



Tests of generalizability can diversify psychology and improve theories

Gerald J. Haefel  and Willie R. Cobb

Diversifying psychology is fundamental to good science. Tests of generalizability using diverse samples can inform and constrain theories, leading to scientific progress and cumulative knowledge.

Psychological scientists and the samples they use tend to be white and Western¹. In the top psychology journals, only 11% of the world's countries are represented with most authors and samples (60%) from the USA². Ironically, the people who might benefit most from psychological science are the least likely to be considered in our theories and studies. For example, 3–6% of all published research articles on mental health include participants from low- and middle-income countries, despite these countries having some of the highest rates of mental illness³. In sum, psychology is a field constrained by the perspectives and methods of a homogenous few⁴.

A second problem with psychological science is poor theorizing. Much of the research in psychology is not theory-driven, and when theories are used they tend to be vague and unfalsifiable⁵. For example, only 15% of articles in *Psychological Science* (between 2009 and 2019) tested predictions derived from theories⁶. This is a problem because testing and refuting theories is the basis for scientific progress⁷. The iterative process of proposing theories, testing hypotheses and revising theories in the face of contradictory evidence is the foundation for generating new knowledge; without it, psychology has little chance of becoming a science with a cumulative character.

Tests of generalizability using diverse samples can help to solve both these problems; they can diversify psychology with more representative participants (that is, samples that are not primarily white and American or European) and facilitate theory building. Tests of generalizability can be conducted using a variety of research designs including both direct and conceptual replications. What is most important is that the study design used to test generalizability allows the researcher to be wrong (that is, not exploratory work).

Studies using more diverse samples are going to expose problems with theories. It is unreasonable to assume that theories derived from white and Western cultures and samples are going to generalize to individuals from all cultures. Results are not going to replicate. But instead of viewing these 'disagreements' as failures, they should be considered opportunities for scientific

progress. Failures to replicate results force researchers to specify the conditions for when their theory will and will not apply, and paves the path for stronger, more-inclusive theories.

It is important to underscore that if a finding does not replicate in a more diverse sample, it does not mean that the original study was invalid. It simply constrains the conclusions that can be made; namely, that the results only apply to a specific population. At the same time, continued overreliance on non-representative samples stalls progress. It is akin to only testing the properties of water at room temperature. The results are valid, but they are a poor representation of nature. It is not a concession or nicety to test the properties of water at different temperatures; it is good science. The same is true for testing psychological theories using more diverse samples.

Examples

To illustrate how tests of generalizability support theory building and have real-world implications, we provide two examples. The first example relates to cognitive theories of depression⁸. According to these theories, some people are at heightened risk for depression because they generate overly negative interpretations of stress. Studies conducted over the past thirty years using US undergraduates have provided strong support for this theory. However, the generalizability of these findings remains unclear. To test for generalizability, the cognitive vulnerability to depression hypothesis was tested in five different populations — Honduran adolescents, Nepali adults, American and European adults, Black US adults and US undergraduates (G.J.H, unpublished work). Results showed that cognitive vulnerability could be measured reliably in all samples, and the distribution of vulnerability scores was similar for people around the globe. However, the association typically found between cognitive vulnerability and depressive symptoms did not generalize to the Honduran and Nepali participants. In other words, the universal tendency to generate negative inferences about stress had different implications for depression, depending on culture. This means that

Department of Psychology,
University of Notre Dame,
Notre Dame, IN, USA.

✉e-mail: ghaefel@nd.edu

<https://doi.org/10.1038/s44159-022-00039-x>

researchers must now try to understand why negative cognitions confer risk for depression in some contexts but not others. It also means that reducing the global burden of depression may require more than ‘transporting’ existing cognitive interventions to other countries.

A second example from a different, but equally important, domain examined children’s learning of mathematics concepts through shared reading of tactile and non-tactile counting books (189 US pre-schoolers, 41% Black; W.R.C., unpublished work). Children worked individually with a tutor in six structured shared book reading sessions in which they counted and labelled the set size on each page. Results showed that race moderated the effect of tactility on early numeracy. Non-Black children showed the greatest growth in numeracy in the non-tactile counting book condition. However, it was the opposite for Black children: they experienced the greatest growth in numeracy in the tactile condition (nearly a one standard deviation improvement in early numeracy). This growth was significantly larger than the growth of non-Black children in the same condition. This effect was specific to Black children (it did not hold for other traditionally underrepresented groups) and could not be attributed to other demographic variables (such as income). These findings demonstrate that basic learning processes are influenced by race and culture, and that failing to include more diverse participants can lead to incomplete understanding about how children learn. This has real-world implications: applying learning strategies derived from one homogenous group of children may not work (or may even backfire) for other children.

Obstacles

There are several obstacles preventing psychological researchers from using more diverse samples. First, there are practical considerations. It is much easier to recruit convenience samples than representative samples. Conducting research with more diverse participants might require travel, new methods of data collection (for example, many people in rural areas lack access to the internet), knowledge of local customs, translators and trusted community partnerships. Solving these practical issues will require universities and funding agencies to value and support this kind of work.

However, the greatest obstacle to diversifying psychology might be systemic bias and tradition. Research in psychology is formulaic — recruit a large convenience sample, find a small *P* value (that is, a ‘positive’ result), create a post hoc explanation for the finding and assume the results generalize. But studies using diverse samples do not fit this formula. Indeed, studies of generalizability are ‘riskier’ than traditional studies because there is a greater chance of contradictory or negative results. This is problematic in a field obsessed

with positive findings⁹. Thus, researchers may be hesitant to conduct tests of generalizability because the results might be difficult to publish (especially in the journals that universities and tenure committees care about⁴). Herein lies the irony — the studies with the greatest potential for creating new knowledge are the ones avoided by scientists.

Solutions

The time is ripe for change; there is a spotlight on systemic racism and a growing choir of voices calling for greater diversity in science. To this end, some journals now require details about sample composition and statements of generalizability¹⁰. However, we suspect these changes will not be enough to shift the dominant practice of publishing only positive results. Further, these new journal requirements might preserve the status quo because researchers can continue to conduct business-as-usual as long as they acknowledge what they are doing. By contrast, we recommend that researchers discuss the theoretical implications of their sample choice. Instead of simply noting the composition of their sample and the constraints on generalizability, researchers should create a ‘theoretical generalizability’ section in which they discuss the relevant literature on culture and race, and then generate a priori hypotheses about how culture and race would affect future tests of the theory. This would promote research on race and culture, generate falsifiable hypotheses and lead to more nuanced and inclusive theorizing.

For real and lasting change to occur, the field must embrace being wrong. Non-replications and theoretical refutations are opportunities to learn. They are not failures, but rather they fuel scientific progress.

1. Clancy, K. B. H. & Davis, J. L. Soylent is people, and WEIRD is white: biological anthropology, whiteness, and the limits of the WEIRD. *Annu. Rev. Anthropol.* **48**, 169–186 (2019).
2. Thalmayer, A. G., Toscanelli, C. & Arnett, J. J. The neglected 95% revisited: is American psychology becoming less American? *Am. Psychol.* **76**, 116–129 (2021).
3. Patel, V. & Sumathipala, A. International representation in psychiatric journals: a survey of 6 leading journals. *Br. J. Psychiatry* **178**, 406–409 (2001).
4. Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D. & Mortenson, E. Racial inequality in psychological research: trends of the past and recommendations for the future. *Perspect. Psychol. Sci.* **15**, 1295–1309 (2020).
5. Meehl, P. E. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J. Consult. Clin. Psychol.* **46**, 806–834 (1978).
6. McPhetres, J. et al. A decade of theory as reflected in *Psychological Science* (2009–2019). *PLoS ONE* **16**, e0247986 (2021).
7. Popper, K. R. *The Logic of Scientific Discovery* (Hutchinson, 1959).
8. Abramson, L. Y., Metalsky, G. I. & Alloy, L. B. Hopelessness depression: a theory-based subtype of depression. *Psychologic. Rev.* **96**, 358–372 (1989).
9. Scheel, A. M., Schijven, M. & Lakens, D. An excess of positive results: comparing the standard psychology literature with Registered Reports. *Adv. Methods Pract. Psychol. Sci.* <https://doi.org/10.1177/25152459211007467> (2021).
10. Simons, D. J., Shoda, Y. & Lindsay, D. S. Constraints on generality (COG): a proposed addition to all empirical papers. *Perspect. Psychol. Sci.* **12**, 1123–1128 (2017).

Competing interests

The authors declare no competing interests.