC = P. Knight Campbell.

as the *Journal of Interactive Marketing* and *Supply Chain Management*. Overall, we identified 62 journals which potentially published literature reviews with recommendations regarding methodological transparency. As described in the next section, 23 journals did not include any articles that met our inclusion criteria, and we excluded 11 additional journals upon closer examination of the articles during the recommendation selection process. Thus, in the end our literature review included 28 journals, which are listed in Table 2.

Steps 3 and 4: Article Selection Procedures

We used a manual search process to identify articles including evidence-based reviews of methodological practices directly related to transparency. Specifically, Nawaf Alabduljader, Ravi S. Ramani, and P. Knight Campbell (hereafter coders), read the title, abstract, and, in some instances, the full text of the article, before deciding on classifying each as "in," "out," or "maybe." In this early stage of the article selection process, the coders erred in the direction of including an article that may not have met the inclusion criteria, rather than excluding an article that did. This allowed us to cast a wide net in terms of inclusivity and then collaboratively eliminate irrelevant articles, rather than missing potentially important information. Each coder also independently categorized the selected articles as primarily related to theory, research design, measurement, data analysis, or reporting of results. Articles that fit multiple categories were labeled accordingly. Herman Aguinis reviewed the list of articles selected and identified those that did not meet the inclusion criteria as they focused on how to conduct high-quality research and the development and appropriate use of new methods rather than transparency. After the third calibration round of coding as described in Step 3 of Table 1, we compared the independent lists of articles using a simple matching function in Excel to determine the overlap between independent selections. In terms of intercoder agreement, results indicated that 90 percent of the articles in each coder's independently compiled lists were the same as those selected by the other coders. The coders then proceeded with the remaining journals. The article selection process resulted in a total of 39 journals containing 165 possibly relevant articles. The 23 journals that did not include review articles with recommendations on how to enhance methodological transparency and were, therefore, excluded from our review are listed at the bottom of Table 2.

Steps 5 and 6: Recommendation Selection Procedures

To select recommendations, the coders read the full text of each of the 165 identified potential articles, and made notes on evidence-based best practice recommendations provided, both in terms of how to make research more transparent, and the rationale and evidence supporting those recommendations. Again, we used a calibration process for recommendation selection to ensure intercoder agreement (described in Table 1). We found that 69 of the articles initially included during the article selection process did not address transparency. After eliminating these 69 articles, a further 11 journals were excluded from our final review (see bottom of Table 2). The final list of articles included in our review is included in the References section and denoted by an asterisk. Overall, our final literature review is based on 96 articles from 28 journals.

TABLE 2
List of Journals Included in Literature Review on Evidence-Based Best-Practice Recommendations to Enhance Transparency (2000–2016)

	Journal Title
1	Academy of Management Journal
2	Annual Review of Organizational Psychology and Organizational Behavior
3	Behavior Research Methods
4	British Journal of Management
5	British Journal of Mathematical & Statistical Psychology
6	Educational and Psychological Measurement
7	Entrepreneurship Theory and Practice
8	European Journal of Work and Organizational Psychology
9	Family Business Review
10	Human Relations
11	Journal of Applied Psychology
12	Journal of Business and Psychology
13	Journal of Business Ethics
14	Journal of International Business Studies
15	Journal of Management
16	Journal of Management Studies
17	Journal of Organizational Behavior
18	Leadership Quarterly
19	Long Range Planning
20	Methodology: European Journal of Research Methods for Behavior and Social Science
21	Multivariate Behavioral Research
22	Organizational Research Methods
23	Personnel Psychology
24	Psychological Methods
25	Psychometrika
26	Psychonomic Bulletin & Review
27	Sociological Methods & Research
28	Strategic Management Journal

Notes: The following journals were initially included but later excluded during the article selection process of the literature review as described in text and Table 1: *Academy of Management Perspectives, Academy of Management Review, Administrative Science Quarterly, Applied Measurement in Education, Applied Psychology-Health and Well Being, Business Ethics Quarterly, Business Strategy and the Environment, Econometrica, Harvard Business Review, Human Resource Management, Journal of Classification, Journal of Occupational and Organizational Psychology, Journal of Vocational Behavior, Journal of World Business, Nonlinear Dynamics Psychology and Life Sciences, Organization Science, Organization Studies, Organizational Behavior and Human Decision Processes, Organizational Psychology Review, Research Policy, Sloan Management Review, Strategic Entrepreneurship Journal*, and *Tourism Management.* The following journals were excluded after the initial literature review as described in text and Table 1: *Academy of Management Annals, Applied Psychological Measurement, International Journal of Management Reviews, Journal of Behavioral Decision Making, Journal of Business Venturing, Journal of Educational and Behavioral Statistics, Journal of Educational Measurement, Journal of Mathematical Psychology, Journal of Occupational Health Psychology, Management Science*, and *Structural Equation Modeling.*

EVIDENCE-BASED BEST-PRACTICE RECOMMENDATIONS

We present our recommendations for authors under five categories: theory, design, measurement, analysis, and reporting of results. Similar to previous research, we found that several recommendations are applicable to more than one of these stages and therefore the choice to place them in a particular category is not clear-cut (Aguinis et al., 2009). For example, issues regarding validity are related to theory, design, measurement, and analysis. Thus, we encourage readers to consider our recommendations taking into account that a particular recommendation may apply to more than the specific research stage in which it has been categorized. Also, as noted earlier, our recommendations are aimed specifically at enhancing inferential reproducibility. However, many of them will also serve the dual purpose of enhancing results reproducibility as well. To make our recommendations more tangible and concrete, we also provide examples of published articles that have implemented some of them. We hope that the inclusion of these exemplars, which cover both micro and macro domains and topics, will show that our recommendations are in fact actionable and doable—not just wishful thinking. Finally, please note that many, but not all, of our recommendations are sufficiently broad to apply to both quantitative and qualitative research, particularly those regarding theory, design, and measurement.

Enhancing Transparency Regarding Theory

Table 3 includes recommendations on how to enhance transparency regarding theory. These recommendations highlight the need to be explicit regarding research questions (e.g., theoretical goal, research strategy, and epistemological assumptions), level of theory, measurement, and analysis (e.g., individual, dyadic, organizational), and specifying the a priori direction of hypotheses (e.g., linear, curvilinear) as well as distinguishing a priori versus post hoc hypotheses.

For example, consider the recommendation regarding the level of inquiry. This recommendation applies to most studies because explicit specification of the focal level of theory, measurement, and analysis is necessary for drawing similar inferences (Dionne et al., 2014; Yammarino, Dionne, Chun, & Dansereau, 2005). Level of theory refers to the focal level (e.g., individual, team, firm) to which one seeks to make generalizations (Rousseau, 1985). Level of

TABLE 3
Evidence-Based Best-Practice Recommendations to Enhance Methodological Transparency Regarding Theory

Recommendations	Improves Inferential Reproducibility by Allowing Others to. . .
1. Specify the theoretical goal (e.g., creating a new theory, extending existing theory using a prescriptive or positivist approach, describing existing theory through interpretive approach); research strategy (e.g., inductive, deductive, abductive); and epistemological orientation (e.g., constructivism, objectivism) [3, 7, 8, 21, 35, 47, 56, 60, 89]	1. Use the same theoretical lens to evaluate how researchers' assumptions may affect the ability to achieve research goals (e.g., postpositivim assumes objective reality exists, focuses on hypothesis falsification; interpretative research assumes different meanings exist, focuses on describing meanings), and conclusions drawn (e.g., data-driven inductive approach versus theory-based deductive approach)
2. Specify level of theory, measurement, and analysis (e.g., individual, dyadic, organizational) [21, 32, 33, 43, 50, 70, 80, 85, 95, 96]	2. Use the same levels to interpret implications for theory (e.g., do results apply to individuals or organizations? Are results influenced by the alignment or lack thereof between the level of theory, measurement, and analysis?)
3. Acknowledge whether there was an expected a priori direction (e.g., positive, plateauing, curvilinear) for the nature of relations as derived from the theoretical framework used. Identify and report any post hoc hypotheses separately from a priori hypotheses. Report both supported and unsupported hypotheses [3, 12, 23, 62, 71, 93, 96]	3. Differentiate between inductive and deductive tests of theory and gather information regarding the precision, relevance, and boundaries of the theory tested (e.g., match between theoretical predictions and results; presence of contradictory findings; does the theory predict a linear or curvilinear relation?)

Notes: Sources used to derive evidence-based recommendations: [3]Aguinis and Vandenberg (2014), [7]Aram and Salipante (2003), [8]Aytug, Rothstein, Zhou, and Kern (2012), [12]Banks et al., (2016a), [21]Bluhm, Harman, Lee, and Mitchell (2011), [23]Bosco, Aguinis, Field, Pierce, and Dalton (2016), [32]Castro (2002), [33]Chandler and Lyon (2001), [35]Chenail (2009), [43]Dionne et al. (2014), [47]Eisenhardt and Graebner (2007), [50]Feng (2015), [56]Gephart (2004), [60]Gibbert, Ruigrok, and Wicki (2008), [62]Haans, Pieters, and He (2016), [70]Klein and Kozlowski (2000), [71]Leavitt, Mitchell, and Peterson (2010), [80]Ployhart and Vandenberg (2010), [85]Schriesheim, Castro, Zhou, and Yammarino (2002), [89]van Aken (2004), [93]Wigboldus and Dotsch (2016), [95]Yammarino, Dionne, Chun, & Dansereau (2005), [96]Zhang and Shaw (2012).

measurement refers to the level at which one collects data. Level of analysis refers to the level at which data is assigned for hypothesis testing and analysis. Transparency regarding the level of theory, measurement, and analysis allows others to recognize potential influences of such decisions on the research question and the meaning of results, such as whether constructs and results differ when conceptualized and tested at different levels, or whether variables at different levels may affect the substantive conclusions reached (Dionne et al., 2014; Hitt, Beamish, Jackson & Mathieu, 2007; Mathieu & Chen, 2011; Schriesheim, Castro, Zhou, & Yammarino, 2002).

Without transparency about the level of inquiry, it is difficult for reviewers and readers to reach similar conclusions about the meaning of results, which compromises the credibility of the findings. For example, a retraction by Walumbwa et al. (2011) was attributed to (among other things) the inappropriate use of levels of analysis, leading to the irreproducibility of the study's conclusions. Had the authors reported the level of analysis, reviewers would have been able to use the same levels to interpret whether results were influenced by the alignment or lack thereof between level of theory, measurement, and analysis. This may have helped reviewers identify the error before publication, thereby avoiding having to eventually retract the paper. Thus, low transparency when specifying levels of inquiry may help account for the difficulty in reproducing conclusions in some research (Schriesheim et al., 2002). Furthermore, authors must be transparent about the decisions they make even when the level of inquiry is explicit, such as when testing the relation between two individual-level variables (Dionne et al., 2014). Being explicit about the level of inquiry is particularly important given the increased interest in developing and testing multilevel models and bridging the micro–macro gap, which adds to model complexities (Aguinis, Boyd, Pierce, & Short, 2011).

Another recommendation is about including post hoc hypotheses separately from a priori ones (Hollenbeck & Wright, 2017). Currently, many authors retroactively create hypotheses after determining the results supported by the data (i.e., Hypothesizing After the Results are Known [HARKing]) (Bosco, Aguinis, Field, Pierce, & Dalton, 2016). HARKing implies that empirical patterns were discovered from data analysis—an inductive approach. By crediting findings to a priori theory, readers assume that the results are based on a random sample of the data, and thus generalizable to a larger population. On the other hand, the truthful and transparent reporting of the use of HARKing (i.e., "tharking") attributes results to the specific sample—a deductive approach. Transparency about HARKing thus changes the inferences drawn from results and makes it a useful investigative technique that provides interesting findings and discoveries (Fisher & Aguinis, 2017; Hollenbeck & Wright, 2017; Murphy & Aguinis, 2017).

Consider the following exemplars of published articles that are highly transparent regarding theory. First, Maitlis (2005) used an interpretive approach to examine social processes involved in organizational sensemaking (i.e., individuals' interpretations of cues from environments) among various organizational stakeholders. Maitlis (2005) explained the theoretical goal ("The aim of this study was theory elaboration," p. 24), the theoretical approach (i.e., describe theory using an interpretive qualitative approach), and the rationale for the choice ("Theory elaboration is often used when preexisting ideas can provide the foundation for a new study, obviating the need for theory generation through a purely inductive, grounded analysis," p. 24). High transparency in stating the theoretical goal, approach, and rationale allows others to use the same theoretical lens to evaluate how researchers' assumptions may affect the ability to achieve research goals and the conclusions drawn. As a second example, transparency about levels of inquiry was demonstrated in an article by Williams, Parker, and Turner (2010), who examined the effect of team personality and transformational leadership on team proactive performance. Williams et al. (2010) stated the level of theory ("Our focus in the current paper is on proactive teams rather than proactive individuals," p. 302), the level of measurement ("Team members (excluding the lead technician) were asked to rate their lead technician, and these ratings were aggregated to produce the team-level transformational leadership score," p. 311), and the level of analysis ("It is important to note that, because the analyses were conducted at the team level (N = 43), it was not appropriate to compute a full structural model," p. 313). Moreover, the authors specified levels in their formal hypotheses (e.g. "The mean level of proactive personality in the team will be positively related to team proactive performance", p. 308), which further enhanced transparency.

Enhancing Transparency Regarding Research Design

Table 4 provides recommendations on how researchers can be more transparent about design, including choices regarding type of research design, data collection procedure, sampling method, power analysis, common method variance, and control variables. Information on issues such as sample size sample type, conducting research using passive observation or experiments, and decisions on including or excluding control variables, influence inferences drawn from these results and, therefore, inferential reproducibility (Aguinis & Vandenberg, 2014; Bono & McNamara, 2011).

Many of the recommendations included in Table 4 are related to the need to be transparent about specific steps taken to remedy often-encountered challenges and imperfections in the research design (e.g., common method variance, possible alternative explanations), and to clearly note the impact of these steps on substantive conclusions, as they may actually amplify the flaws they are intended to remedy (Aguinis & Vandenberg, 2014). Without knowing which corrective actions were taken, others are unable to reach similar inferences from the results obtained (Aguinis & Vandenberg, 2014; Becker, 2005).

For example, because of the practical difficulty associated with conducting experimental and quasi-experimental designs, many researchers measure and statistically control for variables other than their variables of interest to account for the possibility of alternative explanations (Becker, 2005; Bernerth & Aguinis, 2016). Including control variables reduces the degrees of freedom associated with a statistical test, statistical power, and the amount of explainable variance in the outcome (Becker, 2005; Bernerth & Aguinis, 2016; Edwards, 2008). On the other hand, excluding control variables can increase the amount of explainable variance and inflate the relation between the predictor and the outcome of interest (Becker, 2005; Bernerth & Aguinis, 2016; Edwards, 2008). Therefore, the inclusion or exclusion of control variables affects the relation between the predictor and the criterion, and the substantive conclusions drawn from study results. Yet, researchers often do not disclose which control variables were initially considered for inclusion and why, which control variables were eventually included and which excluded, and the psychometric properties (e.g., reliability) of those that were included (Becker, 2005; Bernerth & Aguinis, 2016). As reported by Bernerth and Aguinis (2016), many researchers cite previous work or provide ambiguous statements, such as "it might relate" as a reason for control variable inclusion, rather than providing a theoretical rationale for whether control variables have meaningful relations with criteria and

TABLE 4
Evidence-Based Best-Practice Recommendations to Enhance Methodological Transparency Regarding Research Design

Recommendations	Improves Inferential Reproducibility by Allowing Others to . . . .
1. Describe type of research design (e.g., passive observation, experimental); data collection procedure (e.g., surveys, interviews); location of data collection (e.g., North America/China; at work/in a lab/at home); sampling method (e.g., purposeful, snowball, convenience); and sample characteristics (e.g., students versus full-time employees; employment status, hierarchical level in organization; sex; age; race) (14, 15, 21, 22, 28, 31, 35, 44, 59, 60, 75, 82, 83, 86, 96)	1. Determine influence of study design and sample characteristics on research questions and inferences (e.g., use of cross-sectional versus experimental studies to assess causality), and overall internal and external validity of findings reported (e.g., if theoretical predictions may vary across groups and cultures; if sample is not representative of population of interest or the phenomenon manifests itself differently in sample)
2. If a power analysis was conducted before initiating study or after study completion, report results, and explain if and how they affect interpretation of study's results (1, 3, 10, 24, 29, 36, 52, 76, 86, 90, 91)	2. Draw independent conclusions about the effect of sample size on the ability to detect existing population effects given that low power increases possibility of Type II error (i.e., incorrectly failing to reject the null hypothesis of no effect or relation)
3. If common method variance was addressed, state the theoretical rationale (e.g., failure to correlate with other self-report variables), and study design (e.g., temporal separation and use of self- and other-report measures) or statistical remedies (e.g., Harman one-factor analysis) used to address it (3, 25, 34, 38, 81)	3. Identify the influence, if any, of common method variance preemptive actions and remedies on error variance (i.e., variance attributable to methods rather than constructs of interest), which affects conclusions because it affects the size of obtained effects
4. Provide an explanation of which control variables were included and which were excluded and why, how they influenced the variables of interest, and their psychometric properties (e.g., validity, reliability) (3, 16, 18, 22, 28, 44)	4. Independently determine if conclusions drawn in the study were influenced by choice of control variables because a) Including control variables changes meaning of substantive conclusions to the part of predictor unrelated to control variable, rather than total predictor; b) not specifying causal structure between control variables and focal constructs (e.g., main effect, moderator, and mediator) can cause model misspecification and lead to different conclusions; and c) reporting measurement qualities provides evidence on whether control variables are conceptually valid and representative of underlying construct

Notes: Sources used to derive evidence-based recommendations: [1]Aguinis, Beaty, Boik, and Pierce (2005), [3]Aguinis and Vandenberg (2014), [10]Balluerka, Gómez, and Hidalgo (2005), [14]Bansal and Corley (2011), [15]Baruch and Holtom (2008), [16]Becker (2005), [18]Bernerth and Aguinis (2016), [21]Bluhm, Harman, Lee, and Mitchell (2011), [22]Bono and McNamara (2011), [24]Boyd, Gove, and Hitt (2005), [25]Brannick, Chan, Conway, Lance, and Spector (2010), [28]Brutus, Gill, and Duniewicz (2010), [29]Cafri, Kromrey, and Brannick (2010), [31]Casper, Eby, Bordeaux, Lockwood, and Lambert (2007), [34]Chang, van Witteloostuijn, and Eden (2010), [35]Chenail (2009), [36]Combs (2010), [38]Conway and Lance, 2010; [44]Edwards (2008), [52]Finch, Cumming, Thomason (2001), [59]Gibbert and Ruigrok (2010), [60]Gibbert, Ruigrok, and Wicki (2008), [75]McLeod, Payne, and Evert (2016), [76]Nickerson (2000), [81]Podsakoff, MacKenzie, Lee, and Podsakoff (2003), [82]Rogelberg, Adelman, and Askay (2009), [83]Scandura and Williams (2000), [86]Shen, Kiger, Davies, Rasch, Simon, and Ones (2011), [90]Van Iddekinge and Ployhart (2008), [91]Waldman and Lilienfeld (2016), [96]Zhang and Shaw (2012).

predictors of interest. In addition, some authors may include control variables simply because they suspect that reviewers and editors expect such practice (Bernerth & Aguinis, 2016). Therefore, low transparency regarding the use of control variables reduces inferential reproducibility because it is not known whether conclusions reached are simply an artifact of which specific control variables were included or excluded (Becker, 2005; Bernerth & Aguinis, 2016; Edwards, 2008).

A study by Tsai, Chen, and Liu (2007) offers a good illustration of transparency regarding the use of control variables. They controlled for job tenure when testing the effect of positive moods on task performance. Tsai et al. (2007) provided an explanation of which control variables were included ("We included job tenure (in years) as a control variable"), why they were used ("...meta-analysis showed that the corrected correlation between work experience... and employee job performance was 0.27," p. 1575), and how the control variables might influence the variables of interest ("This positive correlation may be explained by the fact that employees gain more job-relevant knowledge and skills as a result of longer job tenure, which thus leads to higher task performance," p. 1575). An example of high transparency about common method variance is Zhu and Yoshikawa (2016), who examined how a government director's self-identification with both the focal firm and the government influences his or her self-reported governance behavior (managerial monitoring and resource provision). The authors noted why common method variance was a potential problem in their study ("As both the dependent and the independent variables were measured by self-report, common method variance could potentially have influenced the results," p. 1800). In addition, they were highly transparent in describing the techniques used to assess whether common method bias may have influenced their conclusions ("First, we conducted Harman's single factor test and found that three factors were present . . .also controlled for the effects of common method variance by partialling out a general factor score...we tested a model with a single method factor and examined a null model," p. 1800).

Enhancing Transparency Regarding Measurement

Table 5 provides recommendations on how to enhance transparency regarding measurement. In addition to unambiguous construct definitions, providing information about the psychometric properties of all the measures used (e.g., reliability, construct validity), statistics used to justify aggregation, and issues related to range restriction or measurement error are also important. The types of psychometric properties that need to be reported differ based on the conceptual definition of the construct and the scale (e.g., original versus altered) used. For example, when attempting to measure higher-level constructs, transparency includes identifying the focal unit of analysis, whether it differs from the same construct at the lower level, the statistics used to justify aggregation, and the rationale for choice of statistics used (Castro, 2002; Klein & Kozlowski, 2000; Yammarino et al., 2005). This allows others to more clearly understand the meaning of the focal construct of interest and whether aggregation might have influenced the definition of the construct and meaning of results. Transparency here also alleviates concerns on whether authors cherry-picked aggregation statistics to support their decision to aggregate and, therefore, enhances inferential reproducibility.

Another important measurement transparency consideration is the specific instance when no adequate measure exists for the focal constructs of interest. Questions regarding the impact of measurement error on results or the use of proxies of constructs are even more important when using a new measure or an existing measure that has been altered (Bono & McNamara, 2011; Casper, Eby, Bordeaux, Lockwood, & Lambert, 2007; Zhang & Shaw, 2011). In these instances, transparency includes providing details on changes made to existing scales, such as which items were dropped or added, and any changes in the wording or scale items. Without a clear discussion of the changes made, readers may doubt conclusions, as it might appear that authors changed the scales to obtain the desired results, thereby reducing inferential reproducibility.

The article by Wu, Tsui, and Kinicki (2010) is a good example of transparency regarding score aggregation. Wu et al. (2010) examined the consequences of differentiated leadership within groups and described why aggregation was necessary given their conceptualization of the theoretical relationships ("group-focused leadership fits Chan's (1998) referent shift consensus model in which within-group consensus of lower-level elements is required to form higher-level constructs", p. 95), provided details regarding within- and between-group variability, and reported both within-group inter-rater reliability (r_{wg}) and intraclass correlation coefficient (ICC) statistics. An example of high transparency regarding conceptual definitions, choice of particular indicators, and construct validity is the study by Tashman and Rivera (2016) that used resource dependence and institutional theories to examine how firms respond to ecological uncertainty. Tashman and Rivera (2016) explained why they conceptualized ecological uncertainty as a formative construct ("to capture a resort's total efforts at adopting practices related to ecological mitigation"), and how they assessed face and external validity ("we selected Ski Area Citizens Coalition [SACC] ratings when there was a theoretical basis… we calculated variance inflation factors… assessed multicollinearity at the construct level with a condition index", p. 1513). Finally, they provided evidence of construct validity using correlation tables including all variables.

Enhancing Transparency Regarding Data Analysis

Table 6 provides recommendations on how to enhance transparency regarding data analysis. Given the current level of sophistication of data-analytic approaches, offering detailed recommendations on transparency regarding each of dozens of techniques such as meta-analysis, multilevel modeling, structural equation modeling, computational modeling, content analysis, regression, and many others are outside of the scope of our review. Accordingly, we focus on recommendations that are broad and generally applicable to various types of data-analysis techniques. In addition, Table 6 also includes some more specific recommendations regarding issues that are observed quite frequently—as documented in the articles reviewed in our study.

TABLE 5
Evidence-Based Best-Practice Recommendations to Enhance Methodological Transparency Regarding Measurement

Recommendations	Improves Inferential Reproducibility by Allowing Others to. . . .
1. Provide conceptual definition of construct; report all measures used; how indicators (e.g., reflective, formative) correspond to each construct; and evidence of construct validity (e.g., correlation tables including all variables, results of item and factor analysis) (3, 22, 24, 25, 36, 38, 42, 45, 48, 68, 71, 73, 83, 96)	1. Draw independent conclusions about: a) overall construct validity (i.e., do indicators assess underlying constructs); b) discriminant validity of constructs (i.e., are constructs distinguishable from one another); and c) discriminant validity of measures (i.e., do indicators from different scales overlap). Absent such information, inferences have low reproducibility (e.g., small effects could be attributed to weak construct validity, large effects to low discriminant validity)
2. If scales used were altered, report how and why (e.g., dropped items, changes in item referent). Provide psychometric evidence regarding the altered scales (e.g., criterion-related validity) and report exact items used in the revised scale (3, 22, 28, 31, 38, 95, 96)	2. Understand whether conclusions reached are due to scale alterations (e.g., cherry-picking items); independently reach conclusions about the validity of new and revised scales
3. If scores are aggregated, report measurement variability within and between units of analysis; statistics, if any, used to justify aggregation (e.g. r_{wg}, ICC); and results of all aggregation procedures (32, 50, 70, 72, 88, 95)	3. Reach similar conclusions regarding aggregated construct's meaning because different aggregation indices provide distinct information (e.g., r_{wg} represents interrater agreement, and it is used to justify aggregation of data to higher-level; ICC(1) represents interrater reliability, and provides information on how lower-level data are affected by group membership, and if individuals are substitutable within a group)
4. If range restriction or measurement error was assessed, specify the type of range restriction, and provide rationale for decision to correct or not correct. If corrected, identify how (e.g., type of correction used; sequence; and formulas). Report observed effects and those corrected for range restriction and/or measurement error (29, 44, 58, 83, 90)	4. Recognize how the method used to identify and correct for measurement error or range restriction (e.g., Pearson correlations corrected using the Spearman–Brown formula; Thorndike's Case 2) may have led to over- or under-estimated effect sizes

Notes: r_{wg} = Within-group inter-rater reliability; ICC = Intraclass Correlation Coefficient.

Sources used to derive evidence-based recommendations: [3]Aguinis and Vandenberg (2014), [22]Bono and McNamara (2011), [24]Boyd, Gove, and Hitt (2005), [25]Brannick, Chan, Conway, Lance, and Spector (2010), [28]Brutus, Gill, and Duniewicz (2010), [29]Cafri, Kromrey, and Brannick (2010), [31]Casper, Eby, Bordeaux, Lockwood, and Lambert (2007), [32]Castro (2002), [36]Combs (2010), [38]Conway and Lance (2010), [42]Crook, Shook, Morris, and Madden (2010), [44]Edwards (2008), [45]Edwards and Bagozzi (2000), [48]Evert, Martin, McLeod, and Payne (2016), [50]Feng (2015), [58]Geyskens, Krishnan, Steenkamp, and Cunha (2009), [68]Jackson, Gillaspy, and Purc-Stephenson (2009), [70]Klein and Kozlowski (2000), [71]Leavitt, Mitchell, and Peterson (2010), [72]LeBreton and Senter (2008), [73]Mathieu and Taylor (2006), [83]Scandura and Williams (2000), [88]Smith-Crowe, Burke, Cohen, and Doveh (2014), [90]Van Iddekinge and Ployhart (2008), [95]Yammarino, Dionne, Chun, and Dansereau (2005), [96]Zhang and Shaw (2012).

As noted by Freese (2007b), while researchers today have more degrees of freedom regarding data-analytic choices than ever before, decisions made during analysis are rarely disclosed in a transparent manner. Clearly noting the software employed and making available the syntax used to carry out data analysis facilitates our understanding of how the assumptions of the analytical approach affected results and conclusions (Freese, 2007a; Waldman & Lilienfeld, 2016). For example, there are multiple scripts and packages available within the R software to impute missing data. Two of these (Multiple Imputation by Chained Equations [MICE] and Amelia) impute data assuming that the data are missing at random, while another (missForest) imputes data based on nonparametric assumptions. The manner in which data are imputed influences the data values that are analyzed, which affects results and conclusions. Thus, not knowing which precise package was used contributes to inconsistency in results and in the conclusions others draw about the meaning of results (Frese, 2007b).

Another recommendation in Table 6 relates to the topic of outliers. For example, outliers can affect parameter estimates (e.g., intercept or slope coefficients), but many studies fail to disclose whether a dataset included outliers, what procedures were used to identify and handle them, whether analyses were conducted with and without outliers, and whether results and inferences change based on these decisions (Aguinis et al., 2013). Consequently, low transparency about how outliers were defined, identified, and handled means that other researchers will be unable to reach similar conclusions (i.e., reduced inferential reproducibility).

TABLE 6
Evidence-Based Best-Practice Recommendations to Enhance Methodological Transparency Regarding Data Analysis

Recommendations	Improves Inferential Reproducibility by Allowing Others to. . . .
1. Report specific analytical method used and why it was chosen (e.g., EFA versus CFA; repeated measures ANOVA using conventional univariate tests of significance versus univariate tests with adjusted degrees of freedom) (21, 37, 48, 56, 66, 69, 75, 76, 80, 82, 96)	1. Independently verify whether the data analytical approach used influenced conclusions (e.g., using CFA instead of EFA to generate theory)
2. Report software used, including which version, and make coding rules (for qualitative data) and syntax for data analysis available (8, 14, 21, 35, 54, 55, 60, 63, 79, 91)	2. Check whether assumptions of data analytical procedure within the software used (e.g., REML versus FIML) affects conclusions
3. If tests for outliers were conducted, report methods and decision rules used to identify outliers; steps (if any) taken to manage outliers (e.g., deletion, Winsorization, transformation); the rationale for those steps; and results with and without outliers (2, 12, 40, 58, 60, 68)	3. Infer if substantive conclusions drawn from results (e.g., intercept or slope coefficients; model fit) would differ based on the manner in which outliers were defined, identified, and managed

Notes: EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis; ANOVA = Analysis of Variance; REML = Restricted Maximum Likelihood; FIML = Full Information Maximum Likelihood.

Sources used to derive evidence-based recommendations: [2]Aguinis, Gottfredson, and Joo (2013), [8]Aytug, Rothstein, Zhou, and Kern (2012), [12]Banks et al. (2016a), [14]Bansal and Corley (2011), [21]Bluhm, Harman, Lee, and Mitchell (2011), [35]Chenail (2009), [37]Conway and Huffcutt (2003), [40]Cortina (2003), [48]Evert, Martin, McLeod, and Payne (2016), [54]Freese (2007a), [55]Freese (2007b), [56]Gephart (2004), [58]Geyskens, Krishnan, Steenkamp, and Cunha (2009), [60]Gibbert, Ruigrok, and Wicki (2008), [63]Hair, Sarstedt, Pieper, and Ringle (2012), [66]Hogan, Benjamin, and Brezinski (2000), [68]Jackson, Gillaspy, and Purc-Stephenson (2009), [69]Keselman, Algina, and Kowalchuk (2001), [75]McLeod, Payne, and Evert (2016), [76]Nickerson (2000), [79]Pierce, Block, and Aguinis (2004), [80]Ployhart and Vandenberg (2010), [82]Rogelberg, Adelman, and Askay (2009), [91]Waldman and Lilienfeld (2016), [96]Zhang and Shaw (2012).

An illustration of a more specific and technical recommendation included in Table 6 relates to reporting that a study used a "repeated-measures ANOVA," which does not provide sufficient detail to others about whether the authors used a conventional *F* test, which assumes multisample sphericity, or multivariate *F* tests, which assume homogeneity of between-subjects covariance matrices (Keselman, Algina, & Kowalchuk, 2001). Without such information, which refers to the general issue of providing a clear justification for a particular data-analytic choice, consumers of research attempting to reproduce results using the same general analytical method (e.g., ANOVA) may draw different inferences.

An example of high transparency regarding outliers is the study by Worren, Moore, and Cardona (2002), who examined the relationship between antecedents and outcomes of modular products and process architectures. The authors specified how they identified ("We also examined outliers and influential observations using indicators, such as Cook's distance"), defined ("called up the respondent submitting these data, who said that he had misunderstood some of the questions in that he had considered"), and handled the outlier ("we subsequently corrected this company's score on one variable (product modularity)," p. 1132). A second example, which displays high transparency in data analysis when using a qualitative approach, is Amabile, Barsade, Mueller, and Staw's (2005) study that generated theory regarding how affect relates to creativity at work. The authors detailed the coding rules they used when analyzing events ("A narrative content coding protocol was used to identify indicators of mood and creative thought in the daily diary narratives," p. 378). In addition, they were highly transparent about what they coded to develop measures ("we also constructed a more indirect and less obtrusive measure of mood from the coding of the diary narrative, Coder-rated positive mood. Each specific event described in each diary narrative was coded on a valence dimension", p. 379), and how they coded measures ("defined for coders as "how the reporter [the participant] appeared to feel about the event or view the event"..."For each event, the coder chose a valence code of negative, neutral, positive, or ambivalent," p. 379).

Enhancing Transparency Regarding Reporting Results

Table 7 summarizes recommendations on how to enhance transparency regarding reporting of results. The more transparent authors are in reporting results, the better consumers of published work will be able to reach similar conclusions.

The first issue included in Table 7 relates to the need to provide sufficient detail on response patterns so others can assess how they may have affected inferences drawn from the results. While missing data and nonresponses are rarely the central focus of a study, they usually affect conclusions drawn from the analysis (Schafer & Graham, 2002). Moreover, given the variety of techniques available for dealing with missing data (e.g., deletion, imputation), without precise reporting of results of missing data analysis and the analytical technique used, others are unable to judge whether certain data points were excluded because authors did not have sufficient information or because excluding the incomplete responses supported the authors' preferred hypotheses (Baruch & Holtom, 2008). In short, low inferential reproducibility is virtually guaranteed if this information is absent.

Another issue included in Table 7, and one that ties directly to transparency in data analytical choices, is to report the results of any tests of assumptions that may have been conducted. All analytic techniques include assumptions (e.g., linearity, normality, homoscedasticity, additivity), and many available software packages produce results of assumptions tests without the need for additional calculations on the part of researchers. While the issue of assumptions might be seen as a basic concept that researchers learn as a foundation in many methodology courses, most published articles do not report whether assumptions were assessed (Weinzimmer, Mone, & Alwan, 1994). For example, consider the assumption of normality of distributions, which underlies most regression-based analyses. This assumption is violated frequently (Aguinis, O'Boyle, Gonzalez-Mulé, & Joo, 2016; Crawford, Aguinis, Lichtenstein, Davidsson, & McKelvey, 2015; O'Boyle & Aguinis, 2012) and we suspect that many authors are aware of this. Yet, many researchers continue to use software and analytical procedures that assume normality of data, without openly reporting the results of tests of these assumptions. Reporting results of assumptions increases inferential reproducibility by allowing others to independently assess whether the assumption may have been violated based on an examination of graphs or other information provided in the software output.

Another issue included in Table 7 is the need to report descriptive and inferential statistics clearly and precisely. This precision allows others to independently compare their results and conclusions with those reported. Precision in reporting results allows others to confirm that authors did not hide statistically nonsignificant results behind vague writing and incomplete tables. Perhaps the most egregious example of a lack of precision is with regard to *p*-values. Used to denote the probability that a null hypothesis is tenable, researchers use a variety of arbitrary cutoffs to report *p*-values (e.g., $p < .05$ or $p < .01$) as opposed to the exact *p*-value computed by default by most contemporary statistical packages (Bakker & Wicherts, 2011; Finch, Cumming, & Thomason, 2001; Hoekstra, Finch, Kiers, & Johnson, 2006). Using these artificial cutoffs makes it more difficult to assess whether researchers made errors that obscured the true value of the probability (e.g., rounding 0.059 down to 0.05) or reported a statistically significant result when *p*-values do not match, and to judge the seriousness of making a Type I (incorrectly rejecting the null hypothesis) versus Type II (incorrectly failing to reject the null hypothesis) error within the context of the study (Aguinis et al., 2010b; Wicherts, Veldkamp, Augusteijn, Bakker, Van Aert, & Van Assen, 2016).

An example of high transparency with regard to missing data is the article by Gomulya and Boeker (2014) that examined financial earning restatements and the appointment of CEO successors. In addition to providing information as to why they had missing data ("lack of Securities and Exchange Commission (SEC) filings", p. 1765), the authors also reported how their sample size was affected by missing data ("our sample size experienced the following reduction"), the method used to address missing data ("deleted listwise"), how missing data may have affected results obtained ("model is very conservative, treating any missing data as nonrandom (if they are indeed random, it should not affect the analyses)", and the results of a nonresponse analysis ("we did compare the firms that were dropped for the reasons mentioned above with the remaining firms in terms of size, profitability, and year dummies. We found no significant difference", p. 1784). As a second illustration, Makino and Chan's (2017) paper on how skew and heavy-tail distributions influence firm performance is an example of high transparency regarding the testing of assumptions. In answering their research question, the authors explicitly noted why their data might violate the assumptions of homoscedasticity and independence required by the analytical method (Ordinary Least Squares OLS regression), and outlined the steps they took to account for these violations ("we add a lagged value of the dependent variable (Yi(t −1)) to capture possible persistence

TABLE 7
Evidence-Based Best-Practice Recommendations to Enhance Methodological Transparency Regarding Reporting of Results

Recommendations	Improves Inferential Reproducibility by Allowing Others to . . . .
1. Report results of missing-data analysis (e.g., sensitivity analysis); method (e.g., imputation, deletion) used to address missing data; information (even if speculative) as to why missing data occurred; response rate; and if conducted, results of non-response analyses (11, 15, 44, 68, 74, 82, 84, 92)	1. Have greater confidence in inferences and that authors did not cherry-pick data to support preferred hypotheses; verify whether causes of missing data are related to variables of interest; and independently assess external validity (e.g., if survey respondents are representative of the population being studied)
2. Report results of all tests of assumptions associated with analytical method. Examples include: Normality, heteroscedasticity, independence, covariance amongst levels of repeated-measures, homogeneity of the treatment-difference variances, and group size differences in ANOVA (6, 10, 68, 69, 74, 76)	2. Verify whether possible violations of assumptions influenced study conclusions (e.g., based on chi-square statistic, standard errors) and tests of significance upwards or downwards, thereby affecting inferences drawn
3. Report complete descriptive statistics (e.g., mean, standard deviation, maximum, minimum) for all variables; correlation and (when appropriate) covariance matrices (4, 8, 37, 42, 62, 74, 82, 96)	3. Confirm that results support authors claims (e.g. multicollinearity amongst predictors elevating probability of Type I errors; correlations exceeding maximum possible values); gauge if number of respondents on which study statistics are based was sufficient to draw conclusions
4. Report effect size estimates; CI for point estimates; and information used to compute effect sizes (e. g., within-group variances, degrees of freedom of statistical tests) (1, 4, 9, 10, 36, 40, 44, 49, 52, 57, 64, 67, 78, 91)	4. Interpret accuracy (e.g., width of CI reflects degree of uncertainty of effect size; whether sampling error may have led to inflated effect estimates) and practical significance of study results (i.e., allowing for comparison of estimates across studies)
5. Report exact *p*-values to two decimal places; do not report *p*-values compared to cut-offs (e.g., $p < .05$ or $p < .01$) (4, 9, 10, 20, 52, 61, 64, 91)	5. Rule out whether conclusions are due to chance; judge seriousness of making a Type I (incorrectly rejecting the null hypothesis) versus Type II (incorrectly failing to reject the null hypothesis) error within the context of the study
6. Use specific terms when reporting results. Examples include: • Use "statistically significant" when referring to tests of significance; do not use terms such as "significant", "marginally/almost significant", or "highly significant" (10, 52, 64, 76, 82) • Identify precise estimate used when referring to "effect size" (e.g., Cohen's d, *r*, R^2, η^2, partial η^2, Cramer's V) (5, 51) • Provide interpretations of effect sizes and confidence intervals in regard to context of study, effect size measure used, and other studies examining similar relations. Do not use Cohen's (1988) categories as absolute and context-free benchmarks (1, 3, 4, 5, 8, 10, 20, 26, 29, 30, 46, 51, 52, 65, 67, 76, 91)	6. Understand that statistical significance is related to probability of no relation, not effect size; comprehend what authors mean when referring to "effect size" (e.g., strength of relation or variance explained); interpret evidence by comparing with previous findings and weigh importance of effect size reported in light of context of study and its practical significance
7. Report total number of tests of significance (19, 96)	7. Evaluate conclusions because conducting multiple tests of significance on the same dataset increases the probability of obtaining a statistically significant result solely due to chance, and not because of a substantive relation between constructs
8. Report and clearly identify both unstandardized and standardized coefficients (17, 36, 41, 68, 78)	8. Understand the relative size of the relations examined (when using standardized coefficients); understand the predictive power and substantive impact of the explanatory variables (when using unstandardized coefficients)

Notes: ANOVA = Analysis of Variance; CI = Confidence Interval.

Sources used to derive evidence-based recommendations: [1]Aguinis, Beaty, Boik, and Pierce (2005), [3]Aguinis and Vandenberg (2014), [4]Aguinis, Werner, Abbott, Angert, Park, and Kohlhausen (2010b), [5]Alhija and Levy (2009), [6]Antonakis, Bendahan, Jacquart, and Lalive (2010), [8]Aytug, Rothstein, Zhou, and Kern (2012), [9]Bakker and Wicherts (2011), [10]Balluerka, Gómez, and Hidalgo (2005), [11]Bampton and Cowton (2013), [15]Baruch and Holtom (2008), [17]Bergh et al. (2016), [19]Bettis (2012), [20]Bettis, Ethiraj, Gambardella, Helfat, and Mitchell (2016), [26]Brooks, Dalal, and Nolan (2014), [29]Cafri, Kromrey, and Brannick (2010), [30]Capraro and Capraro (2002), [36]Combs (2010), [37]Conway and Huffcutt (2003), [40]Cortina (2003), [41]Courville and Thompson (2001), [42]Crook, Shook, Morris, and Madden (2010), [44]Edwards (2008), [46]Edwards and Berry (2010), [49]Fan and Thompson (2001), [51]Field and Gillett (2010), [52]Finch, Cumming, and Thomason (2001), [57]Gerber and Malhotra (2008), [61]Goldfarb and King (2016), [62]Haans, Pieters, and He (2016), [64]Hoekstra, Finch, Kiers, and Johnson (2006), [65]Hoetker (2007), [67]Hubbard and Ryan (2000), [68]Jackson, Gillaspy, and Purc-Stephenson, 2009; [69]Keselman, Algina, and Kowalchuk (2001), [74]McDonald and Ho (2002), [76]Nickerson (2000), [78]Petrocelli, Clarkson, Whitmire, and Moon (2013), [82]Rogelberg, Adelman, and Askay (2009), [84]Schafer and Graham (2002), [91]Waldman and Lilienfeld (2016), [92]Werner, Praxedes, and Kim (2007), [96]Zhang and Shaw (2012) (autocorrelation), heteroscedasticity"…"we use the generalized method of moments (GMM)", p. 1732).

DISCUSSION

Our article can be used as a resource to address the "research performance problem" regarding low methodological transparency. Specifically, information in Tables 3–7 can be used for doctoral student training and also for researchers as checklists for how to be more transparent regarding judgment calls and decisions in the theory, design, measurement, analysis, and reporting of results stages of the empirical research process. As such, information in these tables addresses one of the two major determinants of antecedents of this research performance problem: KSAs.

But, as described earlier, even if the necessary knowledge on how to enhance transparency is readily available, authors need to be motivated to use that knowledge. So, to improve authors' motivation to be transparent, Table 8 includes recommendations for journals and publishers, editors, and reviewers on how to make transparency a more salient requirement for publication. Paraphrasing Steve Kerr's (1975) famous article, it would be naïve to hope that authors will be transparent if editors, reviewers, and journals do not reward transparency—even if authors *know* and have the ability to be more transparent.

Given the broad range of recommendations in Table 8, we reiterate that we conceptualize transparency as a continuum and not as a dichotomous variable. In other words, the larger the number of recommendations to enhance methodological transparency that are implemented, the more likely it is that the published study will have greater inferential reproducibility. So, our recommendations in Table 8 are certainly not mutually exclusive. While some address actions to be taken before the submission of a manuscript and information authors must certify during the manuscript submission process, others can be used to give "badges" to accepted manuscripts that are particularly transparent (Kidwell et al., 2016). Moreover, implementing as many of these recommendations as possible will reduce the chance of future retractions and, perhaps, decrease the number of "risky" submissions, thereby lowering the workload and current burden on editors and reviewers.

The vast majority of recommendations in Table 8 can be implemented without incurring much cost, encountering practical hurdles, or fundamentally altering the current manuscript submission and review process. To some extent, the implementation of many of our recommendations is now possible because of the availability of online supplemental files, which removes the important page limitation constraint. For example, because of page limitations, the *Journal of Applied Psychology* (JAP) had a smaller font for the Method section (same smaller size as footnotes) from 1954 to 2007 (Cortina et al., 2017a). In addition, the page limitation constraint may have motivated editors and reviewers to ask authors to omit material from their manuscript, resulting in low transparency for consumers of the research. Also, in our personal experience, members of review teams often require that authors expand upon a study's "contributions to theory" (Hambrick, 2007) at the expense of eliminating information on methodological details (e.g., tests of assumption checks, statistical power analysis, properties of measures). Evidence of this phenomenon is that the average number of pages devoted to the Method and Results sections of articles in JAP remained virtually the same from the year 1994 to the year 2013 (Schmitt, 2017). But, the average number of pages devoted to the Introduction section increased from 2.49 to 3.90, and the Discussion section increased from 1.71 to 2.49 pages (Schmitt, 2017). Again, the availability of online supplements will hopefully facilitate the implementation of many of our recommendations, while being mindful of page limitation constraints.

In hindsight, the implementation of some of the recommendations aimed at enhancing methodological transparency and enhanced inferential reproducibility summarized in Table 8 could have prevented the publication of several articles that have subsequently been retracted. For example, an article submission item requesting that authors acknowledge that all members of the research team had access to the data might have prevented a retraction as in the case of Walumbwa et al. (2011). Reassuringly, many journals are in the process of revising their manuscript submission policies. For example, journals published by the American Psychological Association require that all data in their published articles be an original use (Journal of Applied Psychology, 2017). But, although new policies implemented by journals such as *Strategic Management Journal* (Bettis, Ethiraj, Gambardella, Helfat, & Mitchell, 2016) address important issues about "more appropriate knowledge and norms around the use and interpretation of statistics" (p. 257), most are not directly related to transparency. Thus, although we see some progress in the development of

TABLE 8
Evidence-Based Best-Practice Recommendations for Journals and Publishers, Editors, and Reviewers to Motivate Authors to Enhance Methodological Transparency

Transparency Issue	Recommendations
	Prior to submitting a manuscript for journal publication, authors could certify on the submission form that. . .
Data and syntax availability, data access, and division of labor	1. Data have been provided to journal, along with coding rules (for qualitative data) and syntax used for analyses. If data cannot be made available, require authors to explain why [3, 8, 9, 21, 53, 54, 55, 59, 60, 77, 87, 91]
	2. All authors had access to data, and whether data analysis and results were verified by co-author(s) [9, 77]
Hypothesis testing	3. Authors reported all hypotheses tested, even if they found statistically non-significant results [13, 71]
Power analysis	4. If a power analysis was conducted, results were reported in study [1, 3, 36, 52, 76, 86, 90, 91]
Measures	5. All measures used in the study were reported, along with evidence of validity and reliability [3, 22, 28, 31, 36, 39, 42, 60, 82, 96]
Response rate	6. Response rate for surveys and how missing data was handled were reported [11, 15, 82, 84, 92]
Statistical assumptions	7. Results of tests of assumptions of statistical model were reported [10, 68, 74]
Outliers	8. If tests for outliers were conducted, procedure used to identify and handle them was reported [2, 12, 40, 68]
Control variables	9. Justification of why particular control variables were included or excluded was made explicit, and results were reported with and without control variables [3, 16, 18, 22, 28, 44]
Aggregation	10. If scores were aggregated, variability within and across units of analysis and statistics used to justify aggregations were reported [32, 70, 72, 88, 95]
Effect size and confidence intervals	11. Effect sizes and confidence intervals around point estimates were reported [4, 5, 9, 10, 20, 36, 44, 52, 57, 68, 76, 82, 91]
	12. Implications of observed effect size in context of the study were explained [4, 5, 10, 20, 26, 30, 36, 52, 65, 67]
Reporting of *p*-values	13. Exact *p*-values rather than *p*-values compared to statistical significance cutoffs were reported [9, 20, 52, 61, 64, 91]
Precise use of terms when reporting results	14. Precise, unambiguous terms were used when reporting results. Examples include "split-half" or "coefficient alpha" instead of "internal consistency", and "statistically significant" as opposed to "significant", "highly significant", or "marginally significant" [52, 65, 66, 76, 82]
Limitations	15. The implications of the limitations on the results of the study were made explicit [27, 28, 82]
Post hoc analysis	16. Post hoc analyses were included in a separate section [12, 93]
	Reviewer evaluation forms could be revised to require reviewers to. . .
Competence	17. State level of comfort with and competence to evaluate the methodology used in the manuscript they are reviewing [94]
Limitations	18. Identify limitations that impact the inferential reproducibility of the study [27, 28]
Evaluating transparency	19. Evaluate whether authors have provided all information required on manuscript submission form [8, 13, 21,]
	Review process could be revised by. . .
Alternative/complementary review processes	20. Adopting a two-stage review process by requiring pre-registration of hypotheses, sample size, and data-analysis plan, with the results and discussion sections withheld from reviewers until first revise/reject decision is made [12, 13, 23, 57, 61, 93]
Availability of data	21. Using online supplements that include detailed information such as complete data, coding procedures, specific analyses, and correlation tables [8, 13, 54, 55, 77]
Auditing	22. Instituting a policy where some of the articles published each year are subject to communal audits, with data and programs made available, and commentaries invited for publication [57]

Notes: Sources used to derive evidence-based recommendations: [1]Aguinis, Beaty, Boik, and Pierce (2005), [2]Aguinis, Gottfredson, and Joo (2013), [3]Aguinis and Vandenberg (2014), [4]Aguinis, Werner, Abbott, Angert, Park, and Kohlhausen (2010b), [5]Alhija and Levy (2009), [8]Aytug, Rothstein, Zhou, and Kern (2012), [9]Bakker and Wicherts (2011), [10]Balluerka, Gómez, and Hidalgo (2005), [11]Bampton and Cowton (2013), [12]Banks et al. (2016a), [13]Banks et al. (2016b), [15]Baruch and Holtom (2008), [16]Becker (2005), [18]Bernerth and Aguinis (2016), [20]Bettis, Ethiraj, Gambardella, Helfat, and Mitchell (2016), [21]Bluhm, Harman, Lee, and Mitchell (2011), [22]Bono and McNamara (2011), [23]Bosco, Aguinis, Field, Pierce, and Dalton (2016), [26]Brooks, Dalal, and Nolan (2014), [27]Brutus, Aguinis, and Wassmer (2013), [28]Brutus, Gill, and Duniewicz (2010), [30]Capraro and Capraro (2002), [31]Casper, Eby, Bordeaux, Lockwood, and Lambert (2007), [32]Castro (2002), [36]Combs (2010), [39]Cools, Armstrong, and Verbrighe (2014), [40]Cortina (2003), [42]Crook, Shook, Morris, and Madden (2010), [44]Edwards (2008), [52]Finch, Cumming, and Thomason (2001), [53]Firebaugh (2007), [54]Freese (2007a), [55]Freese (2007b), [57]Gerber and Malhotra (2008), [59]Gibbert and Ruigrok (2010), [60]Gibbert, Ruigrok, and Wicki (2008), [61]Goldfarb and King (2016), [64]Hoekstra, Finch, Kiers, and Johnson (2006), [65]Hoetker (2007), [66]Hogan, Benjamin, and Brezinski (2000), [67]Hubbard and Ryan (2000), [68]Jackson, Gillaspy, and Purc-Stephenson (2009), [70]Klein and Kozlowski (2000), [71]Leavitt, Mitchell, and Peterson (2010), [72]LeBreton and Senter (2008), [74]McDonald and Ho (2002), [76]Nickerson (2000), [77]Nuijten, Hartgerink, Assen, Epskamp, and Wicherts (2016), [82]Rogelberg, Adelman, and Askay (2009), [84]Schafer and Graham (2002), [86]Shen, Kiger, Davies, Rasch, Simon, and Ones (2011), [87]Sijtsma (2016), [88]Smith-Crowe, Burke, Cohen, and Doveh (2014), [90]Van Iddekinge and Ployhart (2008), [91]Waldman and Lilienfeld (2016), [92]Werner, Praxedes, and Kim (2007), [93]Wigboldus and Dotsch (2016), [94]Wright (2016), [95]Yammarino, Dionne, Chun, and Dansereau (2005), [96]Zhang and Shaw (2012).

manuscript submission policies, transparency does not seem to be a central theme to date and, therefore, we believe our review's point of view can be useful and influential in the further refinement of such policies.

Regarding ease of implementation, recommendations #20 and #22 in Table 8 are broader in scope and, admittedly, may require substantial time, effort, and resources: Alternative/complementary review processes and auditing. Some journals may choose to implement these two recommendations but not both given resource constraints. In fact, several journals already offer alternative review processes (e.g., *Management and Organization Review*, *Organizational Research Methods*, and *Journal of Business and Psychology*). The review involves a two-stage process. First, there is a "pre-registration" of hypotheses, sample size, and data-analysis plan. If this pre-registered report is accepted, then authors are invited to submit the full-length manuscript that includes results and discussion sections, and the paper is published regardless of statistical significance and size of effects.

Overall, recommendations in Tables 3–8 are aimed at enhancing methodological transparency by addressing authors' KSAs and motivation to be more transparent when publishing research. While our review highlights how enhancing transparency can improve inferential reproducibility, increased transparency also provides other benefits that strengthen the credibility and trustworthiness of research. First, as mentioned previously, enhanced transparency also improves results reproducibility—the ability of others to reach the same results as the original paper using the data provided by the authors. This allows reviewers and editors to check for errors and inconsistencies in results before articles are accepted for publication, thereby reducing the chances of a later retraction. Second, enhanced transparency can contribute to producing higher-quality studies and in quality control (Chenail, 2009). Specifically, reviewers and editors can more easily evaluate if a study departs substantially from best-practice recommendations regarding particular design, measurement, and analytical processes (Aguinis et al., 2013; Aguinis, Gottfredson, & Wright, 2011; Williams et al., 2009), and judge whether the conclusions authors draw from results are unduly influenced by inaccurate judgment calls and decisions. In addition, when there are grey areas (Tsui, 2013) regarding specific decisions and judgment calls, others are able to better evaluate the authors' decisions and draw independent conclusions about the impact on the study's conclusions. Finally, enhanced transparency improves the replicability of research. Replicability is the ability of others to obtain substantially similar results as the original authors by applying the same steps in a different context and with different data. If there is low methodological transparency, low replicability may be attributed to differences in theorizing, design, measurement, and analysis, rather than substantive differences, thereby decreasing the trust others can place in the robustness of our findings (Bergh et al., 2017a; Cuervo-Cazurra, Andersson, Brannen, Nielsen, & Reuber, 2016).

LIMITATIONS

Even if a journal revised its submission and review polices based on our recommendations, it is possible that some authors may choose to not be truthful on the manuscript submission form and report they did something when they did not or vice versa. Our position is that an important threat to the trustworthiness and credibility of management research is not deliberate actions taken by a small number of deviant individuals who actively engage in fraud and misrepresentation of findings but rather widespread practices whereby researchers do not adequately and fully disclose decisions made when confronted with choices that straddle the line between best-practice recommendations and fabrications (Banks et al., 2016a; Bedeian et al., 2010; Honig, Lampel, Siegel & Drnevich, 2014; Sijtsma, 2016).

We acknowledge that the current reality of business school research emphasizing publications in "A-journals" is a very powerful motivator—one that in some cases may supersede the desire to be transparent about the choices, judgment calls, and decisions involved in the research process. This may be the case even if journals decide to revise manuscript submission policies to enhance transparency. Thus, we see our article as a contributor to improvements, but certainly not a silver-bullet solution, to the research performance problem. Also, our recommendations are broad and did not address, for the most part, detailed suggestions about particular methodological approaches and data-analytic techniques. A future direction for this pursuit would be to provide specific recommendations, especially on the more technical aspects of measurement and analysis.

CONCLUDING REMARKS

In many published articles, what you see is not necessarily what you get. Low methodological transparency or undisclosed actions that take place in the "research kitchen" lead to irreproducible research inferences and noncredible and untrustworthy research conclusions. We hope our recommendations for authors regarding how to be more transparent and for journals and publishers as well as journal editors and reviewers on how to motivate authors to be more transparent will be useful in terms of addressing, at least in part, current questions about the relative lack of transparency, inferential reproducibility, and the trustworthiness and credibility of the knowledge we produce.

REFERENCES

Studies preceded by an asterisk were used to generate recommendations inTables 3–8.

Aguinis, H. 2013. ***Performance management***. Upper Saddle River, NJ: Pearson Prentice Hall.

*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., Boyd, B. K., Pierce, C. A., & Short, J. C. 2011. Walking new avenues in management research methods and theories: Bridging micro and macro domains. ***Journal of Management***, 37: 395–403.

Aguinis, H., Cascio, W. F., & Ramani, R. S. 2017. Science's reproducibility and replicability crisis: International business is not immune. ***Journal of International Business Studies***, 48: 653–663.

Aguinis, H., de Bruin, G. P., Cunningham, D., Hall, N. L., Culpepper, S. A., & Gottfredson, R. K. 2010a. What does not kill you (sometimes) makes you stronger: Productivity fluctuations of journal editors. ***Academy of Management Learning & Education***, 9: 683–695.

Aguinis, H., & Edwards, J. R. 2014. Methodological wishes for the next decade and how to make wishes come true. ***Journal of Management Studies***, 51: 143–174.

*Aguinis, H., Gottfredson, R. K., & Joo, H. 2013. Best-practice recommendations for defining, identifying, and handling outliers. ***Organizational Research Methods***, 16: 270–301.

Aguinis, H., Gottfredson, R. K., & Wright, T. A. 2011. Best-practice recommendations for estimating interaction effects using meta-analysis. ***Journal of Organizational Behavior***, 32: 1033–1043.

Aguinis, H., & O'Boyle, E. H. 2014. Star performers in twenty-first century organizations. ***Personnel Psychology***, 67,: 313–350.

Aguinis, H., O'Boyle, E. H., Gonzalez-Mulé, E., & Joo, H. 2016. Cumulative advantage: Conductors and insulators of heavy-tailed productivity distributions and productivity stars. ***Personnel Psychology***, 69: 3–66.

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

Aguinis, H., Shapiro, D. L., Antonacopoulou, E., & Cummings, T. G. 2014. Scholarly impact: A pluralist conceptualization. ***Academy of Management Learning & Education***, 13: 623–639.

*Aguinis, H., & Vandenberg, R. J. 2014. An ounce of prevention is worth a pound of cure: Improving research quality before data collection. ***Annual Review of Organizational Psychology and Organizational Behavior***, 1: 569–595.

*Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. 2010b. Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. ***Organizational Research Methods***, 13: 515–539.

Aiken, L. S., West, S. G., & Millsap, R. E. 2008. Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. ***American Psychologist***, 63: 32–50.

*Alhija, F. N. A., & Levy, A. 2009. Effect size reporting practices in published articles. ***Educational and Psychological Measurement***, 69: 245–265.

Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. 2005. Affect and creativity at work. ***Administrative Science Quarterly***, 50: 367–403.

*Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. 2010. On making causal claims: A review and recommendations. ***Leadership Quarterly***, 21: 1086–1120.

*Aram, J. D., & Salipante, P. F. 2003. Bridging scholarship in management: Epistemological reflections. ***British Journal of Management***, 14: 189–205.

Ashkanasy, N. M. 2010. Publishing today is more difficult than ever. ***Journal of Organizational Behavior***, 31: 1–3.

*Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. 2012. Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. ***Organizational Research Methods***, 15: 103–133.

Bakker, M., van Dijk, A., & Wicherts, J. M. 2012. The rules of the game called psychological science. ***Perspectives on Psychological Science***, 7: 543–554.

*Bakker, M., & Wicherts, J. M. 2011. The (mis) reporting of statistical results in psychology journals. ***Behavior Research Methods***, 43: 666–678.

*Balluerka, N., Gómez, J., & Hidalgo, D. 2005. The controversy over null hypothesis significance testing revisited. ***Methodology: European Journal of Research Methods for the Behavioral and Social Sciences***, 1: 55–70.

*Bampton, R, & Cowton, C. 2013. Taking stock of accounting ethics scholarship: A review of the journal literature. ***Journal of Business Ethics***, 114: 549–563.

*Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., Abston, K. A., Bennett, A. A., & Adkins, C. L. 2016a. Questions about questionable research practices in the field of management. ***Journal of Management***, 42: 5–20.

*Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. 2016b. Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. ***Journal of Business and Psychology***, 31: 323–338.

*Bansal, P., & Corley, K. 2011. The coming of age for qualitative research: Embracing the diversity of qualitative methods. ***Academy of Management Journal***, 54: 233–237.

*Baruch, Y., & Holtom, B. C. 2008. Survey response rate levels and trends in organizational research. ***Human Relations***, 61: 1139–1160.

*Becker, T. E. 2005. Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. ***Organizational Research Methods***, 8: 274–289.

Bedeian, A. G., Taylor, S. G., & Miller, A. N. 2010. Management science on the credibility bubble: Cardinal sins and various misdemeanors. ***Academy of Management Learning & Education***, 9: 715–725.

Bedeian, A. G., Van Fleet, D. D., & Hyman, H. H. 2009. Scientific achievement and editorial board membership. ***Organizational Research Methods***, 12: 211–238.

*Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Boyd, B. K., Su, P., Lau, C. L. L., & Joo, H. 2016. Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. ***Strategic Management Journal***, 37: 477–497.

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. 2017a. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. ***Strategic Organization***, 15: 423–436.

Bergh, D. D., Sharp, B. M., & Li, M. 2017b. Tests for identifying "red flags" in empirical findings: Demonstration and recommendations for authors, reviewers and editors. ***Academy of Management Learning & Education***, 16: 110–124.

*Bernerth, J. B., & Aguinis, H. 2016. A critical review and best-practice recommendations for control variable usage. ***Personnel Psychology***, 69: 229–283.

*Bettis, R. A. 2012. The search for asterisks: Compromised statistical tests and flawed theories. ***Strategic Management Journal***, 33: 108–113.

*Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. 2016. Creating repeatable cumulative knowledge in strategic management. ***Strategic Management Journal***, 37: 257–261.

*Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. 2011. Qualitative research in management: A decade of progress. ***Journal of Management Studies***, 48: 1866–1891.

Blumberg, M., & Pringle, C. D. 1982. The missing opportunity in organizational research: Some implications for a theory of work performance. ***Academy of Management Review***, 7: 560–569.

*Bono, J. E., & McNamara, G. 2011. From the editors: Publishing in AMJ- Part 2: Research design. ***Academy of Management Journal***, 54: 657–660.

*Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. 2016. HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. ***Personnel Psychology***, 69: 709–750.

*Boyd, B. K., Gove, S., & Hitt, M. A. 2005. Construct measurement in strategic management research: Illusion or reality? ***Strategic Management Journal***, 26: 239–257.

*Brannick, M. T., Chan, D., Conway, J. M., Lance, C. E., & Spector, P. E. 2010. What is method variance and how can we cope with it? A panel discussion. ***Organizational Research Methods***, 13: 407–420.

*Brooks, M. E., Dalal, D. K., & Nolan, K. P. 2014. Are common language effect sizes easier to understand than traditional effect sizes? ***Journal of Applied Psychology***, 99: 332–340.

*Brutus, S., Aguinis, H., & Wassmer, U. 2013. Self-reported limitations and future directions in scholarly reports: Analysis and recommendations. ***Journal of Management***, 39: 48–75.

*Brutus, S., Gill, H., & Duniewicz, K. 2010. State of science in industrial and organizational psychology: A review of self-reported limitations. ***Personnel Psychology***, 63: 907–936.

Butler, N., Delaney, H., & Spoelstra, S. 2017. The grey zone: Questionable research practices in the business school. ***Academy of Management Learning & Education***, 16: 94–109.

Byington, E. K, & Felps, W. 2017. Solutions to the credibility crisis in management science. ***Academy of Management Learning & Education***, 16: 142–162.

*Cafri, G., Kromrey, J. D., & Brannick, M. T. 2010. A meta-meta-analysis: Empirical review of statistical power, type I error rates, effect sizes, and model selection of meta-analyses published in psychology. ***Multivariate Behavioral Research***, 45: 239–270.

*Capraro, R. M., & Capraro, M. M. 2002. Treatments of effect sizes and statistical significance tests in textbooks. ***Educational and Psychological Measurement***, 62: 771–782.

*Casper, W. J., Eby, L. T., Bordeaux, C., Lockwood, A., & Lambert, D. A. 2007. Review of research methods in IO/OB work–family research. ***Journal of Applied Psychology***, 92: 28–43.

*Castro, S. L. 2002. Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, rwg(j), hierarchical linear modeling, within-and between-analysis, and random group resampling. ***Leadership Quarterly***, 13: 69–93.

Chan, D. 1998. Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. ***Journal of Applied Psychology***, 83: 234–246.

*Chandler, G. N., & Lyon, D. W. 2001. Issues of research design and construct measurement in entrepreneurship research: The past decade. ***Entrepreneurship: Theory and Practice***, 25: 101–114.

*Chang, S., van Witteloostuijn, A., & Eden, L. 2010. From the editors: Common method variance in international business research. ***Journal of International Business Studies***, 41: 178–184.

*Chenail, R. J. 2009. Communicating your qualitative research better. ***Family Business Review***, 22: 105–108.

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

*Combs, J. G. 2010. Big samples and small effects: Let's not trade relevance and rigor for power. ***Academy of Management Journal***, 53: 9–13.

*Conway, J. M., & Huffcutt, A. I. 2003. A review and evaluation of exploratory factor analysis practices in organizational research. ***Organizational Research Methods***, 6: 147–168.

*Conway, J. M., & Lance, C. E. 2010. What reviewers should expect from authors regarding common method bias in organizational research. ***Journal of Business & Psychology***, 25: 325–334.

*Cools, E., Armstrong, S. J., & Verbrigghe, J. 2014. Methodological practices in cognitive style research: Insights and recommendations from the field of business and psychology. ***European Journal of Work & Organizational Psychology***, 23: 627–641.

Corley, K. G., & Schinoff, B. S. 2017. Who, me? An inductive study of novice experts in the context of how editors come to understand theoretical contribution. ***Academy of Management Perspectives***, 31: 4–27.

*Cortina, J. M. 2003. Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. ***Organizational Research Methods***, 6: 415–439.

Cortina, J. M., Aguinis, H., & DeShon, R. P. 2017a. Twilight of dawn or of evening? A century of research methods in the Journal of Applied Psychology. ***Journal of Applied Psychology***, 102: 274–290.

Cortina, J. M., Green, J. P., Keeler, K. R., & Vandenberg, R. J. 2017b. Degrees of freedom in SEM: Are we testing the models that we claim to test? ***Organizational Research Methods***, 20: 350–378.

*Courville, T., & Thompson, B. 2001. Use of structure coefficients in published multiple regression articles: β is not enough. ***Educational and Psychological Measurement***, 61: 229–248.

Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P., & McKelvey, B. 2015. Power law distributions in entrepreneurship: Implications for theory and research. ***Journal of Business Venturing***, 30: 696–713.

*Crook, T. R., Shook, C. L., Morris, M. L., & Madden, T. M. 2010. Are we there yet? An assessment of research design and construct measurement practices in entrepreneurship research. ***Organizational Research Methods***, 13: 192–206.

Cuervo-Cazurral, A., Andersson, U., Brannen, M. Y., Nielsen, B. B., & Reuber, A. R. 2016. From the editors: Can I trust your findings? Ruling out alternative explanations in international business research. ***Journal of International Business Studies***, 47: 881–897.

Davis, G. F. 2015. What is organizational research for? ***Administrative Science Quarterly***, 60: 179–188.

*Dionne, S. D., Gupta, A., Sotak, K. L., Shirreffs, K. A., Serban, A., Hao, C., Kim, D. H., & Yammarino, F. J. 2014. A 25-year perspective on levels of analysis in leadership research. ***Leadership Quarterly***, 25: 6–35.

*Edwards, J. R. 2008. To prosper, organizational psychology should. . . overcome methodological barriers to progress. ***Journal of Organizational Behavior***, 29: 469–491.

*Edwards, J. R., & Bagozzi, R. P. 2000. On the nature and direction of relationships between constructs and measures. ***Psychological Methods***, 5: 155–174.

*Edwards, J. R., & Berry, J. W. 2010. The presence of something or the absence of nothing: Increasing theoretical precision in management research. ***Organizational Research Methods***, 13: 668–689.

*Eisenhardt, K. M., & Graebner, M. E. 2007. Theory building from cases: Opportunities and challenges. ***Academy of Management Journal***, 50: 25–32.

*Evert, R. E., Martin, J. A., McLeod, M. S., & Payne, G. T. 2016. Empirics in family business research. ***Family Business Review***, 29: 17–43.

*Fan, X., & Thompson, B. 2001. Confidence intervals for effect sizes confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. ***Educational and Psychological Measurement***, 61: 517–531.

*Feng, G. C. 2015. Mistakes and how to avoid mistakes in using intercoder reliability indices. ***Methodology: European Journal of Research Methods for the Behavioral and Social Sciences***, 11: 13–22.

*Field, A. P., & Gillett, R. 2010. How to do a meta-analysis. ***British Journal of Mathematical and Statistical Psychology***, 63: 665–694.

*Finch, S., Cumming, G., & Thomason, N. 2001. Colloquium on effect sizes: The roles of editors, textbook authors, and the publication manual reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. ***Educational and Psychological Measurement***, 61: 181–210.

*Firebaugh, G. 2007. Replication data sets and favored-hypothesis bias: Comment on Jeremy Freese (2007) and Gary King (2007). ***Sociological Methods & Research***, 36: 200–209.

Fisher, G., & Aguinis, H. 2017. Using theory elaboration to make theoretical advancements. ***Organizational Research Methods***, 20: 438–464.

*Freese, J. 2007a. Overcoming objections to open-source social science. ***Sociological Methods & Research***, 36: 220–226.

*Freese, J. 2007b. Replication standards for quantitative social science: Why not sociology. ***Sociological Methods & Research***, 36: 153–172.

George, G. 2014. Rethinking management scholarship. ***Academy of Management Journal***, 57: 1–6.

*Gephart, R. P. 2004. Qualitative research and the Academy of Management Journal. ***Academy of Management Journal***, 47: 454–462.

*Gerber, A. S., & Malhotra, N. 2008. Publication bias in empirical sociological research: Do arbitrary significance levels distort published results? ***Sociological Methods & Research***, 37: 3–30.

*Geyskens, I., Krishnan, R., Steenkamp, J. M., & Cunha, P. V. 2009. A review and evaluation of meta-analysis practices in management research. ***Journal of Management***, 35: 393–419.

*Gibbert, M., & Ruigrok, W. 2010. The 'what' and 'how' of case study rigor: Three strategies. ***Organizational Research Methods***, 13: 710–737.

*Gibbert, M., Ruigrok, W., & Wicki, B. 2008. What passes as a rigorous case study? ***Strategic Management Journal***, 29: 1465–1474.

*Goldfarb, B., & King, A. A. 2016. Scientific apophenia in strategic management research: Significance tests & mistaken inference. ***Strategic Management Journal***, 37: 167–176.

Gomulya, D., & Boeker, W. 2014. How firms respond to financial restatement: CEO successors and external reactions. ***Academy of Management Journal***, 57: 1759–1785.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. 2016. What does research reproducibility mean? ***Science Translational Medicine***, 8: 341–353.

Grand, J. A., Rogelberg, S. G., Allen, T. D., Landis, R. S., Reynolds, D. H., Scott, J. C., Tonidandel, S., & Truxillo, D. M. in press. A systems-based approach to fostering robust science in industrial-organizational psychology. ***Industrial and Organizational Psychology: Perspectives on Science and Practice***.

Green, J. P., Tonidandel, S., & Cortina, J. M. 2016. Getting through the gate: Statistical and methodological issues raised in the reviewing process. ***Organizational Research Methods***, 19: 402–432.

*Haans, R. F., Pieters, C., & He, Z. L. 2016. Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. ***Strategic Management Journal***, 37: 1177–1195.

*Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. 2012. The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. ***Long Range Planning***, 45: 320–340.

Hambrick, D. C. 2007. The field of management's devotion to theory: Too much of a good thing? ***Academy of Management Journal***, 50: 1346–1352.

Hitt, M. A., Beamish, P. W., Jackson, S. E., & Mathieu, J. E. 2007. Building theoretical and empirical bridges across levels: Multilevel research in management. ***Academy of Management Journal***, 50: 1385–1399.

*Hoekstra, R., Finch, S., Kiers, H. L., & Johnson, A. 2006. Probability as certainty: Dichotomous thinking and the misuse of p values. ***Psychonomic Bulletin & Review***, 13: 1033–1037.

*Hoetker, G. 2007. The use of logit and probit models in strategic management research: Critical issues. ***Strategic Management Journal***, 28: 331–343.

*Hogan, T. P., Benjamin, A., & Brezinski, K. L. 2000. Reliability methods: A note on the frequency of use of various types. ***Educational and Psychological Measurement***, 60: 523–531.

Hollenbeck, J. R., & Wright, P. M. 2017. Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. ***Journal of Management***, 43: 5–18.

Honig, B., Lampel, J., Siegel, D., & Drnevich, P. 2014. Ethics in the production and dissemination of management research: Institutional failure or individual fallibility? ***Journal of Management Studies***, 51: 118–142.

*Hubbard, R., & Ryan, P. A. 2000. Statistical significance with comments by editors of marketing journals: The historical growth of statistical significance testing in psychology—and its future prospects. ***Educational and Psychological Measurement***, 60: 661–681.

Hunton, J. E., & Rose, J. M. 2011. Retraction. Effects of anonymous whistle-blowing and perceived reputation threats on investigations of whistle-blowing allegations by audit committee Members. ***Journal of Management Studies***, 48: 75–98. (Retraction published 2015, Journal of Management Studies, 52, 849).

*Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. 2009. Reporting practices in confirmatory factor analysis: An overview and some recommendations. ***Psychological Methods***, 14: 6–23.

John, L. K., Loewenstein, G., & Prelec, D. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. ***Psychological Science***, 23: 524–532.

Joo, H., Aguinis, H., & Bradley, K. J. 2017. Not all non-normal distributions are created equal: Improved theoretical and measurement precision. ***Journal of Applied Psychology***, 102: 1022–1053.

Journal of Applied Psychology. 2017. Data transparency policy. http://www.apa.org/pubs/journals/apl/?tab=4. Accessed March 7, 2017.

Karabag, S. F., & Berggren, C. 2016. Misconduct, marginality and editorial practices in management, business and economics journals. ***PloS One***, 11: e0159492.

Kerr, S. 1975. On the folly of rewarding A, while hoping for B. ***Academy of Management Journal***, 18: 769–783.

*Keselman, H. J., Algina, J., & Kowalchuk, R. K. 2001. The analysis of repeated measures designs: A review. ***British Journal of Mathematical and Statistical Psychology***, 54: 1–20.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. 2016. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. ***PLoS Biol***, 14: e1002456.

*Klein, K. J., & Kozlowski, S. W. 2000. From micro to meso: Critical steps in conceptualizing and conducting multilevel research. ***Organizational Research Methods***, 3: 211–236.

*Leavitt, K., Mitchell, T. R., & Peterson, J. 2010. Theory pruning: Strategies to reduce our dense theoretical landscape. ***Organizational Research Methods***, 13: 644–667.

*LeBreton, J. M., & Senter, J. L. 2008. Answers to 20 questions about interrater reliability and interrater agreement. ***Organizational Research Methods***, 11: 815–852.

Lichtenthaler, U. 2008. Retraction. Externally commercializing technology assets: An examination of different process stages. ***Journal of Business Venturing***, 23: 445–464 (Retraction published 2013, Journal of Business Venturing, 28, 581).

Maier, N. R. F. 1955. ***Psychology in industry: A psychological approach to industrial problems***. Boston, MA: Houghton-Mifflin.

Maitlis, S. 2005. The social processes of organizational sensemaking. ***Academy of Management Journal***, 48: 21–49.

Makino, S., & Chan, C. M. 2017. Skew and heavy-tail effects on firm performance. ***Strategic Management Journal***, 38: 1721–1740.

Mathieu, J. E., & Chen, G. 2011. The etiology of the multilevel paradigm in management research. ***Journal of Management***, 37: 610–641.

*Mathieu, J. E, & Taylor, S. R. 2006. Clarifying conditions and decision points for mediational type inferences in organizational behavior. ***Journal of Organizational Behavior***, 27: 1031–1056.

*McDonald, R. P., & Ho, M. H. R. 2002. Principles and practice in reporting structural equation analyses. ***Psychological Methods***, 7: 64–82.

*McLeod, M., Payne, G., & Evert, R. 2016. Organizational ethics research: A systematic review of methods and analytical techniques. ***Journal of Business Ethics***, 134: 429–443.

Murphy, K. R., & Aguinis, H. 2017. How badly can cherry picking and question trolling produce bias in published results? Determinants and size of different forms of HARKing. Working paper. ***Journal of Business and Psychology***.

Min, J., & Mitsuhashi, H. 2012. Retraction. Dynamics of unclosed triangles in alliance networks: Disappearance of brokerage positions and performance consequences. ***Journal of Management Studies*** 49: 1078–1108 (Retraction published 2016, Journal of Management Studies, 51, 868).

*Nickerson, R. S. 2000. Null hypothesis significance testing: A review of an old and continuing controversy. ***Psychological Methods***, 5: 241–301.

Nosek, B. A., Spies, J. R., & Motyl, M. 2012. Scientific utopia II: Restructuring incentives and practices to promote truth over publishability. ***Perspectives on Psychological Science***, 7: 615–631.

*Nuijten, M. B., Hartgerink, C. H., Assen, M. A., Epskamp, S., & Wicherts, J. M. 2016. The prevalence of statistical reporting errors in psychology (1985–2013). ***Behavior Research Methods***, 48: 1205–1226.

O'Boyle, E. H, & Aguinis, H. 2012. The best and the rest: Revisiting the norm of normality of individual performance. ***Personnel Psychology***, 65: 79–119.

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. 2017. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. ***Journal of Management***, 43: 376–399.

*Petrocelli, J. V., Clarkson, J. J., Whitmire, M. B., & Moon, P. E. 2013. When ab≠ c–c′: Published errors in the reports of single-mediator models. ***Behavior Research Methods***, 45: 595–601.

*Pierce, C. A., Block, R. A., & Aguinis, H. 2004. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. ***Educational and Psychological Measurement***, 64: 916–924.

*Ployhart, R. E., & Vandenberg, R. J. 2010. Longitudinal research: The theory, design, and analysis of change. ***Journal of Management***, 36: 94–120.

*Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. ***Journal of Applied Psychology***, 88: 879–903.

*Rogelberg, S., Adelman, M., & Askay, D. 2009. Crafting a successful manuscript: Lessons from 131 reviews. ***Journal of Business & Psychology***, 24: 117–121.

Rousseau, D. M. 1985. Issues of level in organizational research: Multi-level and cross-level perspectives. ***Research in Organizational Behavior***, 7: 1–37.

*Scandura, T. A., & Williams, E. A. 2000. Research methodology in management: Current practices, trends, and implications for future research. ***Academy of Management Journal***, 43: 1248–1264.

*Schafer, J. L., & Graham, J. W. 2002. Missing data: Our view of the state of the art. ***Psychological Methods***, 7: 147–177.

Schmitt, N. 2017. Reflections on the Journal of Applied Psychology for 1989 to 1994: Changes in major research themes and practices over 25 years. ***Journal of Applied Psychology***, 102: 564–568.

*Schriesheim, C. A., Castro, S. L., Zhou, X. T., & Yammarino, F. J. 2002. The folly of theorizing "A" but testing "B": A selective level-of-analysis review of the field and a detailed leader–member exchange illustration. ***Leadership Quarterly***, 12: 515–551.

Schwab, A., & Starbuck, W. 2017. A call for openness in research reporting: How to turn covert practices into helpful tools. ***Academy of Management Learning & Education***, 16: 124–141.

*Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. 2011. Samples in applied psychology: Over a decade of research in review. ***Journal of Applied Psychology***, 96: 1055–1064.

*Sijtsma, K. 2016. Playing with data: Or how to discourage questionable research practices and stimulate researchers to do things right. ***Psychometrika***, 81: 1–15.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. ***Psychological Science***, 22: 1359–1366.

*Smith-Crowe, K., Burke, M. J., Cohen, A., & Doveh, E. 2014. Statistical significance criteria for the rwg and average deviation interrater agreement indices. ***Journal of Applied Psychology***, 99: 239–261.

Stapel, D. A., & Semin, G. R. 2007. Retraction. The magic spell of language: Linguistic categories and their perceptual consequences. ***Journal of Personality & Social Psychology***, 93: 23–33. (Retraction published 2013, Journal of Personality & Social Psychology, 104, 197).

Tashman, P., & Rivera, J. 2016. Ecological uncertainty, adaptation, and mitigation in the US ski resort industry: Managing resource dependence and institutional pressures. ***Strategic Management Journal***, 37: 1507–1525.

Tett, R. P., Walser, B., Brown, C., Simonet, D. V., & Tonidandel, S. 2013. The 2011 SIOP graduate program benchmarking survey part 3: Curriculum and competencies. ***The Industrial-Organizational Psychologist***, 50: 69–89.

Tsai, W. C., Chen, C. C., & Liu, H. L. 2007. Test of a model linking employee positive moods and task performance. ***Journal of Applied Psychology***, 92: 1570–1583.

Tsui, A. S. 2013. The spirit of science and socially responsible scholarship. ***Management and Organization Review***, 9: 375–394.

*van Aken, J. E. 2004. Management research based on the paradigm of the design sciences: The quest for field-tested and grounded technological rules. ***Journal of Management Studies***, 41: 219–246.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. 2017. A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. ***Journal of Management***. Advance online publication. doi: https://doi.org/10.1177/0149206317702220

*Van Iddekinge, C. H., & Ployhart, R. E. 2008. Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. ***Personnel Psychology***, 61: 871–925.

Vroom, V. H. 1964. ***Work and motivation***. New York: Wiley.

*Waldman, I. D., & Lilienfeld, S. O. 2016. Thinking about data, research methods, and statistical analyses: Commentary on Sijtsma's (2014) "Playing with Data". ***Psychometrika***, 81: 16–26.

Walumbwa, F. O., Luthans, F., Avey, J. B., & Oke, A. 2011. Retraction. Authentically leading groups: The mediating role of collective psychological capital and trust. ***Journal of Organizational Behavior***, 32: 4–24 (Retraction published 2014, Journal of Organizational Behavior, 35, 746).

Weinzimmer, L. G., Mone, M. A., & Alwan, L. C. 1994. An examination of perceptions and usage of regression diagnostics in organization studies. ***Journal of Management***, 20: 179–192.

*Werner, S., Praxedes, M., & Kim, H. G. 2007. The reporting of nonresponse analyses in survey research. ***Organizational Research Methods***, 10: 287–295.

Wicherts, J. M., Bakker, M., & Molenaar, D. 2011. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. ***PloS one***, 6, e26828.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R. C., & Van Assen, M. A. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. ***Frontiers in Psychology***, 7: 1832.

*Wigboldus, D. H., & Dotsch, R. 2016. Encourage playing with data and discourage questionable reporting practices. ***Psychometrika***, 81: 27–32.

Williams, H. M., Parker, S. K., & Turner, N. 2010. Proactively performing teams: The role of work design, transformational leadership, and team composition. ***Journal of Occupational and Organizational Psychology***, 83: 301–324.

Williams, L. J., Vandenberg, R. J., & Edwards, J. R. 2009. Structural equation modeling in management research: A guide for improved analysis. ***Academy of Management Annals***, 3: 543–604.

Worren, N., Moore, K., & Cardona, P. 2002. Modularity, strategic flexibility, and firm performance: A study of the home appliance industry. ***Strategic Management Journal***, 23: 1123–1140.

*Wright, P. M. 2016. Ensuring research integrity: An editor's perspective. ***Journal of Management***, 42: 1037–1043.

Wu, J. B., Tsui, A. S., & Kinicki, A. J. 2010. Consequences of differentiated leadership in groups. ***Academy of Management Journal***, 53: 90–106.

*Yammarino, F. J., Dionne, S. D., Chun, J. U., & Dansereau, F. 2005. Leadership and levels of analysis: A state-of-the-science review. ***Leadership Quarterly***, 16: 879–919.

*Zhang, Y., & Shaw, J. D. 2012. From the editors: Publishing in AMJ-part 5: Crafting the methods and results. ***Academy of Management Journal***, 55: 8–12.

Zhu, H., & Yoshikawa, T. 2016. Contingent value of director identification: The role of government directors in monitoring and resource provision in an emerging economy. ***Strategic Management Journal***, 37: 1787–1807.

Herman Aguinis (haguinis@gwu.edu; www.hermanaguinis.com) is the Avram Tucker Distinguished Scholar and Professor of Management at George Washington University School of Business. His research is interdisciplinary and addresses the acquisition and deployment of talent in organizations and organizational research methods. He has published about 150 academic articles and five books and received numerous awards including the Academy of Management (AOM) Research Methods Division Distinguished Career Award for lifetime contributions, Michael R. Losey Human Resource Research Award by Society for Human Resource Management Foundation for lifetime achievement in human resource management research, and best-article-of-the-year awards from Journal of Management, Personnel Psychology, Organizational Research Methods, Journal of Organizational Behavior, and Academy of Management Perspectives. He is a Fellow of AOM (also serving as AOM Fellows Deputy Dean), the American Psychological Association, the Association for Psychological Science, and the Society for Industrial and Organizational Psychology and currently serves on the editorial board of 14 journals.

Ravi S. Ramani is a PhD Candidate in the Department of Management at the George Washington University School of Business. His research focuses on employees' emotional experience of work, organizational research methods, and meaningful scholarship. His research has been published in Academy of Management Annals, Journal of International Business Studies, and Industrial and Organizational Psychology: Perspectives on Science and Practice. He received a Best Paper award from the Academy of Management Entrepreneurship Division, and has delivered has delivered several presentations at the annual meetings of the Academy of Management, Society for Industrial and Organizational Psychology, and other conferences.

Nawaf Alabduljader is a PhD Candidate in the Department of Management at George Washington University School of Business and will join Kuwait University as an assistant professor in September 2018. His research focuses on entrepreneurship and human resource management. His work has been published in peer-reviewed journals, edited books, and conference proceedings and he has delivered several presentations at the annual meetings of the Academy of Management and several other conferences.