

Handbook on Conducting Reproduction and Replication Studies

Lukas Röseler	Lukas Wallrich	Helena Hartmann
Flavio Azevedo	Joachim Hüffmeier	Janik Goltermann
Charlotte R. Pennington	Veronica Boyce	Sarahanne M. Field
Merle-Marie Pittelkow	Don van Ravenzwaaij	Priya Silverstein

2025-08-01

Table of contents

1 Abstract	3
I Foundations	4
2 Background	5
II The Replication Process	8
5 Planning and Conducting Reproductions and Replications	10
6 Planning and Conducting Reproductions and Replications	11
6.1 Post Publication Conversations	12
6.2 Reproduction before Replication	13
6.3 Close replication before conceptual replication	14
III Advanced Topics and Conclusion	19
References	23
Appendices	25

1 Abstract

The practice of repeatedly testing published results with the same data (reproduction) or new data (replication) is currently gaining traction in the social sciences, owing to multiple failures to reproduce and replicate published findings. Along with increased skepticism have come guidelines for the repeated testing of hypotheses from various disciplines and fields. This guide aims to enable researchers to conduct high-quality reproductions and replications across social science disciplines. First we summarize recent developments, then provide a comprehensive guide to carrying out reproductions and replications, and finally present an example for how guidance needs to be tailored for specific fields. Our guide covers the entire research process: choosing a target study, deciding between different types of reproductions and replications, planning and running the new study, analyzing the results, discussing outcomes in the light of potential differences, and publishing a report.

Keywords: replication, repetitive research, reproducibility, meta-science, meta-research, open science, open research, open scholarship

Part I

Foundations

2 Background

“The proof established by the test must have a specific form, namely, repeatability. The issue of the experiment must be a statement of the hypothesis, the conditions of test, and the results, in such form that another experimenter, from the description alone, may be able to repeat the experiment. Nothing is accepted as proof, in psychology or in any other science, which does not conform to this requirement.” – (Dunlap 1926)

Repeatability is the cornerstone of many sciences: A majority of the scientific progress rests on the successful accumulation of evidence for claims through reproduction and replications to establish robust discoveries. Reproductions and replications, that is repeated testing of a hypothesis with the same (reproduction) or different (replication) data, are necessary.

Cumulative science without repetition is costly. The aim of this guide is to empower researchers to conduct high-quality reproductions and replications and thereby contribute to making their fields of research more cumulative and robust. Issues of replicability have been discussed across many disciplines, such as psychology ((Open Science Collaboration 2015)), economics ((Dreber and Johannesson 2024)), biology ((Errington et al. 2021)), marketing ((Urminsky and Dietvorst 2024)), linguistics ((McManus 2024)), computer science ((Hummel and Manner 2024)) and epidemiology ((Lash, Collin, and Van Dyke 2018)) and the number of replications has been rising sharply (see Figure 1).

Figure 1

Number of replication studies by year of publication based on the FORRT Replication Database (FReD, (Röseler et al. 2024)) based on the version from July 16, 2025. Code to reproduce the figure: <https://osf.io/dznrb>.

![[image1](img/bVY_Image_1.jpeg)]

While the number of replication and reproduction studies has increased, the overall proportion of them is still very small, with reviews finding yearly replication rates of up to 1% ((Perry, Morris, and Lea 2022)). Moreover, much of the guidance on replications is being developed actively ((Clarke et al. 2024)) and in narrow parts of science, which leads to fragmentation, siloing, and potentially inconsistent information.

Here we attempt to integrate useful guidelines (e.g., (Block and Kuckertz 2018; Jekel et al. 2020)) into a comprehensive overview that allows diverse fields to profit from each other. In sum, this guide provides information about the entire process of research allowing researchers at all career stages to plan, conduct, and publish reproduction and replication studies. We limit

our scope to quantitative research, given that the concept of reproducibility and replicability itself is highly contested among qualitative researchers (see Makel, Plucker, and Hegarty (2012); Cole et al. (2024); Pownall (2022); Bennett (2021)).

3

Part II

The Replication Process

4

5 Planning and Conducting Reproductions and Replications

6 Planning and Conducting Reproductions and Replications

Planning depends on whether the focus is on a certain method or a theory, that is whether the replication will be close or conceptual. Table 6.1 provides an overview of reproduction and replication types, or more generally “repetitive research” ((Schöch 2023)), drawn from different resources (e.g., (Dreber and Johannesson 2024; Hüffmeier, Mazei, and Schultze 2016); for an alternative taxonomy see also (Cortina, Köhler, and Aulisi 2023)). The decision between these types is the first step in planning.

In addition, the formation of the replication team is important, as replications can take substantial resources. Notably, repetitive research has successfully been conducted collaboratively with graduate and undergraduate students (e.g., (Boyce et al. 2024; Hawkins et al. 2018; Jekel et al. 2020; Moreau and Wiebels 2023)) and we recommend the use of replication studies to engage students of different levels in conducting and publishing research.

Table 6.1: Types of repetitive research ordered by reproduction and replication and respective closeness to the original study.

Type	Description	Goals
Computational Reproduction	Reanalysis of the same data with the same code	Correctness of original report
Recoding reproduction	Reanalysis of the same data, with new (equivalent) code	Correctness of original report
Robustness Reproduction	Reanalysis of the same data with new coding choices; can vary in closeness	Robustness of original finding and sensitivity to different analytical decisions or software
Multiverse analysis	Analyze data in all sensible ways (i.e., a large number of different robustness reproductions)	Robustness and generalizability of original finding, identification of potential moderators or sources for effect variability

Type	Description	Goals
Internal replication	Replicate one of your own studies as closely as possible	Demonstrate one's findings' generalizability across studies and rule out fear of false-positives (e.g., for new discoveries)
Close / direct / exact replication	Conduct a new study (based on work by other researchers) that is as close as possible to the original study	Rule out fear of the original finding being a false-positive, validate original materials or design, check generalizability/external validity for theoretically irrelevant variables (e.g., population, year of data collection)
Close replication with extension	Add a variable or procedure to a close replication	Rule out fear of the original finding being a false-positive, test generalizability of original finding
Conceptual / constructive replication	Conduct a study with changes that may be <i>theoretically</i> relevant but that tests the same hypothesis (e.g., different operationalization)	Generalizability of original finding, validity of theory
Multilab / multisite replications	Conduct studies that adhere to a predetermined replication protocol in many different locations at the same time	Robustness, generalizability over different locations, and receive a precise estimate of effect size and heterogeneity

6.1 Post Publication Conversations

When planning the replication study, additional knowledge should be taken into account such as any discussions of the original finding. There can be other studies citing the original studies, criticizing them, disconfirming their underlying theory, identifying errors, reinterpreting the finding, or making suggestions for replications. All of these might highlight considerations that need to be taken into account when designing a replication study that *robustly* tests the original claim or its generalisability.

Thus, replication researchers should look for post-publication discussions on the target study such as published comments and reviews, blog posts, or discussions on social media. These can often be found via Altmetric (<https://www.altmetric.com>) or other tools that allow researchers to quickly identify discussions on social media or news outlets beyond scientific

journals (PubPeer [<https://pubpeer.com>], Hypothes.is [<https://web.hypothes.is>]), or the in-development platform Alphaxiv.org [<https://www.alphaxiv.org/>]; for a review see (Henriques et al. 2023)).

6.2 Reproduction before Replication

Many features of a replication study rest on the correctness of the original report. A reproduction allows researchers to investigate this by being able to uncover coding errors, fraud, robustness to analytical decisions, and generalizability. To make efficient use of resources, we encourage researchers to investigate the original finding’s reproducibility and robustness first. In other words, ideally, reproductions should take place before planning and conducting a replication study. Depending on the availability of the code and data, these can take several minutes to weeks.

If the original code and dataset are available, researchers can try to *numerically reproduce* the results. Beware, however, that differences in software versions or default settings may lead to slight deviations or require corrections in some cases (for a large-scale test of reproducibility see Brodeur et al., 2024). Similarly, the lack of a set seed for random number generators can mean that analyses relying on random numbers (e.g., bootstrapping) cannot be exactly reproduced. If no analysis script is available, analyses need to be recreated from the descriptions in the report (*recoding reproduction*). In this case, special attention should be paid to processing steps such as exclusion of outliers, transformation of variables, and handling of missing data. However, in many research areas information on these steps is often incomplete (Field et al., 2019); older research tends to be especially limited in terms of the methodological details they provide. In addition, we recommend testing the robustness of the original finding by making small alterations to the data processing and analyses procedure (*robustness reproductions*). For example, if the analyses were run for a subset of the data (e.g., participants aged 21 to 30 or without outliers ± 3 standard deviations), this subset can be changed (e.g., participants aged 18 to 30 or without outliers ± 2 standard deviations). Here, the initial focus should be on choices that are not determined by the *theory* that is presented, though this can also be used to explore the generalisability of some aspects of theory. Finally, if the original study was preregistered and the original code is available, reproduction researchers can check whether the original analyses adhere to the preregistered analysis plan.

If neither code nor data are available (or shared by the authors), no reproduction is possible. Researchers can still use automated tools to compare reported *p*-values with those that can be computed from test statistics via the website statcheck.io (where documents may be uploaded) or the corresponding R package ((Nuijten and Polanin 2020)).

6.3 Close replication before conceptual replication

If the goal is to increase the generalizability of a specific finding, we also suggest starting with replications that adhere as close as possible to the original study (e.g., close replications) and only later conduct conceptual replications. Based on (Hüffmeier, Mazei, and Schultze 2016), we propose the typology and order of replication attempts in Figure 6.1. Note that there may be cases where the sequence of replications is not necessary, or where the context of the replication team requires a focus on generalisability to a specific context (see ?@sec-interpretation).

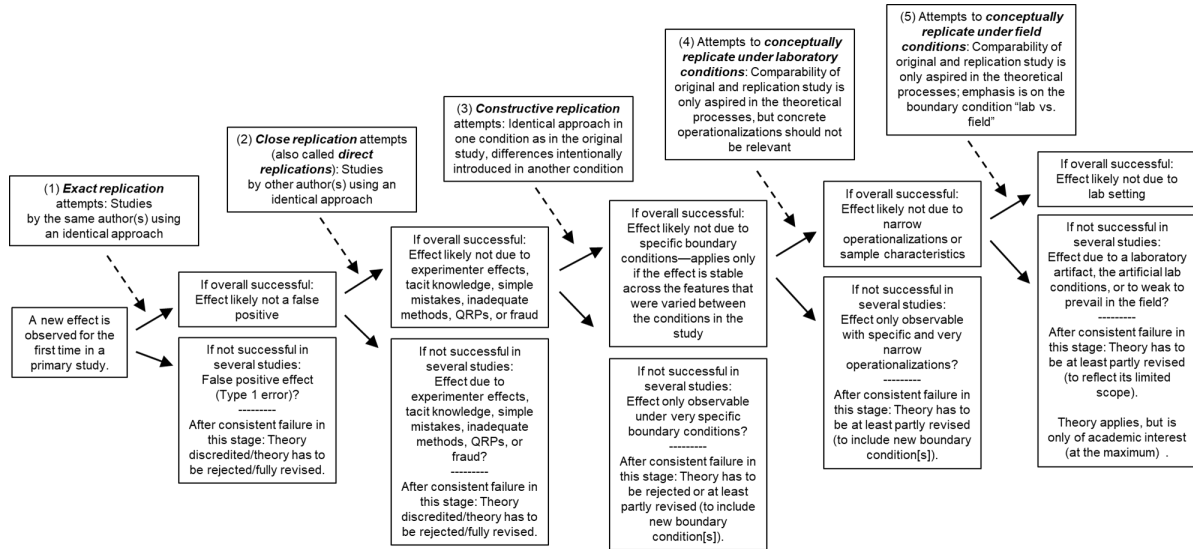


Figure 6.1: Sequence of replications from exact replications to conceptual replications under field conditions. Note: This an adaptation and update of the typology of replication studies by (Hüffmeier, Mazei, and Schultze 2016). The typology is conceptualized as a hierarchy of studies that together help to (i) establish the validity and replicability of new effects, (ii) exclude alternative explanations, (iii) test relevant boundary conditions, and (iv) test generalizability. Importantly, replications at any stage should not compromise any aspects of an original study, but rather (at the latest from the third study stage [constructive replications] onwards) try to improve one or more aspects of the original study, such as “[...] more valid measures, more critical control variables, a more realistic task, a more representative sample, or a design that allows for stronger conclusions regarding causality”, see (Köhler and Cortina 2021, 494)). (Köhler and Cortina 2021) term such replications “constructive replications” and caution against the conduct of “quasi-random” replications that vary features without clear rationale.

At the start, you will need to collect all available materials and data, and contact the original authors if there is something missing. It will often be necessary to contact the authors more than once because missing descriptions of details of the original study only become apparent

once the replication study is planned. In most cases, the original paper identifies one of the authors as “corresponding author” with an e-mail address. We recommend a quick web search to check if this is the current email address, as researchers frequently change institutions and thus e-mail addresses. Sometimes, it may be most helpful to write to the last authors instead, who tend to have more stable e-mail addresses, or to copy all authors into the email. Templates for asking for materials and sharing replication results are in the appendix. Note that original authors may not respond due to institutional changes or not being active in academia anymore.

7

8

9

Part III

Advanced Topics and Conclusion

10

11

12

References

- Bennett, E. A. 2021. "Open Science from a Qualitative, Feminist Perspective: Epistemological Dogmas and a Call for Critical Examination." *Psychology of Women Quarterly* 45 (4): 448–56. <https://doi.org/10.1177/03616843211036460>.
- Block, J., and A. Kuckertz. 2018. "Seven Principles of Effective Replication Studies: Strengthening the Evidence Base of Management Research." *Management Review Quarterly* 68 (4): 355–59. <https://doi.org/10.1007/s11301-018-0149-3>.
- Boyce, V., B. Prystawski, A. B. Abutto, E. M. Chen, Z. Chen, H. Chiu, and M. C. and Frank. 2024. "Estimating the Replicability of Psychology Experiments After an Initial Failure to Replicate," May. <https://doi.org/10.31234/osf.io/an3yb>.
- Clarke, B., P. Y. (K.) Lee, S. R. Schiavone, M. Rhemtulla, and S. Vazire. 2024. "The Prevalence of Direct Replication Articles in Top-Ranking Psychology Journals." *American Psychologist*. <https://doi.org/10.1037/amp0001385>.
- Cole, N. L., S. Ulpts, A. Bochynska, E. Kormann, M. Good, B. Leitner, and T. Ross-Hellauer. 2024. "Reproducibility and Replicability of Qualitative Research: An Integrative Review of Concepts, Barriers and Enablers." https://doi.org/10.31222/osf.io/n5zkw_v1.
- Cortina, J. M., T. Köhler, and L. C. Aulisi. 2023. "Current Reproducibility Practices in Management: What They Are Versus What They Could Be." *Journal of Management Scientific Reports* 1 (3-4): 171–205. <https://doi.org/10.1177/27550311231202696>.
- Dreber, A., and M. Johannesson. 2024. "A Framework for Evaluating Reproducibility and Replicability in Economics." *Economic Inquiry*. <https://doi.org/10.1111/ecin.13244>.
- Dunlap, K. 1926. "The Experimental Methods of Psychology." In *Psychologies of 1925*, edited by C. Murchison, 331–51. Clark University Press. <https://doi.org/10.1037/11020-022>.
- Errington, T. M., M. Mathur, C. K. Soderberg, A. Denis, N. Perfito, E. Iorns, and B. A. Nosek. 2021. "Investigating the Replicability of Preclinical Cancer Biology." *eLife* 10: e71601. <https://doi.org/10.7554/eLife.71601>.
- Hawkins, R. X., E. N. Smith, C. Au, J. M. Arias, R. Catapano, E. Hermann, and M. C. and Frank. 2018. "Improving the Replicability of Psychological Science Through Pedagogy." *Advances in Methods and Practices in Psychological Science* 1 (1): 7–18. <https://doi.org/10.1177/2515245917740427>.
- Henriques, S. O., N. Rzaeva, S. Pinfield, and L. Waltman. 2023. "Preprint Review Services: Disrupting the Scholarly Communication Landscape?" <https://doi.org/10.31235/osf.io/8c6xm>.
- Hüffmeier, J., J. Mazei, and T. Schultze. 2016. "Reconceptualizing Replication as a Sequence of Different Studies: A Replication Typology." *Journal of Experimental Social Psychology* 66: 81–92. <https://doi.org/10.1016/j.jesp.2015.09.009>.

- Hummel, T., and J. Manner. 2024. “A Literature Review on Reproducibility Studies in Computer Science.” In *Proceedings of the 16th ZEUS Workshop on Services and Their Composition (ZEUS 2024)(CEUR)*. Vol. 3673.
- Jekel, M., S. Fiedler, R. Allstadt Torras, D. Mischkowski, A. R. Dorrough, and A. Glöckner. 2020. “How to Teach Open Science Principles in the Undergraduate Curriculum—the Hagen Cumulative Science Project.” *Psychology Learning & Teaching* 19 (1): 91–106. <https://doi.org/10.1177/1475725719868149>.
- Köhler, T., and J. M. Cortina. 2021. “Play It Again, Sam! An Analysis of Constructive Replication in the Organizational Sciences.” *Journal of Management* 47 (2): 488–518. <https://doi.org/10.1177/0149206319843985>.
- Lash, T. L., L. J. Collin, and M. E. Van Dyke. 2018. “The Replication Crisis in Epidemiology: Snowball, Snow Job, or Winter Solstice?” *Current Epidemiology Reports* 5: 175–83.
- Makel, M. C., J. A. Plucker, and B. Hegarty. 2012. “Replications in Psychology Research: How Often Do They Really Occur?” *Perspectives on Psychological Science* 7 (6): 537–42. <https://doi.org/10.1177/1745691612460688>.
- McManus, K. 2024. “Replication Studies in Second Language Acquisition Research: Definitions, Issues, Resources, and Future Directions: Introduction to the Special Issue.” *Studies in Second Language Acquisition* 46 (5): 1299–319. <https://doi.org/10.1017/S0272263124000652>.
- Moreau, D., and K. Wiebels. 2023. “Ten Simple Rules for Designing and Conducting Undergraduate Replication Projects.” *PLOS Computational Biology* 19 (3): e1010957. <https://doi.org/10.1371/journal.pcbi.1010957>.
- Nuijten, M. B., and J. R. Polanin. 2020. “‘Statcheck’: Automatically Detect Statistical Reporting Inconsistencies to Increase Reproducibility of Meta-analyses.” *Research Synthesis Methods* 11 (5): 574–79. <https://doi.org/10.1002/jrsm.1408>.
- Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349 (6251): aac4716. <https://doi.org/10.1126/science.aac4716>.
- Perry, T., R. Morris, and R. Lea. 2022. “A Decade of Replication Study in Education? A Mapping Review (2011–2020).” *Educational Research and Evaluation* 27 (1-2): 12–34. <https://doi.org/10.1080/13803611.2021.2022315>.
- Pownall, M. 2022. “Is Replication Possible for Qualitative Research?” <https://doi.org/10.31234/osf.io/dwxeg>.
- Röseler, L., L. Kaiser, C. Doetsch, N. Klett, C. Seida, A. Schütz, and Y. and Zhang. 2024. “The Replication Database: Documenting the Replicability of Psychological Science.” *Journal of Open Psychology Data* 12 (1): 8. <https://doi.org/10.5334/jopd.101>.
- Schöch, C. 2023. “Repetitive Research: A Conceptual Space and Terminology of Replication, Reproduction, Revision, Reanalysis, Reinvestigation and Reuse in Digital Humanities.” *International Journal of Digital Humanities* 5 (2): 373–403. <https://doi.org/10.1007/s42803-023-00073-y>.
- Urminsky, O., and B. J. Dietvorst. 2024. “Taking the Full Measure: Integrating Replication into Research Practice to Assess Generalizability.” *Journal of Consumer Research* 51 (1): 157–68. <https://doi.org/10.1093/jcr/ucae007>.

A

B