8-2017

More Multi-Study Articles Wanted

Michael Lynn
Cornell University, wml3@cornell.edu

Recommended Citation
Lynn, W. M. (2017). More multi-study articles wanted. *Cornell Hospitality Quarterly, 58*(3), 228. doi: 10.1177/1938965517714372

More Multi-Study Articles Wanted

Abstract
[Excerpt] Cornell Hospitality Quarterly (CQ) readers may have noticed that the lead article for this issue and for each of the previous two issues has been a multi-study paper. The lead article for the next issue of CQ will also be a multi-study paper, and this will be true for future issues as long as I have enough accepted multi-study papers to make it so. I want to use this editorial to explain my preference for multi-study articles and to encourage CQ authors to write and submit more of them.

Keywords
Cornell, editorial, publishing, empirical articles, writing

Disciplines
Hospitality Administration and Management | Scholarly Publishing

More Multi-Study Articles Wanted

Cornell Hospitality Quarterly
2017, Vol. 58(3) 228

DOI: 10.1177/1938965517714372
journals.sagepub.com/home/cqx
SAGE

Cornell Hospitality Quarterly (*CQ*) readers may have noticed that the lead article for this issue and for each of the previous two issues has been a multi-study paper. The lead article for the next issue of *CQ* will also be a multi-study paper, and this will be true for future issues as long as I have enough accepted multi-study papers to make it so. I want to use this editorial to explain my preference for multi-study articles and to encourage *CQ* authors to write and submit more of them.

Multi-study articles contribute more than typical, single-study articles. One reason is that the use of different measures or manipulations, research settings (field vs. lab), service contexts, and moderators across studies helps to establish the generalizability of effects, which speaks to their practical value and often to the processes underlying them. However, multi-study articles contribute more than single-study articles in another way as well: The literal or conceptual replication of main or simple-main effects across studies helps to increase our confidence that those effects are reliable and not due to chance. This latter contribution turns out to be particularly valuable as explained below.

Science is facing a much-discussed replication crisis. Across diverse fields, such as economics, psychology, and medicine, researchers have discovered that many effects in their literatures are less reliable or replicable than commonly believed. This lower-than-expected reliability of findings is partly attributable to misbehavior on the part of researchers including the use of questionable statistical practices that capitalize on chance (called *p*-hacking) and outright fraud. However, there is another, more fundamental, and widespread cause of the lower-than-expected reliability of published research findings. Specifically, it is also a consequence of two facts: (a) conducting more tests increases the probability of making a Type 1 error, and (b) science has a bias toward publishing "significant" results.

We typically think of the *p* value associated with a test-statistic as indicating the maximum probability of making a Type 1 error (of being wrong IF we conclude the tested effect is reliable). By requiring a *p* value to fall below some predetermined level (the alpha level) before we conclude that the associated effect is reliable, we can control our maximum probability of making Type 1 errors. However, if we conduct multiple tests, then the maximum probability of making at least one Type 1 error across all those tests exceeds our chosen alpha level. The rate of Type 1 errors across all published and unpublished studies and statistical tests equals the average of the alpha levels used in those tests, but science's publication bias favoring "significant" results means that the rate of Type 1 errors in the published literature exceeds the average of the alpha levels used. In other words, our bias against publishing null results causes published effects to be less reliable than their reported statistical tests suggest.

It is in this context that multi-study replications are particularly valuable as a way to help to assess the reliability of effects. Of course, replications can be Type 1 errors too, and the bias against null results extends to replications, so the reliability of even replicated effects is lower than suggested by their reported statistical tests. However, researchers studying effects that occur at only chance levels are likely to rapidly move on to other more promising and productive areas of inquiry, so multiple replications of an effect in a single paper or by a single scholar are reasonably good indications that the effect is reliable. Therefore, I encourage *CQ* authors to conduct and report on multi-study investigations of the effects they study.

Single study papers are still welcome and are likely to outnumber multi-study papers in *CQ* for the foreseeable future, but multi-study papers are more informative and will be preferred. Please note that not every study in a multi-study paper has to produce a significant effect to be accepted at *CQ*. Even high-powered studies can produce occasional nonsignificant results for real effects (i.e., Type 2 errors). It is better to include such null-result studies in multi-study papers rather than leave them out because they still provide valuable information about the size and reliability of effects. To maximize learning from a set of studies, researchers should include all the studies conducted and (especially if the significance of results varies across studies) report meta-analyses across the studies in their paper in addition to—or sometimes instead of—separate analyses of each study. See McShane and Bockenholt (2017) for a tutorial on one way to do this.

Michael Lynn

Reference

McShane, B. B., & Bockenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, *43*, 1048-1063.