

Despite Stronger Vetting and Sampling, Certain Psychological Research Results Elude Replication

January 06, 2021



The “gold standard” for all scientific research—from physics and biology to psychology and medicine—is the ability to replicate experimental results. If studies do not hold up under new investigation, then the discrepancies must either be accounted for or the original conclusions may be called into question.

In a surprising and, to many, disquieting discovery, the 2015 “[Reproducibility Project: Psychology](#)” found that out of 100 published psychology studies, only 40% could be successfully reproduced. These results sparked debate about the credibility of psychological research and prompted global interest in finding the reasons behind the lack of reproducibility.

“If the original findings are replicable, then the conditions necessary to observe them are not yet understood.”

Christopher Chartier (Ashland University)

Some [researchers proposed](#) that this lack of reproducible results was possibly a consequence of inadequate sample size and the replicators’ not adhering to experts’ insight when designing the replication studies.

A new [collection of 11 articles](#) published in the Association for Psychological Science's journal *Advances in Methods and Practices in Psychological Science* (AMPPS), however, found that a dramatic increase in sample size and prior expert peer review of replication designs did not increase replicability of the original findings.

"If the original findings are replicable, then the conditions necessary to observe them are not yet understood," said Christopher Chartier, associate professor of psychology at Ashland University, Ohio, and co-author on the study.

The new replication project, a multiteam effort known as Many Labs 5, examined a specific subset of studies from the original replication attempts. The Many Labs 5 project selected 10 out of 11 findings that were "not endorsed" by the original authors. These were studies in which the original authors had expressed reservations about the replication methodology that the original replication team did not completely address.

The replication teams submitted proposed study protocols for formal peer review and revised the protocols accordingly before conducting their studies. These efforts allowed for a direct comparison of whether the expert feedback improved replicability of the original findings.

"We tested whether revising the replication protocols based on expert reviews could improve replicability of the findings, and we found that it had no meaningful impact on these findings," said Charlie Ebersole, lead author of the project and postdoctoral associate at the University of Virginia. "Overall, the effects generated by the original replications were very similar to those generated by our revised protocols. Looking at all of these replications, our evidence suggests that the original studies may have exaggerated the existence or size of the findings."

"These results do not suggest that expertise is irrelevant. It could be that this particular selection of studies was unlikely to improve no matter what expert feedback was provided," said Hans IJzerman, co-author and associate professor at Université Grenoble Alpes, France. "It will be interesting to conduct follow-up research on findings that are known to be replicable but have complex methodologies to help assess the role of expertise in achieving replicable results."

"There were hints that some of the findings may be replicable, and perhaps even slightly more so with the revised protocols for one or two of them," said Hugh Rabagliati, co-author and reader in psychology at Edinburgh University, U.K. "However, because we had very large samples, our findings had much more precision than the original studies."

The findings are evidence against the hypothesis that the earlier attempts to replicate these 10 studies were hindered by deficiencies in sample size or adherence to expert feedback.

Future research may yet identify conditions that improve replicability of these findings. "For now, the cumulative evidence suggests that the effects are weaker than the original research suggested or not yet established as a reliable finding," concluded Erica Baranski, co-author and Postdoctoral Researcher at the University of Houston.

"If psychology's reform continues to improve the transparency and rigor of research, I expect that future

replication efforts will demonstrate the tangible impact of those improvements on research credibility,” concluded Brian Nosek, senior author and executive director of the [*Center for Open Science*](#), Charlottesville, Virginia.