Prestige Does not Equal Quality: Lack of Research Quality in High-Prestige Journals

Tine Köhler
The University of Melbourne

Justin A. DeSimone
The University of Alabama

Jeremy L. Schoen
The University of Auckland

Author note: All authors contributed equally to the work on this paper.
Prestige Does not Equal Quality: Lack of Research Quality in High-Prestige Journals

Publishing in high-prestige journals is a top priority on most researchers’ to do list, especially if they are early in their career, want to get or keep a job, get tenure, get promoted, or build a reputation in their field. This seems to be a sound strategy as one might say that selection, promotion, and tenure committees always count, but they don’t always read (e.g., Aguinis, Cummings, Ramani, & Cummings, 2019). I.e., the criteria by which a researcher’s output and quality of work is assessed is by the journals in which they have published and by the impact factors that these journals have accrued at the time of the author’s publication, rather than by the actual content of the work (e.g., Addo, 2017). Publishing in “top” journals is further encouraged (financially or in other ways) by department chairs and deans as some university rankings depend heavily on how often faculty members publish in them (UTD, 2020).

One of the underlying assumptions in the practice of evaluating the quality of a researcher (or university) by the citation count of the journals they publish in is the perception that high-prestige journals only accept and publish work of a high quality and that they do not publish lower quality work. Publishing in high-prestige journals is certainly difficult. Space for published articles in each of these journals is limited, and strong incentive structures exist for researchers around the world to target the same small number of journals. As a consequence, academics compete for a limited resource, a publication in a coveted high-prestige journal. This is likely to create an incentive for researchers to do whatever it takes to get in.

Over the last 10 years, an increasing amount of research has documented that competition for journal space results in a range of undesirable research practices. Our main contention in this commentary is that the practice of equating impact factors with prestige and consequently coercing researchers to (only) publish in the top journals subsequently contributes
to a potential decrement in the quality of the research submitted to and published in these journals (i.e., by cutting corners and engaging in questionable research practices (QRPs)). If we assume that the work published in high-prestige journals must be of high-quality, then we are pulling the wool over our own eyes as to the kinds of QRPs researchers are willing to engage in when their career and livelihood depend on it. In this commentary, we (a) provide an overview of the kinds of research practices that previous research has found to be common in top journals, yet are detrimental to research quality and (b) question at a very fundamental level if the research published in these journals has any value at all beyond providing a boost to the author’s career.

Unrealistic theoretical models – Feeding the beasts called novelty and theoretical contribution

Many of the high-prestige journals identified by Highhouse, Zickar, and Melick (e.g., JAP, AMJ, ASQ) strongly value novelty and theoretical contribution. For example, excerpts from the mission statements of some of the top journals read:

*To be published in AMJ, the research must make strong empirical and theoretical contributions [...]. Authors should strive to produce original, insightful, interesting, important, and theoretically bold research that demonstrates a significant "value-added" contribution to the field's understanding of an issue or topic. (AMJ, 2020)*

*ASQ seeks to advance the understanding of management, organizations, and organizing in a wide variety of contexts, including teams, business and nonprofit organizations, government agencies, and markets. Contributions to this enterprise can include the discovery and analysis of new phenomena, new theoretical accounts informed by empirical analysis, or the disconfirmation of existing theory. (ASQ, 2020)*

*The Journal of Applied Psychology® emphasizes the publication of original investigations that contribute new knowledge and understanding to fields of applied psychology [...]. The journal publishes several types of articles, including: Rigorously conducted empirical investigations that extend conceptual understanding (original investigations or meta-analyses); Theory development articles as well as integrative conceptual reviews that synthesize literature and create new theory of psychological phenomena that will stimulate novel research; Rigorously conducted qualitative*
research on phenomena that are difficult to capture with quantitative methods, or on phenomena that warrant inductive theory building. (JAP, 2020)

This focus on novelty and theoretical contribution leads researchers to add constructs, add mediating paths, search for new boundary conditions, or simply include “more” than previous models did. The strong focus on theoretical novelty and contribution is also reflected in the reluctance of most high prestige journals (and their reviewers) to publish replications (Hunter, 2001), leading to a perceived need to demonstrate something new with every paper.

Along these lines, Saylors and Trafimow (in press) survey AMJ, OBHDP, and ASQ (each of which was in the top quartile (and top 8) of the journals as ranked in the focal article) over a decade beginning in 2008 and demonstrate that researchers are publishing increasingly complex models that consist of more boxes and arrows than ever before. Saylors and Trafimow then show that as model complexity increases, the probability of model veracity decreases. They conclude based on their data that what is published in the top journals is likely not actually a true representation of reality, which in turn undoubtedly contributes to the replicability crisis in and beyond our field (e.g., Christensen & Miguel, 2018; Ioannidis, 2005).

Reviewers and editors are complicit in this push for increased model complexity, as high-prestige journals seem to value complex research designs, including multi-wave, multi-source, and multi-level data collection. At first glance, these seem to be good things as this kind of data is appropriate when studying complex phenomena or when one wants to demonstrate causality. Furthermore, this kind of data, if appropriately and rigorously collected, can provide stronger evidence for a proposed theoretical model. However, it is rare that such data can be obtained in organizational settings.

Complex research designs are tough to execute for research participants and for researchers. Participants do not like to participate in studies driven by complex designs as
participation in them often requires more time and effort due to multiple data collections and complex measurement, such as convincing a supervisor or co-worker to complete a paired assessment or simply filling out more measures for more variables included in the underlying model. This can lead to respondent fatigue and apathy in responding (e.g., Burchell & March, 1992). This also means that the kind of research participants we really want to get (employees, managers, people in high-stress situations, etc.) are less likely to agree to participate in studies based on complex designs because of the effort and time involved. Furthermore, organizations may not allow us access to their employees if the amount of time required to participate in the study is too extensive and will interfere too much with people’s work time. As such, many complex models are tested on populations that are more easily available but have lower fidelity, for example, student or MTurk samples. Even when such models are tested in organizational samples, the measurement and research design often involve sacrificing quality to achieve brevity (e.g., shortened scales or unvalidated ad hoc measures, see Heggestad et al., 2019).

For researchers, these designs are similarly demanding and often require collecting more data. As an example, in the case of multilevel data collections, sample size can quickly increase to the equivalent of months of data collection, especially if these data collections should happen in organizations. Multi-wave and longitudinal designs require a lot of time, as several data collection points with time lag have to be implemented. This means that conducting a study with a complex design may involve risking a job or a promotion if it takes too long to conduct and subsequently publish. In addition, these highly complex designs need a high level of research methods skills. So, authors with more experience, certain research methods skill sets, and training in specific research traditions have an advantage in getting published in high-prestige journals. Consequently, authors need to invest a lot of effort and time into training up
on these designs. This would not be as big a problem were it not for the fact that publication in high-prestige journals is especially important for junior career academics.

In short, the investment for researchers is very big. If at the end of a complex study the obtained data does not support the theoretical model one has proposed, the risk is that considerable valuable research time has been lost and that one might not be employed for much longer. Given the pending doom that ensues from the required complex theoretical models and research designs necessitated by the top journals and from the very serious publication pressures to keep one’s job, researchers are motivated to find ways to make it work. As existing literature implies, there are many ways in which researchers can “make their data work.”

**Trade-offs and QRPs**

Organizational scholarship lacks a “gold standard” methodology (as opposed to, e.g., double-blind randomized control trials in medicine, see Moher et al., 1995). Instead, our field often relies on correlational or quasi-experimental design, and our measurement often lacks precision. The phenomena we study and the populations in which we are interested often force organizational researchers to consider methodological trade-offs. For example, researchers who use experience sampling may rely on shorter scales to reduce memory or fatigue effects. Our field has also increasingly turned to online data collection to provide data quickly from individuals across multiple organizations (Porter, Outlaw, Gale, & Cho, 2019), but a lack of direct participant contact increases concerns about data quality and participant honesty (Chandler & Paolacci, 2017; Harms & DeSimone, 2015).

There is nothing inherently wrong with making well-reasoned decisions regarding trade-offs in research. Researchers tend to think positively about their own work and may genuinely believe that what they gain in a trade-off is worth far more than what they may lose. No
research design is perfect, and researchers may have excellent reasons for making these decisions. However, this practice becomes problematic when researchers fail to disclose the decisions they make and their associated alternatives or potential consequences. This practice may be exacerbated by the “publish or perish” culture and reward systems that value successful publication in high-prestige journals above the conduct and reporting of high-quality scholarship. To increase the probability of successful publication, researchers may neglect to disclose the negative aspects of such trade-offs. When alternatives or potential consequences are mentioned, they are typically relegated to a cursory mention in a “limitations” section. Failing to present readers with sufficient information (or burying it) runs counter to recent laudable efforts to improve transparency in organizational scholarship (Banks et al., 2018).

A lack of transparency in published research is one of the culprits in the aforementioned replicability crisis (Cramerer et al., 2016; Hunter, 2001; OSC, 2015). However, opaque research practices are only one of many potential problems with the current system. Even well-meaning researchers may inadvertently fail to disclose details or remove them due to space constraints. Less scrupulous researchers may find more creative ways to game the system. There is also a growing awareness of QRPs that pervade organizational scholarship (though the problem is certainly not unique to our field; see Neuroskeptic, 2012).

For quite some time, researchers have been aware of (and bemoaned) the tendency for journals to favor the publication of statistically significant results over null results (Rosenthal, 1979). Motivated by the existing incentive structure and armed with the knowledge that it helps to be “right” if one wants to be published, some researchers may attempt to pass off exploratory practices as confirmatory through practices such as p-hacking (Head, Holman, Lanfear, Kahn, & Jennions, 2015; Nuzzo, 2014) or HARKing (Murphy & Aguinis, 2019; Rubin, 2017).
However, a lack of disclosure is far from the only transgression committed by organizational scholars in the pursuit of publication. Researchers have also documented the remarkable metamorphosis of results from non-significant to significant between the stages of dissertation and publication (O’Boyle, Banks, & Gonzales-Mulé, 2017). Others have noted that the models reported in top journals commonly misreport degrees of freedom for the models they test (Cortina, Green, Keeler, & Vandenberg, 2017, Credé & Harms, 2015). Each of these QRPs may vary in both prevalence and severity, but the presence of any of these practices calls into question the accuracy of reported findings and undermines the credibility of our field.

**Consequences**

The use of QRPs (Head et al., 2015; Murphy & Aguinis, 2019; Neuroskeptic, 2012; Nuzzo, 2014; Rubin, 2017), the difficulty replicating our work (OSC, 2015), and the murky nature of research reporting (Aguinis, Ramani, & Alabduljader, 2018; Heggestad et al., in press), at least in part driven by pressure to publish in high-prestige journals, have serious impacts on the quality of the research we produce, including the usability but also the believability of our works. Of notable example is the apparent lack of concern regarding measurement in applied psychology over the past few decades (c.f., Cortina, Aguinis, & DeShon, 2017). The lack of emphasis on measurement means that the complex and novel theories evident in the organizational sciences are tested with the weakest foundation possible. A recent review of measurement practices in the organizational sciences found that the majority of our research is tested with measures that have been adapted in some way or another (Heggestad et al., in press). It is important to note that any scale adaptation can change the validity of the new scale, with some adaptations being more severe than others. Of greater concern to us is the timing of when various scale adaptation occurs. Some of the adaptations are
likely conducted *a priori*. Examples of adaptations that almost certainly occur *a priori* include changing scale anchor points and shifting the number of demarcations in a Likert scale. Other scale adaptations are less certain to be *a priori*. The kinds of changes that can (and we suspect do) happen after data has been collected include dropping of items that do not work well (which can negatively affect scale reliability, scale validity, or both) and indicating that the target of items has been changed to assess a group construct (e.g., saying that a scale was modified from an individual level assessment to a team level assessment). These changes are akin to exploratory model fitting, and if these processes are not disclosed to readers, they represent yet another way in which researchers pass off exploratory analyses as confirmatory.

The work of Heggestad and colleagues along with our own work (Schoen, Köhler, DeSimone, & Davison, 2019) also suggests a large proportion of changes to scales go unreported. Examples include (a) researchers who report that a scale contained 5 items when the original scale contained 12 items and (b) statements that scales were adapted with no accompanying information on how the scales were adapted or whether the adapted scale was revalidated. We have a strong suspicion that items are frequently dropped from scales and that such scale modifications are not always disclosed to readers. A recent review of research published in the *Journal of Applied Psychology* demonstrates that the number of items in scales from recent works are not correlated with the reliabilities reported for those scales whereas the expected correlation between scale length and reliability does appear in older works (Cortina, Sheng, List, Katell, & Keeler, 2019). The degree to which a shortened scale taps the original construct as cited by researchers likely ranges from “quite considerably” to “little at all.”

Finally, our own review of the research in high-prestige journals also suggests a large proportion of studies are tested with ad hoc scales supported with minimal validation work.
These scales may include items from multiple existing scales combined together in an attempt to create measures for new constructs. In many cases, the number of items developed and tested for these new construct measures are unknown. All we know is that the new scale meets some minimal reliability and exhibits a few, supposedly hypothesized, correlations with criterion measures. It appears that many researchers (and reviewers too) believe that even this minimal amount of scale validation is sufficient for testing complex models and informing theory. However, nearly all recommendations for scale validation (e.g., Hinkin, 1995; Wright, Quick, Hannah, & Hargrove, 2017) suggest the need for far more rigorous validation efforts.

The ultimate consequence then is not just research that is irreproducible but research that is ultimately unbelievable. HARKing and p-hacking are occurring at not just the level of our constructs but also the level of the scales. Thus, a large part of what is produced by the organizational sciences may be little more than an effort in model fitting. While the results of hypotheses tests may be significant and result in publications, the utility of the result of our HARKed and hacked research made to fit our proposed model and sold as confirmatory research when it clearly is not, is neither accurate nor useful to other researchers or practitioners.

**Conclusion**

As the previously highlighted research has shown, many QRPs are common in top journals, much more common than we can allow as a field if we want to remain credible. If journals and reviewers value novelty and theoretical contribution as opposed to sound theorizing, methodology, empirical quality, and robust results, they breed and publish low quality research. Judging a researcher’s output and performance by the citation count (here termed prestige) of the journal they publish in is not a robust assessment of the quality of their
research. In fact, it may be warranted to read that work extra closely for incorporated shortcuts and QRPs. At the same time, we need academic institutions to acknowledge the system they are creating. Citation counts are not a sign of quality of work published in these journals. By coercing researchers to engage in the highly competitive publishing game for top-journal space, they tacitly encourage bad practices for all the reasons mentioned above.

Moreover, if the ultimate goal is to advance our cumulative knowledge about organizational scholarship, then we suggest that the quality and accuracy of published research is far more important than the journal in which it is published. Once an article is published, it can inform future work and influence the direction of the field. This is certainly a desirable feature of publication when our research is trustworthy, but the same feature can serve as a detriment to our field when incorrect or low-quality work is published. Especially when the topic and findings are perceived to be relevant but the methods with which they were obtained are suspect, we must ask ourselves if we are doing more harm than good. Research across multiple fields has demonstrated that retracted papers can continue to be cited years after the authors or publisher retracts the paper (Bar-Ilan & Halevi, 2017; Fernández & Vadillo, unpublished manuscript). The potential misleading influence of corrected and retracted articles underscores the importance of ensuring the publication of high-quality research. Scholars in our field build on previous research to inform knowledge creation, and as such, it is critical that the review and publication processes are aligned with the goal of maximizing the accuracy of published research, more so than maximizing novelty and theoretical contribution.
References


Author/s:
Köhler, T;DeSimone, J;Schoen, J

Title:
Prestige Does not Equal Quality: Lack of Research Quality in High-Prestige Journals

Date:
2020

Citation:

Persistent Link:
http://hdl.handle.net/11343/241657