Journal of Expertise
2018. Vol. 1(3)
© 2018. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.
ISSN 2573-2773

Statistical and Methodological Considerations for Research on Expertise: Remedies from Open Science and the Reproducibility Crisis

Samuel T. McAbee
Department of Psychology, Bowling Green State University

Correspondence: smcabee@bgsu.edu

Abstract
Researchers who study expertise are faced with a number of statistical and methodological challenges, including small sample sizes, range restriction, and concerns surrounding research design. Recent trends in psychological science point to the need for increased transparency and a greater emphasis on the reproducibility of research findings. In this article, I discuss lessons that expertise researchers can draw from the greater open science community and the reproducibility crisis to combat challenges inherent to the study of expertise.

Keywords
expertise, open science, transparency, replication, reproducibility, methodology

Introduction

Experts are those rare individuals who demonstrate exceptional knowledge and skill in one or more domains of life activity, such as in music, sports, and education (e.g., Macnamara, Hambrick, & Oswald, 2014; Macnamara, Moreau, & Hambrick, 2016). By virtue of their rarity, the study of experts presents a number of significant obstacles to researchers. In a 2017 book chapter, Fred Oswald and I described a number of statistical and methodological challenges that are characteristic of the study of expertise, largely arising from concerns over small sample sizes, restricted variability among participants, and research design decisions (McAbee & Oswald, 2017). Building on the reflections of individual differences researchers in the area of skill acquisition and expertise (e.g., Ackerman, 2014), we articulated four common concerns for research in this area:

1. The small sample sizes typical of studies of expertise (a) limit statistical power and (b) result in large amounts of sampling error variance around obtained effect sizes, collectively reducing confidence in the magnitude of observed effects.
2. Because experts are, by definition, rare individuals who exist at the upper extremes of the performance distribution, observed effects are often attenuated within this population due to range restriction in the variance of performance and its determinants (i.e., direct range restriction), external variables that are not directly of interest but are nonetheless related to expert performance and its determinants (i.e., incidental range restriction), or some combination of these.

3. The use of extreme groups designs (e.g., studies comparing performance between groups of experts and novices), although efficient for detecting the direction and presence of statistically significant effects, can inflate obtained estimates of effect size through a process known as range enhancement, whereby variance on the performance dimension under study is greater among participants than it is within the general population of interest.
4. The designation of expert status is often arbitrary and ad hoc, with few areas of study classifying experts and non-experts on standard, accepted definitions of performance within the domain (e.g., Elo ratings in chess).

Importantly, although the four aforementioned concerns are critical to research on expertise, there are other areas that our chapter did not address that are of equal importance, including concerns over the psychometric qualities (e.g., reliability, construct validity) of the measurements used (Clark & Watson, 1995; Marcondes & Marcondes, 2018), and broader methodological issues that limit the generalizability and replicability of findings within this area. It is the last of these areas of concern that I discuss here. Drawing lessons from the open science community (e.g., McAbee, Grubbs, & Zickar, 2018; Nosek et al., 2015; Open Science Collaboration, 2015) and the reproducibility crisis (e.g., Baker, 2016; Munafò et al., 2017; Tackett et al., 2017), I first briefly review recent trends in replication research and research transparence. I then offer a number of suggestions for ways that researchers can mitigate some of the widespread statistical and methodological challenges inherent to the study of expertise.

The Reproducibility Crisis and Open Science

Beginning in the mid-2000s, a confluence of events began to unearth a number of questions about the robustness and reproducibility of findings from psychological research. In a provocative paper titled *Why Most Published Research Findings are False*, Ioannidis (2005) articulated a number of statistical issues that lead to increased rates of Type I error (i.e., asserting that an effect is present when it is not) among published research articles, including small sample sizes, small effect sizes, increased numbers of statistical tests, and other publication-related concerns (see also, Brysbaert & Stevens, 2018, for a discussion of the effects of sample size and number of experimental conditions on statistical power). Based on the results of a simulation study, Ioannidis argued that for most research designs this false-positive rate might exceed 50% of all published effects.

This startling realization was quickly followed by a number of large public incidents concerning data fabrication, such as the 2011 case of Diederik Stapel (see Stroebe, Postmes, & Spears, 2012); papers reporting outlandish, yet statistically significant results, such as the publication of an article by Bem (2011) asserting evidence supporting extrasensory perception; and papers raising concerns of the rate and role of questionable research practices (QRPs) in major psychological outlets (e.g., John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011; see also Pashler & Wagenmakers, 2012; Sijtsma, 2016; Waldman & Lilienfeld, 2016; Wigboldus & Dotsch, 2016). Collectively, the aforementioned events gave rise to psychology's recent reproducibility crisis, wherein increasing attention has been paid to replicating research findings and issues surrounding transparency in the research process within psychology (e.g., Zwaan, Etz, Lucas, & Donnellan, 2018) and other fields (e.g., ecology: Fidler et al., 2017; the social sciences: Mullinix, Leeper, Druckman, & Freese, 2015).

A number of efforts have resulted from the reproducibility crisis, largely facilitated by the Open Science Framework (OSF: https://osf.io/) out of the Center for Open Science (https://cos.io/). Based on principles of transparency as a fundamental characteristic of scientific engagement (Nosek & Bar-Anan, 2012; Nosek, Spied, & Motyl, 2012), the OSF

was developed as an online repository allowing researchers to collaborate internationally and, centrally, to pre-register research hypotheses, to publish pre-prints of articles, and to share supporting materials for existing or ongoing studies. Alongside the OSF, the Center for Open Science has produced the Transparency and Openness (TOP) Guidelines, which are a set of eight standards that researchers, reviewers, editors, and publishers can follow to increase transparency throughout the research process, from study inception to publication. Importantly, each standard is described at four alternative levels of stringency allowing for a more continuous classification of transparency. These standards include (1) requirements for citation of data sources and other materials, (2) data transparency (i.e., the extent to which data is available publically), (3) analytic methods transparency (i.e., sharing of syntax code), (4) transparency and sharing of research materials, (5) transparency in describing the study design and analyses conducted, (6) preregistration of study hypotheses or research questions with the journal, (7) preregistration of all planned analyses associated with a research study, and (8) explicit encouragement of direct and conceptual replication reports for publication by journals (see Nosek et al., 2015).

Around the same time that the OSF was being developed, researchers at the Center for Open Science (e.g., Open Science Collaboration, 2012, 2015) and elsewhere (e.g., Klein et al., 2014) began to undertake a number of alternative paths for increasing the robustness and replicability of psychological research. This is exemplified by the publication of several special issues (e.g., *Journal of Research in Personality*; Donnellan & Lucas, 2018; *Psychological Science*; Pashler & Wagenmakers, 2012) and current calls for papers (e.g., *Journal of Abnormal Psychology*, *Psychological Bulletin*) highlighting reproducibility, and calling for direct and conceptual replications of classic findings in various areas of psychological inquiry. As one route towards increased replication research, several large-scale "multilab" studies have been published (e.g., Frank et al., 2017; Milcu et al., 2018; Open Science Collaboration, 2012), wherein researchers from multiple universities (often internationally) pool resources to individually and collectively replicate the findings of past research. As one example, Hagger et al. (2016) presented the results of 23 individual studies conducted by various researchers and a subsequent integrative meta-analysis on the ego-depletion effect (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven, Tice, & Baumeister, 1998; Sripada, Kessler, & Jonides, 2014). The findings of this collaborative effort indicated that the magnitude of the ego-depletion effect was quite small with confidence intervals that include zero (d = .04, 95% CI = -.07, .15), thus calling into question the broader resource depletion model of self-control. Notably, only *two* of the 23 replication studies conducted obtained positive effects for which the confidence interval did not include zero (Ns = 50 and 55, respectively), and one study (N = 56) observed a *negative* effect for which the confidence interval did not include zero. This study underscores the variability of observed effects, particularly when obtained from small samples (e.g., those typical of expertise research), and the importance of replication for drawing broader conclusions about the robustness of psychological findings.

Collectively, the trends toward increasing research reducibility and transparency following the reproducibility crisis highlight a number of critical issues relevant to research on expertise. In the next section, I will address the four statistical and methodological challenges outlined in the introduction to this article and argue that a shift towards open science and replication practices can help to alleviate these concerns for research on expertise, particularly as they relate to the typically small and uncommon samples studied, research reproducibility, and ways to increase research transparency.

Transparency, Replication, and Expertise Research

One theme common to both the concerns for research on expertise raised at the beginning of

this article and broader concerns for reproducibility in psychology as a whole is the role of sample size. As noted, studies of experts often rely on small sample sizes. Although some authors have recommended alternative means of reducing the rate of publication of false positive effects (i.e., Type I errors) that often accompany small sample sizes, such as lowering the threshold for statistical significance (Benjamin et al., 2018), these alternatives are not without criticism (e.g., Trafimow et al., 2018); thus, increasing sample sizes typical of research remains the most viable alternative for increasing confidence in statistical conclusions. In addition, by virtue of their rarity and extreme standing on the performance domain of interest, research on expertise is often characterized by range restriction relative to the population as a whole (McAbee & Oswald, 2017; Vaci, Gula, & Bilalić, 2014), which can also lead to reduced confidence in the accuracy of obtained effects.

There are a number of ways in which recent trends following the reproducibility crisis and the shift towards open science might help to alleviate difficulties that expertise researchers face when obtaining access to participants. In particular, expertise researchers might consider forming coalitions across institutions to conduct multilab research studies, similar to that described by Hagger et al. (2016). In such experiments, researchers across several labs conduct identical studies of a phenomenon in parallel, and the results of these studies are reported separately and in combination through techniques such as meta-analysis (e.g., Schmidt & Hunter, 2014) within a single publication. Multilab studies are likely to represent a particular advantage to expertise researchers, where individual researchers (or research groups) are likely to have limited access to particular populations of interest, often resulting from geographic constraints (e.g., musicians training at top-tier conservatories). New technologies out of the open science movement make the design and implementation of such multilab research more accessible than ever before. For example, the Psychological Science Accelerator (PSA) provides an online infrastructure allowing researchers to advertise and recruit collaborators and participants for research studies through crowdsourcing (Moshontz et al., 2018). Expertise researchers could leverage collaboration networks such as those provided by the PSA and OSF to develop communities of researchers studying similar topics within expertise, thereby increasing access to experts of various types and offering the opportunity to produce evidence based on more larger and more representative samples of experts.

A second theme that has emerged from the reproducibility crisis is the greater emphasis on direct replications of classical psychological findings. Direct replications are characterized by the researchers attempting to emulate, as closely as possible, the methods and conditions undertaken by the initial researchers (Makel, Plucker, & Hegarty, 2012). Conversely, conceptual replications purposefully alter the methods and measures used in an initial study in order to test the robustness of the findings across design variations. Research on expertise has long been characterized by its adherence to theory for driving scientific inquiry, and major theories of expertise and skill acquisition (e.g., Ackerman, 1987; Chase & Simon, 1973a, 1973b; Ericsson, Krampe, & Tesch-Römer, 1993; Gobet & Simon, 1996) have received support across numerous individual studies. Indeed, reproducibility is one of the most fundamental prerequisites for ensuring the robustness of scientific theory (e.g., Popper, 1959), and some have argued that the role of direct replications is to "determine if a specific way of testing a theoretical ideal will produce a similar result in a subsequent attempt" (Zwaan et al., 2018, p. 4). Thus, one interpretation of the present reproducibility crisis is that direct replication is *sine qua non* for advancing scientific theory. Notably, however, examples of direct replications in research on expertise are relatively uncommon (e.g., Calin-Jageman, 2018; Gobet & Simon, 1998; Schneider, Gruber, Gold, & Opwis, 1993). Although existing theory can alleviate some concerns surrounding the reproducibility of research on expertise, the frequency of direct replication reports and the role of direct replication for supporting theory

remain important topics for substantive debate.

The need for direct replications of research findings in the study of expertise is also evident given the reliance on small samples in many such studies. As we noted in our chapter (McAbee & Oswald, 2017), effects obtained in smaller samples demonstrate a greater degree of sampling error variance, resulting in wider confidence intervals, and hence, a larger plausible range of potential values for a given effect under study. It is exactly this situation in which direct replications are most helpful: Through direct replication across multiple (even small) samples, researchers can have greater confidence in the aggregate effects obtained—although any given sample is still subject to concerns over the degree of sampling error present. As a result, expertise researchers might consider conducting direct replications of classic studies; particularly those whose effects were demonstrated with limited sample sizes. Outlets such as the *Journal of Expertise* could facilitate the publication of such direct replication reports by creating special sections or special issues of their journals dedicated to publishing high-quality direct replication reports, often of the pre-registered variety described in the TOP guidelines (Nosek et al., 2015).

One final emergent trend from the reproducibility crises and open science community that would be of great benefit to expertise researchers is the increased recognition of research transparency evident in the broader psychological sciences. In particular, expertise researchers would greatly benefit by sharing preprints of manuscripts, materials used in conducting research studies, and data collected and published in online repositories, such as the OSF or PsyArXiv (https://psyarxiv.com/). As noted at the beginning of this article, relatively few domains of expertise hold common, accepted standards for classifying experts from non-experts, and the measures used to characterize expertise and its correlates are often study-specific. By publishing materials along with research findings, or sharing these in online repositories, researchers can aid future studies that seek to replicate and extend existing findings by reducing the time and resources researchers must spend to develop such materials (McAbee et al., 2018; Munafò et al., 2017), potentially increasing the standardization in the ways that expert performance is measured and assessed across studies.

Although researchers might be reticent to publish de-identified data, access to larger archival datasets can increase the robustness with which future studies are able to examine phenomenon in the expertise literature. Importantly, reexamining primary data can lead to unseen insights not tested by the initial researchers, including a greater understanding of how decisions made during data analysis can affect the robustness of statistical conclusions (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). In addition, despite concerns over being "scooped," recent research has suggested that this occurs less often than expected (e.g., LeBel, Campbell, & Loving, 2017), and there is initial evidence that sharing data in a published public archive can lead to potential advantages for the researchers involved, such as increased citation rates (e.g., Piwowar & Vision, 2013). Collectively, the sharing of materials and data can help to increase the transparency, confidence, and efficiency with which research is conducted. As a result, some publications, including the *Journal of Experimental Psychology*, have moved towards encouraging authors to make their research more open to all consumers of research.

Conclusion

Research on expertise is often criticized as a result of typically small samples sizes, restricted variance in the variables under study, and methodological design concerns. These concerns, in part, mirror general difficulties facing psychological science as a whole that have contributed to a lack of confidence in the robustness of research findings. Importantly, however, recent trends as a result of the reproducibility crisis and the shift towards open scientific practices offer a number of mechanisms by which expertise researchers can increase confidence in their results: by

encouraging researchers to be more open and transparent with study materials and data, by encouraging direct replication of scientific research, particularly when conducted in small samples, and by providing avenues through which researchers can share and collaborate throughout the entirety of the research process. It is my sincere hope that expertise researchers, reviewers, editors, and publishers heed this call towards a more open and reproducible science of expertise to ensure the viability and longevity of the field.

Author's Declaration

The author declares that there are no personal or financial conflicts of interest regarding the research in this article.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3-27. doi:10.1037/0033-2909.102.1.3

Ackerman, P. L. (2014). Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence*, *45*, 6-17. doi:10.1016/j.intell.2013.04.009

Baker, M. (2016). Is there a reproducibility crisis? *Nature*, *533*, 452-454.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? Journal of Personality and Social Psychology, 74, 1252-1265. doi:10.1037/0022-3514.74.5.1252

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407-426. doi:10.1037/a0021524

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers E.-J., Berk, R., … & Johnson, V. E. (2018). Redifine statistical significance. *Nature Human Behaviour*, *2*, 6-10. doi:10.1038/s41562-017-0189-z

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effect models: A tutorial. *Journal of Cognition*, *1*, 9. doi:10.5334/joc.10

Calin-Jageman, R. J. (2018). Direct replications of Ottati et al. (2015): The earned dogmatism effect occurs only with some manipulations of expertise. *Journal of Experimental Social Psychology*, *78*, 240-249. doi:10.1016/j.jesp.2017.12.008

Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology*, *4*, 55-81. doi:10.1016/0010-0285(73)90004-2

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319. doi:10.1037/1040-3590.7.3.309

Donnellan, M. B., & Lucas, R. E. (2018). Introduction to the special issue – A replication project in personality psychology. *Journal of Research in Personality*, *72*, 1-4. doi:10.1016/j.jrp.2017.11.004

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363-406. doi:10.1037/0033-295X.100.3.363

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A. Floccia, C., Gervain, J., … & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*, 421-435. doi:10.1111/infa.12182

Fidler, F., Chee, Y. E., Wintle, B. C., Burgman, M. A., McCarthy, M. A., & Gordon, A. (2017). Metaresearch for evaluating reproducibility in ecology and evolution. *BioScience*, *67*, 282-289. doi:10.1093/biosci/biw159

Gobet, F., & Simon, H. A., (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1-40. doi:10.1006/cogp.1996.0011

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*, 225-255. doi:10.1080/741942359

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., . . . Zwienenberg, M. (2016). A multi-lab pre-registered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*, 546-573. doi:10.1177/1745691616652873

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.

doi:10.1177/0956797611430953

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B. Jr., Bahník, S., Bernstein, M. J., … & Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*, 142-152. doi:10.1027/1864-9335/a000178

LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. *Journal of Personality and Social Psychology, 113*, 230-243. doi:10.1037/pspi0000049

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, *25*, 1608-1618. doi:10.1177/0956797614535810

Macnamara, B. N., Moreau, D., & Hambrick, D. Z. (2016). The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspectives on Psychological Science*, *11*, 333-350. doi:10.1177/1745691616635591

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537-542. doi:10.1177/1745691612460688

Marcondes, D., & Marcondes, N. R. (2018). A nonparametric statistical approach to content analysis of items. *Stats*, *1*(1), 1-13. doi:10.3390/stats1010001

McAbee, S. T., Grubbs, J. B., & Zickar, M. J. (2018). Open science is robust science. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 11*, 54-61. doi:10.1017/iop.2017.85

McAbee, S. T., & Oswald, F. L. (2017). Primer: Statistical methods in the study of expertise. In D. Z. Hambrick, G. Campitelli, & B. Macnamara (Eds.), *The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 13-30). New York: Routledge.

Milcu, A., Puga-Freitas, R., Ellison, A. M., Blouin, M., Scheu, S., Fraschet, G. T., … Roy, J. (2018). Genotypic variability enhances the reproducibility of an ecological study. *Nature Ecology & Evolution*, *2*, 279-284. doi:10.1038/s41559-017-0434-x

Moshontz, H., Campbell, L. Ebersole, C. R., Ijzerman, H., Urry, H. L., Forscher, P. S., … Chartier, C. R., (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*. doi:10.1177/2515245918797607

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*, 109-138. doi:10.1017/XPS.2015.19

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., . . . Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. doi:10.1038/s41562-016-0021

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*, 774-789. doi:10.1037/0022-3514.74.3.774

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S. J., . . . Contestabile, M. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425. doi:10.1126/science.aab2374

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*, 217-243. doi:10.1080/1047840X.2012.692215

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. doi:10.1177/1745691612459058

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657-660. doi:10.1177/1745691612462588

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528-530. doi:10.1177/1745691612465253

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. doi:10.7717/peerj.175

Popper, K. R. (1959). *The logic of scientific discovery*. London, England: Hutchinson

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Schneider, W., Gruber, H., Gold, A., & Opwis, K.

(1993). Chess expertise and memory for chess positions in children and adults. *Journal of Experimental Child Psychology*, *56*, 328-349. doi:10.1006/jecp.1993.1038

Sijtsma, K. (2016). Playing with data—Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, *81*, 1-15. doi:10.1007/s11336-015-9446-0

Simmons, J. O., 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. doi:10.1177/0956797611417632

Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science*, *25*, 1227-1234. doi:10.1177/0956797614526415

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702-712. doi:10.1177/1745691616658637

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*, 670-688. doi:10.1177/1745691612460687

Tackett, J. L., Lilienfeld, S.O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J.D., . . . Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, *12*, 742-756. doi:10.1177/1745691617690042

Trafimow, D., Amrhein, V., Areshenkoff, C. N., Barrera-Causil, C. J., Beh, E. J., Bligiç, Y. K., … & Marmolejo-Ramos, F. (2018). Manipulating the alpha level cannot cure significance testing. *Frontiers in Psychology*, *9*, 699. doi:10.3389/fpsyg.2018.00699

Vaci, N., Gula, B., & Bilalić, M. (2014). Restricting range restricts conclusions. *Frontiers in Psychology*, *5*(569). doi:10.3389/fpsyg.2014.00569

Waldman, I. D., & Lilienfeld, S. O. (2016). Thinking about data, research methods, and statistical analyses: Commentary on Sijtsma's (2014) "playing with data". *Psychometrica*, *81*, 16-26. doi:10.1007/s11336-015-9447-z

Wigboldus, D. H. J., & Dotsch, R. (2016). Encourage playing with data and discourage questionable reporting practices. *Psychometrika*, *81*, 27-32. doi:10.1007/s11336-015-9445-1

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. doi:10.1017/S0140525X17001972

Received: 30 October 2018
Revision received: 29 November 2018
Accepted: 29 November 2018