

USER GUIDE

Replication Navigator

User guide, replication handbook and resource pack

Getting started

No sign-in needed. Open Replication Navigator in your browser, answer four quick questions and start. It is free and open access.

Version 0.2 · October 2026 · Built on 322 evidence cards drawn from nine key sources on replication

Contents

1. Quick start	4
2. What Replication Navigator is.....	5
Why an “evidence-grounded” coach?	5
The nine sources	5
3. Using the platform	6
The screen at a glance.....	6
Each stage page.....	6
Working with the AI coach.....	6
Other features.....	7
Saving and moving your work.....	8
Working as a team	8
4. Ways to use Replication Navigator	9
5. The replication process, stage by stage	10
Stage 1. Choose & justify the target.....	10
Stage 2. Define aim & replication type	11
Stage 3. Gather materials & contact authors.....	12
Stage 4. Reproduce before you replicate.....	14
Stage 5. Check measurement & construct validity	15
Stage 6. Plan design & sample size	16
Stage 7. Document differences from the original	17
Stage 8. Preregister	18
Stage 9. Analysis & success criteria.....	20
Stage 10. Interpret, report & publish.....	21
6. Quick-reference tables.....	24
Key terms	24
Six replication goals (Anderson & Maxwell, 2016).....	24
Sample-size approaches.....	25
Describing outcomes in balanced language.....	25
7. Checklists and templates	26
The FORRT reproduction and replication checklist.....	26
Emails to the original authors	26
The Replication Recipe (Brandt et al., 2014).....	27
8. Key resources	28
Core readings (the evidence base).....	28
Further reading	28

Tools, databases and registries	29
9. Research participation and privacy	30
10. Troubleshooting and FAQ	31
Frequently asked questions	31

If page numbers are missing, right-click the table and choose "Update field".

1. Quick start

Everything you need to get going in five minutes. Planning stages 1–9 takes about 1½–2 hours, less if some stages do not apply to you. The rest of this manual explains each part in more detail.

1. **Open the app** in your web browser, ideally on a laptop. There is no sign-in and no account.
2. **Answer four quick questions on the start screen** (your experience, whether the original data are available, whether you translate materials, quantitative or qualitative). The tool hides what you do not need and adjusts how much it explains.
3. **Tick the research box if you are happy to help.** It is optional and the tool works fully either way. See section 9.
4. **Optional: upload the paper you are replicating** on Stage 1. The AI pulls out the sample size, effect size, design and measures to pre-fill your worksheets, and the coach tailors its advice to that study.
5. **Work through the stages.** Each stage starts with a plain-language explanation. Click “?” next to a field for help, and click any underlined term for a short definition.
6. **Check the evidence.** Click any grey-blue citation label (e.g. isager2021-02) to see the source, page and exact quote behind the coach’s advice.
7. **Finish & download.** At the end, “Finish & download” (in the sidebar) gives you your plan and a draft preregistration as Word documents. Then fill in the short feedback survey.

Three things to remember

- Your work is saved automatically in your browser, but only on that device and browser. Export regularly.
- The first coach reply of the day can take up to a minute while the server wakes up. After that, replies start appearing within a few seconds.
- The coach can be wrong. Treat it as a well-read colleague, not an authority, and check citations and key decisions with an expert.

2. What Replication Navigator is

Replication Navigator is a step-by-step guide and AI coach for planning a rigorous replication or reproduction study. It turns the best available guidance on replication into ten practical stages. Each stage has a worksheet, a checklist, resources and a coach. By the end you have a written replication plan that you can discuss with a supervisor, turn into a preregistration, or develop into a paper.

Why an “evidence-grounded” coach?

General chatbots answer from everything they were trained on, so you cannot tell where their advice comes from. This coach is different: it may only use a curated evidence base of 322 evidence cards drawn from nine sources on replication, and it must cite the card behind every recommendation. Each card contains:

- **a short piece of guidance**, paraphrased in plain language;
- **an action** you can tick off;
- **a short verbatim quote** from the source, with its page number, so you can check it.

Every quote has been automatically checked against the original PDF. If the evidence base does not cover a question, the coach is instructed to say so rather than improvise.

The nine sources

Source	What it contributes
Röseler, Wallrich et al. (2025). FORRT Handbook for Reproduction and Replication Studies	The backbone of the process, from choosing a target to publishing, plus the 10-item checklist and email templates
Isager et al. (2023). Deciding what to replicate	How to choose a target: value, uncertainty, costs
Obenauer (2024). Designing, executing, and publishing replication research	Practical advice from purpose to publication, especially in management
Anderson & Maxwell (2016). There’s more than one way to conduct a replication study	Six different replication goals and the analysis each one needs
Bonett (2021). Design and analysis of replication studies	Sample size, equivalence tests and confidence-interval criteria
Brandt et al. (2014). The Replication Recipe	What makes a replication convincing; the 36-question template
Flake, Davidson, Wong & Pek (2022). Construct validity and the validity of replication studies	Measurement and construct validity in replications
Irvine (2021). The role of replication studies in theory building	How replications connect to theory
Schwab et al. (2023). How replication studies can improve doctoral student education	Replication as training for doctoral researchers

Full references with DOIs are in section 8.

3. Using the platform

The screen at a glance

Area	What it does
Left sidebar	The ten stages with a time estimate for each (stages that don't apply to you are greyed out). A number turns into a tick when all checklist items are done. Below: Finish & download, Replication Recipe, Evidence base, Feedback survey, About.
Top bar	Overall progress, Evidence base, Export plan, Preregistration and Settings.
Centre	The current stage: a plain-language explanation, worksheet, checklist, templates and resources.
Right panel	The AI coach. On a phone, open it with the "Ask the coach" button.

Each stage page

- **What this stage is about:** a short plain-language explanation of what you decide and why it matters, with the key questions underneath.
- **Worksheet:** fields for your decisions. Click "?" for an explanation and an example. Underlined terms open a one-sentence definition. Fields marked "optional" add detail; newcomers see them only on request. "Evidence" labels link to the cards behind each field.
- **Checklist:** tick items when your team has genuinely done them. Each item is in plain language, with the wording used in the literature underneath and links to its evidence.
- **Templates:** in stages 3 and 10, email templates for contacting the original authors.
- **Resources:** links to databases, tools and registries.
- **Expert check-points:** stages 1, 6 and 9 carry a ★ badge. Ask an expert to review your decisions before moving on.

Working with the AI coach

Ask your own question, click one of the suggested questions, or use the buttons at the top of the coach panel. The coach sees your worksheet for the current stage, a summary of earlier stages, your setup answers and (if you uploaded it) a summary of your target paper. It answers your question first and keeps replies short; ask it to go deeper if you want more. Replies appear word by word as they are written. If a long reply stops early, click "Continue".

Button	Use it when
Review my worksheet	You have written something and want the 2–3 most important problems named, with concrete fixes.
What am I missing?	You want the most common pitfalls checked against your plan.
Challenge me	You want a critical but fair "reviewer" to test your choices.
Draft author email	Stages 3 and 10: drafts an email to the original authors from the template.

Good questions to ask

- “Is my justification for this target strong enough? What would a reviewer say?”
- “Which replication goal fits what we want to show, and what does that mean for our analysis?”
- “We plan to translate the scale into Italian. What should we check?”
- “Our original study had $N = 60$. How should we think about our sample size?”
- “How should we word our conclusion if the effect is smaller but still positive?”

Checking citations

Every recommendation should end with one or more citation labels, such as isager2021-02. Click a label to open the card: its guidance, the exact quote, the source and the page. If a card does not actually support what the coach said, click “Flag: doesn’t support the coach’s claim”. Flags are valuable data for improving the tool. A label that is struck through means the coach cited a card that does not exist, so treat that statement with extra caution.

What the coach cannot do

- It can misread your situation or cite a card that only partly supports its point.
- It will not calculate your sample size for you. It explains which approach to use and points to tools.
- It knows little about qualitative replication. For those questions, ask a supervisor or methods expert.
- Do not paste personal or confidential data into the coach.

Other features

- **Replication Recipe:** the full 36-question template from Brandt et al. (2014). Filling it in gives you most of a preregistration. Your answers are included when you export.
- **Preregistration:** the “Preregistration” button (top bar, Stage 8 and the Replication Recipe page) downloads a draft preregistration as a Word document in the Replication Recipe format. It uses your Recipe answers and fills blanks from your stage worksheets. Drafted answers are marked “check and edit”; questions still empty are highlighted in yellow. Review everything with your supervisor, then copy it into the Replication Recipe template on OSF Registries.
- **Evidence base:** search and filter all 322 cards by stage and source. Useful for reading around a topic.
- **Export plan:** downloads your whole plan as a formatted Word document: an overview of progress, all your answers stage by stage, your checklists, your Replication Recipe answers, next steps and the evidence cards referenced.
- **Finish & download:** a final page with your plan, your draft preregistration, a completeness overview, next steps and the feedback survey.
- **Upload your target paper:** on Stage 1. The PDF is sent to the AI service only to extract the key details and is not stored. Always check the extracted details against the paper.
- **Settings:** change your setup answers, switch research logging on or off, download or import a backup (.json), or start over.
- **Feedback survey:** a four-minute questionnaire. Your feedback shapes the next version.

Saving and moving your work

Your answers are stored in your browser as you type. They stay there if you close the tab, but they do not follow you to another device or browser, and clearing your browser data deletes them. To continue elsewhere, go to Settings → Download backup, then on the new device go to Settings → Import plan and choose the .json file.

Working as a team

Teams work best with one person “driving” the app on a shared screen while the others read the cited cards, check the sources and challenge decisions. Replications are well suited to collaborative student teams, where members can debate how to implement the study and what to infer from it.

- **Let students replicate collaboratively.** Students working together can debate how to implement the replication, what to infer, and how to write it up, with more experienced students helping less experienced ones. Students can also choose studies using methods they know, or ones that make them try newly learned methods. (Schwab et al. (2023), p. 26; card schwab2023-31)

4. Ways to use Replication Navigator

Stages 1–9 take about 1½–2 hours in total. You do not have to do them in one sitting: your work is saved in your browser, so you can stop after any stage and come back on the same device.

Situation	Suggested approach
On your own (e.g. a PhD student planning a replication)	Upload the paper you are replicating on Stage 1, then work through one or two stages per session. Use “Review my worksheet” before moving on, and take your exported plan to your supervisor at the expert check-points (stages 1, 6 and 9).
As a team	One person “drives” on a shared screen while the others read the cited evidence and challenge decisions. Export the plan at the end of each session and share it with the team.
In a course or workshop	Give each team one pre-selected target study so time goes into the process rather than the search. Plan about two hours for stages 1–9, and have an expert visit each team at the check-points. Ask teams to put their team name in the file name of the exported plan.
After data collection	Come back to Stage 10 to interpret and report your results, invite the original authors to comment, and log your replication in FReD.

5. The replication process, stage by stage

This section is a portable version of the ten stages. For each stage you will find its goal, the key recommendations and pitfalls from the evidence base (with source, page and card id so you can find the card in the app), the checklist, and the main resources.

Stage 1. Choose & justify the target

★ **Expert check-point:** have an expert review your decisions before moving on.

Goal. Pick one clearly identified study and the specific claim you will test, and write down why it is worth replicating now.

Output of this stage. A named target study and claim, with a written rationale covering value, uncertainty, feasibility and existing replications.

Questions to answer

- Which study, and which exact claim or effect, will you test?
- How valuable is it to know whether this claim is true?
- How uncertain is the current evidence?
- Is it feasible with your time, money, participants and skills?

Key recommendations

- **Justify and transparently report target selection.** There is no single correct way to choose what to replicate; it depends on goals (e.g., building on a finding vs. doubting it) and practical constraints. What matters is that the choice is well justified and openly communicated. A structured replication target selection checklist (Pittelkow et al., 2023) can help. (Röseler, Wallrich et al. (2025), p. 14; card forrt2025-09)
- **Replication value: value times uncertainty.** Isager et al. define a claim's replication value as the most expected utility a replication could add. It depends on two things: how valuable it would be to know for certain whether the claim is true, and how uncertain the current evidence leaves us. A candidate scores high only when both are substantial. (Isager et al. (2023), p. 1; card isager2021-02)
- **Weigh expected gain against cost.** When candidates have similar replication value, prefer the one that delivers the most knowledge per unit of resources, the idea of marginal utility per dollar. Lower cost, all else equal, means a higher gain per resource spent, so cost belongs on your worksheet next to value and uncertainty. (Isager et al. (2023), p. 8; card isager2021-08)
- **Search for existing reproductions and replications.** The first reproduction or replication reduces uncertainty the most, so check what already exists. Useful resources include the Institute for Replication discussion papers, ReplicationWiki, the CODECHECK register, the Social Science Reproduction Platform, and the FORRT Replication Database, supplemented by manual searches. (Röseler, Wallrich et al. (2025), p. 19; card forrt2025-15)
- **Run an outcome thought experiment.** After preliminary target selection, list every combination of replicated and non-replicated findings (e.g., three hypotheses give eight combinations) and state the contribution each would make. If you cannot identify a contribution for every combination, reconsider whether the target is appropriate. (Obenauer (2024), p. 14; card obenauer2024-12)

Pitfalls to avoid

- **Don't replicate what is already settled.** A replication adds little if the claim is either not worth knowing or already firmly established; the authors illustrate this with repeated replications of a classroom-staple

effect. Before investing effort, check whether prior replications already leave little uncertainty about your candidate claim. (Isager et al. (2023), p. 8; card isager2021-07)

- **A flawed original design limits what replication can teach.** A claim can be valuable and uncertain yet still be a poor target: if the uncertainty stems mainly from a low-quality original design that the replication would repeat, replicating it will barely reduce uncertainty. Check that the original design is good enough for a repetition of it to be informative. (Isager et al. (2023), p. 10; card isager2021-13)
- **Replicators are biased too.** Original authors may have confirmation bias, motivated reasoning or financial conflicts, but replicators are equally prone to bias: interest in failed replications may tempt 'null hacking', and personal admiration or envy may drive target choice. Preregistration and Registered Reports largely address p-hacking and null-hacking. (Röseler, Wallrich et al. (2025), p. 20; card forrt2025-16)
- **Replication value supports ranking, not ratios.** Because replication value assumes all uncertainty could be removed, it overestimates the actual gain; it is useful for ranking candidates, not for saying one is twice as worth replicating as another. Use your scores to order candidates, not to make interval-scale claims. (Isager et al. (2023), p. 8; card isager2021-24)

Checklist

- ☐ We named one study and the one specific result we will test obenauer2024-06
- ☐ We searched for existing replications of this result (e.g. in FReD) forrt2025-15
- ☐ We explained why the result matters and why we are still unsure about it isager2021-02, forrt2025-10
- ☐ We checked we have the time, money and participants to do it isager2021-10
- ☐ We checked the original study's design is good enough to be worth repeating isager2021-13, flake2023-09
- ☐ We wrote down why we chose this study forrt2025-133, isager2021-22

Resources

- [FORRT Replication Database \(FReD\) explorer](#)
- [FORRT Replication Hub](#)
- [Isager et al. \(2023\) decision model](#)

Stage 2. Define aim & replication type

Goal. State the purpose of your study, whether it is a reproduction or replication, how close it is to the original, and which replication goal your analysis will serve.

Output of this stage. A stated purpose, replication type and replication goal (1–6) that will drive your sample size and analysis.

Questions to answer

- What do you want to learn: whether the effect exists, how large it is, whether it generalises?
- Close (direct) or conceptual (constructive) replication, and why?
- How does this connect to theory?

Key recommendations

- **Identify your purpose before your target.** Decide why you are replicating before choosing what to replicate; the purpose guides target selection and how you frame the work for your audience. Possible

purposes include updating a theory after a relevant event, testing an established theory with stronger methods, probing boundary conditions, advancing a methodological precedent, doctoral training, or meeting a special-issue call. (Obenauer (2024), p. 4; card obenauer2024-03)

- **Replication has more than one legitimate goal.** Anderson and Maxwell argue that most replication studies judge success only by whether $p < .05$ comes out the same way as in the original. That is one goal among several. Before you design your study, ask what question your replication is really meant to answer: whether the effect exists, whether it is null, how big it is, or how it compares with the original. (Anderson & Maxwell (2016), p. 2; card anderson2016-01)
- **Close replications test credibility; conceptual ones test generalizability.** Since deviations from the original have mostly uncertain consequences, close replications most directly test whether original results are credible. Conceptual replications that vary design features instead address generalizability. Seemingly irrelevant design choices (e.g., text vs. image stimuli) have produced opposite conclusions in past work. (Röseler, Wallrich et al. (2025), p. 11; card forrt2025-05)
- **Close replication before conceptual replication.** To build generalizability, start with replications as close as possible to the original, and only later move to conceptual ones. From constructive replications onward, try to improve the original (more valid measures, better controls, more representative samples) and avoid 'quasi-random' changes lacking clear rationale. (Röseler, Wallrich et al. (2025), p. 27; card forrt2025-24)

Pitfalls to avoid

- **Procedural fidelity may undermine theoretical equivalence.** Contextually sensitive topics (dependent on time, culture, or population) replicated less often even with high methodological fidelity. Reusing outdated stimuli (e.g., old celebrities) may be psychologically inert, so strict copying can make a replication less valid. Consider whether adapting stimuli better preserves the intended psychological process. (Röseler, Wallrich et al. (2025), p. 10; card forrt2025-04)
- **Close replication will not fix original confounds.** A close replication inherits any confounds or theoretical limitations of the original design. To remove confounds or extend generalizability, add conceptual replications or replication-plus-extension designs. When your interest is a theoretical prediction rather than an effect in one paradigm, a mix of close and conceptual replications builds the most confidence. (Brandt et al. (2014), p. 222; card brandt2014-25)

Checklist

- ☐ We wrote down our purpose before planning the details obenauer2024-03
- ☐ We chose a replication type (close or conceptual) and said why obenauer2024-28, forrt2025-134
- ☐ We chose a replication goal (it decides our analysis in Stage 9) anderson2016-03
- ☐ If we change the design, we also include a close replication of the original (or explain why not) forrt2025-13, forrt2025-24
- ☐ We checked our study really tests the original claim forrt2025-06

Resources

- [FORRT Handbook — types of replication](#)

Stage 3. Gather materials & contact authors

Goal. Collect everything needed to rerun the original (materials, measures, procedure, data, code) and contact the original authors professionally.

Output of this stage. A materials inventory, a log of contact with the original authors, and a record of how gaps were filled.

Questions to answer

- Which materials are public, which are missing?
- Have you written to the original authors?
- Have you read post-publication discussion of the study?

Key recommendations

- **Start by contacting the original authors.** To follow the original methods as exactly as possible, begin by asking the original authors for their materials (Recipe Question 10). If they are not available, check whether the materials exist elsewhere, such as previously published scales, and record how any materials you had to create were built. Where feasible, work with the original authors; the authors note that much of the Recipe is easier with their cooperation. (Brandt et al. (2014), p. 219; card brandt2014-06)
- **Contacting original authors effectively.** Expect to contact authors more than once, as missing details surface during planning. Verify the corresponding author's current email by web search, consider writing to the last author or copying all authors, and use the handbook's email templates (Appendix E). Some authors may not respond. (Röseler, Wallrich et al. (2025), p. 30; card forrt2025-26)
- **When original materials cannot be obtained.** If the original authors are uncooperative or cannot be reached, rebuild the methods as well as you can from the original methods section, assuming the original study was highly rigorous. Fill any gaps using best practice for that paradigm (e.g., for a lexical decision task), and ask expert colleagues in the area whether your version faithfully recreates the study as described. Brandt et al. say replications should still go ahead when collaboration is not possible. (Brandt et al. (2014), p. 219; card brandt2014-07)
- **Search post-publication discussions.** Comments, critiques, error reports, reinterpretations and replication suggestions about the target may change how you should design your study. Look for published comments, reviews, blog posts and social media discussion via Altmetric, PubPeer, or Hypothes.is. (Röseler, Wallrich et al. (2025), p. 25; card forrt2025-18)

Pitfalls to avoid

- **Expect missing information in the original.** Studies cited by Schwab et al. could not reproduce about 70% of sampled published articles because crucial information was missing, and replications that need new data likely face even bigger gaps. Check how completely your target is reported before committing. (Schwab et al. (2023), p. 22; card schwab2023-07)

Checklist

- ☐ We have, or have asked for, all the original materials forrt2025-135
- ☐ We emailed the original authors (template below) forrt2025-143, brandt2014-06
- ☐ We wrote down anything missing and how we filled the gap schwab2023-07, flake2023-07
- ☐ We considered asking the original authors for feedback on our design forrt2025-144, forrt2025-102, forrt2025-101

Resources

- [PubPeer \(post-publication discussion\)](#)
- [Open Science Framework](#)

Stage 4. Reproduce before you replicate

Goal. Where data are available, rerun the original analysis on the original data to check the reported result before collecting new data.

Output of this stage. A reproduction of the original result (or a reason why that is not possible), with software versions documented.

Questions to answer

- Are data and code available?
- Do you get the same numbers?
- Is the result robust to small, reasonable analytic changes?

Key recommendations

- **Reproduce before you replicate.** Many features of a replication rest on the original report being correct. A reproduction can reveal coding errors, fraud, and sensitivity to analytic decisions, and can take minutes to weeks. To use resources well, check reproducibility and robustness before planning a replication. (Röseler, Wallrich et al. (2025), p. 25; card forrt2025-19)
- **Identify and justify the claim you reproduce.** Analyses always serve to evaluate a claim. Focus typically on results essential to the paper's main argument, though secondary results can be justified if relevant elsewhere. Either way, justify your choice of claim in the report. (Röseler, Wallrich et al. (2025), p. 30; card forrt2025-27)
- **Document your computing environment.** Reproduction success can depend on the machine's software, program versions, and packages. In R, include `sessionInfo()` output in supplementary materials; Python users can use `session_info.show()`. (Röseler, Wallrich et al. (2025), p. 31; card forrt2025-29)
- **Run small robustness alterations.** Test the finding's robustness with small changes to processing and analysis, e.g., a different age range or a ± 2 instead of ± 3 SD outlier cut-off. Start with choices not determined by theory. If the original was preregistered, check whether its analyses followed the plan. (Röseler, Wallrich et al. (2025), p. 26; card forrt2025-22)

Pitfalls to avoid

- **Software versions and missing seeds cause discrepancies.** Even with original code and data, different software versions or defaults can produce small deviations, and without a set random seed, analyses like bootstrapping cannot be reproduced exactly. Anticipate and document such differences rather than treating them as errors. (Röseler, Wallrich et al. (2025), p. 26; card forrt2025-20)

Checklist

- ☐ We tried to reproduce the original result with the original data (or noted why we can't) forrt2025-19, forrt2025-136, obenauer2024-16
- ☐ We checked that the numbers reported in the paper are consistent with each other forrt2025-23
- ☐ We noted the software and versions we used forrt2025-29
- ☐ We tried a few small, reasonable changes to the analysis to see if the result holds forrt2025-22

Resources

- [BITSS Guide for Advancing Computational Reproducibility \(ACRE\)](#)
- [Institute for Replication](#)

- [CODECHECK register](#)

Stage 5. Check measurement & construct validity

Goal. Make sure the measures in the original and in your replication produce valid scores. Without that, a success or failure is hard to interpret.

Output of this stage. A summary of validity evidence for each key measure, and the checks you will run in your own data.

Questions to answer

- What validity evidence exists for the original measures?
- Will you translate or adapt anything?
- How will you check the measure works in your sample?

Key recommendations

- **Validity belongs to score interpretations, not instruments.** Validity is not a yes/no property of a questionnaire; it is a judgment about how scores are interpreted, based on accumulating evidence. Because the same instrument can behave differently in a new context or population, evidence gathered in the original study does not automatically carry over to your replication sample. (Flake et al. (2022), p. 4; card flake2023-02)
- **Build minimum validity checks into the replication.** If the original provides little validity evidence and you still replicate it, gather evidence yourself. At a minimum: review items or stimuli for face and content validity, check the assumed scoring structure (e.g., component or factor analysis), and report an appropriate reliability estimate. Treat these like a manipulation check or sample size plan. (Flake et al. (2022), p. 21; card flake2023-11)
- **Test measurement invariance after translation.** Translation standards vary: personality scales are usually validated with invariance tests, while social psychology often uses only back-translation. In any field, comparing effect sizes across samples requires measurement invariance, which should be tested rather than assumed where possible. (Röseler, Wallrich et al. (2025), p. 38; card forrt2025-42)

Pitfalls to avoid

- **Cronbach's alpha is not validity evidence.** Many researchers rely on alpha as the only support for a score, but reliability shows consistency, not meaning: scores can be reliable yet not represent the intended construct. Alpha is also only interpretable under a single-factor model with equal loadings, so do not treat a high alpha as sufficient validation. (Flake et al. (2022), p. 19; card flake2023-05)
- **Translated scales need validation.** Forty RPP scales were translated, and only 20% used an existing translation with validity evidence. Even careful translation with multiple translators and back-translation is not enough: scores from translated scales need validation to confirm they capture the construct comparably. Validated translations of some scales (e.g., PANAS) have shown non-equivalence. (Flake et al. (2022), p. 17; card flake2023-19)
- **Don't interpret the effect if measures failed.** If evidence shows that foundational aspects (like measurement) are not comparable between original and replication, interpreting the downstream hypothesis test does not make sense, just as you would not test group differences after a failed manipulation. Measurement differences also offer an explanation for replication failure beyond a Type I error in the original. (Flake et al. (2022), p. 24; card flake2023-23)

Checklist

- ☐ For each main measure we noted the evidence that it works: at least reliability (e.g. α) and ideally a validation study or factor structure flake2023-08, flake2023-12
- ☐ We planned basic checks of the measures in our own data (e.g. reliability) flake2023-11
- ☐ If we translate a measure, we planned how to check the translation works the same way flake2023-21, forrt2025-42
- ☐ We decided what we will do if a measure does not work in our sample flake2023-23, flake2023-18

Resources

- [Flake et al. \(2022\) review materials \(OSF\)](#)

Stage 6. Plan design & sample size

★ **Expert check-point:** have an expert review your decisions before moving on.

Goal. Choose a sample-size strategy that fits your replication goal. Don't simply reuse the original N.

Output of this stage. A planned sample size with a full justification that fits your replication goal.

Questions to answer

- What is the original effect size and its confidence interval?
- Which planning approach fits your goal?
- What is the smallest effect size of interest?

Key recommendations

- **Use a conservative effect-size planning value.** The original estimate contains sampling error, so using it directly as the planning value can mislead. An alternative is the 80% one-sided lower confidence limit (for positive estimates), approximately the estimate $- 0.84 \times SE1$. If SE1 was not reported, derive it from other reported statistics. (Bonett (2021), p. 525; card bonett2021-19)
- **Small telescopes approach.** Test whether the replication effect is smaller than the effect the original study had 33% power to detect; if so, the original can be treated as uninformative. This suits imprecise originals with wide confidence intervals. It needs 2.5 times the original N for 80% power, but the handbook recommends 95% power, requiring about 4.5 times the original N. (Röseler, Wallrich et al. (2025), p. 35; card forrt2025-36)
- **Equivalence testing with a SESOI.** If you can justify a smallest effect size of interest (SESOI), plan to test whether the replication effect is significantly smaller than it using a one-sided test in the 'lesser' direction, since the original fixes the direction. Lakens et al. (2018) provide a tutorial. (Röseler, Wallrich et al. (2025), p. 36; card forrt2025-37)
- **Goal 3: plan sample size for precision (AIPE).** If you care about how big the effect is, use accuracy in parameter estimation (AIPE) to plan your sample. Choose n to reach a target confidence interval width, not a target power. For example, estimating d within a half-width of 0.15 took about 350 participants per group. (Anderson & Maxwell (2016), p. 6; card anderson2016-15)

Pitfalls to avoid

- **Don't reuse the original sample size.** Although you should stay close to the original otherwise, replication samples usually need to be larger: an informative failure must provide evidence for a null or a

substantially smaller effect. See Lakens (2022) for general sample size justification. (Röseler, Wallrich et al. (2025), p. 34; card forrt2025-34)

- **Power based on original effect sizes is inflated.** Published effect sizes tend to be too large because of publication bias and sampling error. If you base your power analysis on the original effect size, your real power can be much lower than you planned. A study that reports power of .80 may in fact have much less. The authors point to Taylor and Muller (1996) for a power method that takes bias and uncertainty into account. (Anderson & Maxwell (2016), p. 3; card anderson2016-13)
- **Showing a null effect needs a very large sample.** To show that the follow-up effect is trivially small, the whole CI must fall inside a narrow range around zero, which can require a large sample. The required sample size grows sharply as the true effect approaches the boundary. If this is your goal, plan the sample with that in mind or accept that it may not be realistic. (Bonett (2021), p. 522; card bonett2021-13)
- **Never derive effect sizes from small pilots.** Pilot when resources are high or materials are new, but not as an underpowered version of the study. Use pilots to find methodological flaws or test assumptions; small pilots are too imprecise for power analyses. A few pilot participants plus a debriefing survey can check instructions, procedure and data recording. (Röseler, Wallrich et al. (2025), p. 40; card forrt2025-45)

Checklist

- ☐ We did not just copy the original sample size forrt2025-34
- ☐ We allowed for the original effect probably being overestimated anderson2016-13, bonett2021-19
- ☐ Our sample-size approach matches our goal from Stage 2 anderson2016-02
- ☐ If we want to show there is no effect: our sample is large enough for that bonett2021-13, anderson2016-12
- ☐ We did not base our sample size on a small pilot study forrt2025-45

Resources

- [G*Power](#)
- [TOSTER \(equivalence tests, R\)](#)
- [MBESS \(AIPE, R\)](#)

Stage 7. Document differences from the original

Goal. List every difference between your study and the original, rate each one, and think ahead about which differences could change the result.

Output of this stage. A table of every difference from the original (Exact / Close / Different) and which ones might matter.

Questions to answer

- What differs: population, setting, language, time, materials, procedure?
- Which differences might plausibly matter, and why?
- Can you test any of them?

Key recommendations

- **Rate each component Exact, Close, or Different.** For the instructions, measures, stimuli, procedure, location, remuneration and participant population (Recipe Questions 17–23), rate your study as exact,

close, or conceptually different from the original, and justify each rating. This lets reviewers judge whether deviations were justified, shows where your study sits on the close-to-conceptual continuum, and helps later work identify boundary conditions once several replications exist. (Brandt et al. (2014), p. 219; card brandt2014-13)

- **Anticipate and test influential differences.** Identify differences between your study and the original that might change the size or direction of the effect (Recipe Question 24), such as participants' socioeconomic status or mood at the time of testing. Measure such a variable at the end of the study so it does not interfere with the close replication, then test it as a moderator within your sample (Question 25). (Brandt et al. (2014), p. 220; card brandt2014-14)
- **Discuss every deviation and its likely impact.** Evaluate each replication outcome in light of how close the replication was. Discuss all deviations, the reasons for them and, where possible, their potential impact on results. Use existing theory to judge whether a deviation should matter; for instance, most psychological theories are agnostic to age, so a different age distribution is usually unproblematic. (Röseler, Wallrich et al. (2025), p. 54; card forrt2025-121)
- **Keep intentional changes focused and narrow.** Every intentional change to the precedent design is a 'slippery slope' because it adds alternative explanations for any difference in results. To keep comparisons with the precedent study interpretable, keep deliberate changes highly focused and narrow. (Schwab et al. (2023), p. 21; card schwab2023-04)

Pitfalls to avoid

- **Same scale, same protocol, different measurement.** Measurement differences arise not only from changed item wording, presentation or response options, but also when the same scale and protocol are used. Psychometric analyses in the replication sample may show that you are not measuring the same construct as the original, which undermines comparing scores across studies. (Flake et al. (2022), p. 14; card flake2023-15)
- **Seemingly minor factors get questioned later.** Commentaries on the Many Labs projects questioned many kinds of difference: associations with national symbols, which groups count as out-groups, ambient lab temperature, running several studies in one session, and changes in how stimulus features are rated. Expect such factors to be raised and record them in advance. (Irvine (2021), p. 846; card irvine2021-07)

Checklist

- ☐ We rated every part of the study as Exact, Close or Different brandt2014-13
- ☐ We identified the differences that could change the result brandt2014-14
- ☐ We kept changes to the minimum needed and explained each one forrt2025-137, forrt2025-121
- ☐ We considered differences in how things are measured, not only in procedure flake2023-15, flake2023-20

Resources

- [Replication Recipe template \(OSF\)](#)

Stage 8. Preregister

Goal. Fix hypotheses, design, exclusions, analysis and success criteria before seeing the data, or submit as a Registered Report.

Output of this stage. A preregistration or Registered Report plan covering hypotheses, exclusions, analyses, success criteria and contingencies.

Questions to answer

- Where will you preregister?
- Is a Registered Report possible?
- What will you do if things go wrong (contingencies)?

Key recommendations

- **Preregister analysis plan and success criteria.** Because replications meet skepticism, preregister the replication including the analysis plan (ideally code tested on simulated or pilot data) and criteria distinguishing success from failure. A preregistration without an analysis plan does not guard against p-hacking. Criteria can be sequential, e.g., requiring a working manipulation check. (Röseler, Wallrich et al. (2025), p. 33; card forrt2025-31)
- **Consider a registered report.** In a registered report, peer review takes place before data collection. Obenauer's team benefited because their response to a failed manipulation pretest had already been defined and accepted by reviewers, and a minimum sample size had been pre-approved. (Obenauer (2024), p. 16; card obenauer2024-22)
- **Plan for contingencies in advance.** Pretest failures, attention and manipulation checks, and response rates matter even more in replication because they can determine whether you report success. Specify ahead of time what you will do if, for example, a manipulation pretest fails or a minimum sample size is not reached. (Obenauer (2024), p. 16; card obenauer2024-20)
- **Preregister one central exclusion rule.** Original exclusion criteria may conflict with current best practice. Running the analysis both ways is a sensible robustness check, but you must preregister one of the options as the central analysis so the choice is not made after seeing the data. (Röseler, Wallrich et al. (2025), p. 43; card forrt2025-108)

Pitfalls to avoid

- **Handle deviations from the preregistration openly.** Deviating from your plan is not wrong in itself, but list and discuss all changes and show how they affected the results. If you notice changes during data collection, upload amendments with version history on your registration platform. (Röseler, Wallrich et al. (2025), p. 34; card forrt2025-33)

Checklist

- ☐ Our analysis plan and success criteria are written down and registered forrt2025-31, forrt2025-138, brandt2014-16
- ☐ We specified the main result(s) we will test forrt2025-111
- ☐ We decided our exclusion rules in advance forrt2025-108
- ☐ If we test for 'no effect': we set the equivalence range in advance anderson2016-09
- ☐ We have a plan for reporting any deviations from the preregistration forrt2025-33

Resources

- [OSF Registries](#)
- [AsPredicted](#)
- [PCI Registered Reports](#)

Stage 9. Analysis & success criteria

★ **Expert check-point:** have an expert review your decisions before moving on.

Goal. Choose analyses that match your replication goal and define success and failure before you analyse the data.

Output of this stage. A primary analysis matched to your goal, and success / failure / inconclusive criteria fixed in advance.

Questions to answer

- Which analysis matches your goal from stage 2?
- What counts as success, failure, and inconclusive?
- Will you run the original analysis, an updated one, or both?

Key recommendations

- **Six replication goals, analyses and success criteria.** The article's Table 1 pairs each goal with an analysis and a success criterion. (1) Show the effect exists: repeat the original analysis, success = $p < .05$. (2) Show a null effect: equivalence test, success = CI entirely inside the region of equivalence. (3) Estimate the effect size precisely: AIPE plus a CI for the effect size. (4) Combine with the original: CI for the average effect size. (5) Show clear inconsistency: CI for the difference in effect sizes excludes 0. (6) Show clear consistency: equivalence test on the difference in effect sizes. (Anderson & Maxwell (2016), p. 2; card anderson2016-02)
- **Fix exclusions, analysis plan, and success definition.** Before data collection, write down your exclusion criteria, such as outlier handling (Question 26), and your analysis plan, justifying any differences from the original analysis (Question 27). Also define in advance what counts as a successful replication (Question 28). (Brandt et al. (2014), p. 219; card brandt2014-18)
- **Run both the original and an updated test.** If methodological advances suggest the original test was not robust, run the original test (for comparability) and the analysis most appropriate by current standards, such as both a t-test and a mixed-effects model. If you can get the original data, you may also re-run the updated analysis on it, which makes interpreting the comparison easier. (Röseler, Wallrich et al. (2025), p. 42; card forrt2025-106)
- **Test similarity with an equivalence test (ROPE).** To show that the original and follow-up effects are similar, use an equivalence test. Choose h , a difference small enough to be unimportant, which defines a range of practical equivalence (ROPE) from $-h$ to h . If the CI for the difference falls entirely inside the ROPE, the effects are equivalent; if it falls entirely outside, they differ meaningfully; otherwise the result is inconclusive. A $100(1 - 2\alpha)\%$ CI (e.g., 90%) is typically used. (Bonett (2021), p. 517; card bonett2021-07)

Pitfalls to avoid

- **Treat nonsignificant results as inconclusive.** Use the confidence interval for directional two-sided tests. If the upper limit is below 0, conclude the effect is negative; if the lower limit is above 0, conclude it is positive; otherwise the result is inconclusive, not evidence that the effect is zero. The same logic applies to the CI for the difference between the original and follow-up effects. (Bonett (2021), p. 516; card bonett2021-06)
- **Nonsignificance is not evidence of no effect.** Many published replications treat $p > .05$ as showing that the effect is absent, but a significance test cannot show that. If your analysis was built to detect an effect, describe a nonsignificant result as a failure to detect it, not as proof of a null. If you want to claim a null, use equivalence tests or Bayesian methods. (Anderson & Maxwell (2016), p. 2; card anderson2016-07)

- **Do not compare p values or eyeball CI overlap.** Saying the studies differ because one had $p < .05$ and the other $p > .05$ is not valid, because the difference between significant and nonsignificant may not itself be significant. Checking whether two confidence intervals overlap is also wrong: intervals that overlap can still come from estimates that differ significantly. (Anderson & Maxwell (2016), p. 8; card anderson2016-22)
- **Don't just check if your CI covers the original estimate.** Another common criterion calls a replication a failure if the follow-up confidence interval excludes the original point estimate. This ignores the fact that the original estimate has its own sampling error. Compare the two studies using a confidence interval for the difference between the effect sizes instead. (Bonett (2021), p. 514; card bonett2021-02)

Checklist

- ☐ We defined success and failure before seeing the data forrt2025-139, brandt2014-18
- ☐ Our analysis matches the goal we chose in Stage 2 anderson2016-02
- ☐ We don't rely only on 'significant vs. not significant' bonett2021-01, anderson2016-07, brandt2014-19
- ☐ We won't compare p-values or eyeball overlapping confidence intervals anderson2016-22, bonett2021-02
- ☐ We planned how to handle careless responders and failed manipulation checks forrt2025-104, forrt2025-105

Resources

- [TOSTER \(R\)](#)
- [APA JARS — quantitative replication reporting table](#)

Stage 10. Interpret, report & publish

Goal. Interpret the results fairly, report everything openly, invite the original authors to comment, and find the right outlet.

Output of this stage. A balanced interpretation, open materials and data, comments invited from the original authors, and a target outlet.

Questions to answer

- What does the evidence, taken as a whole, say?
- Which explanations for divergence are plausible, and testable?
- How will you frame the paper constructively?

Key recommendations

- **Use descriptive, impersonal language.** A failed replication can harm the original authors' reputation, so write descriptively and impersonally. When criticizing poor documentation, unavailable data or brief methods sections, keep the historical context in mind; sharing data was much harder in the 1990s and was not required in many fields until recently. (Röseler, Wallrich et al. (2025), p. 57; card forrt2025-125)
- **Frame replication as building, not attacking.** Motivate replication as contributing to cumulative science rather than as correcting sampling error or questionable practices in the original. Arguments about flaws are falsifiable and can be rebutted, while a constructive framing invites reviewers to be partners in building better science. (Obenauer (2024), p. 18; card obenauer2024-25)
- **Invite original authors to comment.** If your results do not match the original, you can ask the original authors for a comment to publish alongside your replication report; the appendix has a template. Some

journals, such as the Journal of Comments and Replications in Economics, require such statements at submission. (Röseler, Wallrich et al. (2025), p. 55; card forrt2025-123)

- **Share data, syntax, and all analyses.** After the study, make your data, analysis syntax, and all analyses available so others can fully evaluate the replication and test alternative explanations (Questions 34–35). Exceptions can be made for data protection, e.g., data that are hard to anonymize. The authors suggest the Open Science Framework for sharing. (Brandt et al. (2014), p. 220; card brandt2014-23)
- **Log results in FReD and PubPeer.** Beyond a paper, make your result visible. Enter it in the FORRT Replication Database (FReD) through its guided survey, and post a comment on the original study on PubPeer describing the outcome with links or DOIs. Readers using the PubPeer browser plug-in will then see that a replication exists. These options can be combined with formal publication. (Röseler, Wallrich et al. (2025), p. 58; card forrt2025-126)

Pitfalls to avoid

- **Don't overstate replication failure.** It is tempting to overstate findings, for example by writing 'we failed to replicate Study X' when some ancillary findings did replicate. Be precise, and follow best practice in reporting boundary conditions and limitations; this builds confidence and makes the work read as less adversarial. (Obenauer (2024), p. 23; card obenauer2024-34)
- **Move beyond vague hidden-moderator claims.** Unknown or 'hidden' moderators, ranging from lab brightness to cultural differences, can influence results, and they are often used to dismiss unwelcome replication results. The same argument could cast doubt on any original study, so a skeptic who stops there is not taken seriously. Discuss concrete, theory-based candidate moderators instead. (Röseler, Wallrich et al. (2025), p. 54; card forrt2025-119)
- **A single replication is never conclusive.** Hidden moderators mean that one replication is never entirely conclusive, although it can raise enough doubt to question further investment. A failure shows that the original claim does not generalize to the replication's setting; whether it also calls the original setting into question must be argued on theoretical grounds. (Röseler, Wallrich et al. (2025), p. 54; card forrt2025-120)
- **A smaller effect is not automatically a nonreplication.** Finding a smaller or nonsignificant effect, even with a bigger sample, does not show that your result differs from the original. In the authors' example, a replication with $d = 0.15$ against an original $d = 0.5$ gave a difference CI of $[-0.17, 0.87]$, so no difference could be claimed. Precise comparisons need adequately large samples in both studies. (Anderson & Maxwell (2016), p. 9; card anderson2016-23)

Checklist

- ☐ We describe the outcome in neutral, balanced language forrt2025-140, obenauer2024-34
- ☐ We carefully considered reasons for any differences from the original forrt2025-141
- ☐ We share our materials, data and code openly forrt2025-142, forrt2025-124
- ☐ We invited the original authors to comment on our results forrt2025-123, forrt2025-145
- ☐ We logged our result in the FORRT Replication Database (FReD) forrt2025-126
- ☐ We don't treat one replication as the final word forrt2025-120, irvine2021-23

Resources

- [Submit a replication to FReD](#)
- [Journal of Management Scientific Reports](#)
- [Peer Community In \(preprint review\)](#)

- [ARIM — How-to resources](#)

6. Quick-reference tables

Key terms

Term	Meaning
Reproduction	Re-running the original analysis on the original data. It tests whether the reported result follows from the data.
Replication	Repeating the study with new data. It tests whether the finding holds again.
Close (direct, literal) replication	As similar to the original as possible. It tests the credibility of the original finding.
Conceptual (constructive) replication	Tests the same claim with deliberately different methods, samples or settings. It tests generalisability.
Replication-extension	A close replication plus added conditions or measures that extend the original.
Replication value	How much a replication could add: the value of knowing the claim's truth combined with how uncertain the current evidence is (Isager et al.).
SESOI	Smallest effect size of interest: the smallest effect you would consider meaningful. It is needed for equivalence tests.
Equivalence test (TOST)	Tests whether an effect is small enough to be considered negligible, using two one-sided tests.
Small telescopes	Tests whether the replication rules out an effect the original study could have reliably detected. A common rule of thumb is a sample about 2.5 times the original.
Preregistration / Registered Report	Fixing your hypotheses and analysis plan before seeing the data. A Registered Report is peer-reviewed and accepted in principle before data collection.

Six replication goals (Anderson & Maxwell, 2016)

Decide your goal in Stage 2. It determines your sample size in Stage 6 and your analysis in Stage 9.

Goal	Analysis	Success looks like
1. Show the effect exists	Repeat the original analysis	Significant effect in the same direction
2. Show the effect is absent or negligible	Equivalence test (TOST, 90% CI) or Bayes factor	CI lies entirely inside a region of equivalence set in advance
3. Estimate the effect size precisely	Plan for precision (AIPE); report a CI around the effect	A CI narrow enough to be informative

Goal	Analysis	Success looks like
4. Combine with the original	Small meta-analysis of original + replication	A pooled estimate and CI (pooling does not fix publication bias)
5. Show the results are clearly inconsistent	CI for the difference between the two effect sizes	The CI for the difference excludes 0
6. Show the results are clearly consistent	Equivalence test on the difference in effect sizes	The difference is within equivalence bounds

Source: Anderson & Maxwell (2016), p. 2; card anderson2016-02.

Sample-size approaches

Approach	When to use it	Card
Power for a conservative effect size	Goal 1, when the original effect is likely inflated	bonett2021-19
Small telescopes ($\approx 2.5 \times$ original N)	Testing whether the original study could have detected the effect	forrt2025-36, brandt2014-12
Equivalence with a SESOI	Goal 2: showing the effect is negligible	forrt2025-37
Precision / AIPE	Goal 3: estimating the effect size accurately	anderson2016-15
Bayesian planning	When you will quantify evidence with Bayes factors	forrt2025-38
Multilab / pooled	When one site cannot reach the needed sample	forrt2025-39

Describing outcomes in balanced language

Instead of...	Consider...
"The replication failed."	"The replication did not find evidence for the effect under the conditions tested."
"Non-significant, so there is no effect."	"The result is inconclusive: the confidence interval includes both zero and meaningful effects."
"The original authors got it wrong."	"Our results diverge from the original; possible reasons include ..."
"This proves the effect is real."	"This adds to the evidence for the effect; one replication is not conclusive."

Based on forrt2025-125, bonett2021-14, obenauer2024-34 and forrt2025-120.

7. Checklists and templates

The FORRT reproduction and replication checklist

The ten-item checklist from the FORRT Handbook (Röseler, Wallrich et al., 2025). Use it as a final check before preregistering and again before submitting.

- ☐ 1. Justify choice of target study and claims
- ☐ 2. Choose a reproduction/replication type that aligns with your aims
- ☐ 3. Gather and review all relevant materials
- ☐ 4. Reproduce before you replicate, where possible
- ☐ 5. Discuss all updates, changes, and extensions of the original materials (as close as possible, as updated as necessary)
- ☐ 6. Preregister your study and analysis plan
- ☐ 7. Predetermine conditions for success and failure
- ☐ 8. Use balanced language when describing the outcomes
- ☐ 9. Carefully evaluate outcomes and potential reasons for divergences
- ☐ 10. Report your research comprehensively and openly accessible

Source: FORRT Handbook, Röseler, Wallrich et al. (2025), p. 68–69.

Emails to the original authors

Contacting the original authors is professional courtesy and often essential for getting materials. These outlines follow the templates in the FORRT Handbook appendix. The coach can draft them for you (“Draft author email” in stages 3 and 10).

Requesting materials and data

Address the original authors by name and explain that you are replicating their research, naming the phenomenon and study and briefly saying why it matters. List exactly which materials, data and code you need so you can design a replication as close as possible to the original, cite the original study, and close politely.

Feedback on your replication paradigm

Tell the original authors which study you plan to replicate and why, and share a link to a mock-up of your paradigm. Ask whether it accurately captures their design and invite comments on any details that differ. State your planned sample, and if it differs from the original population, acknowledge the deviation and say you will note it prominently in any write-up.

Inviting comment on replication results

Tell the original authors which study you replicated and summarize your results. Offer them the chance to comment, explain where you plan to publish (e.g., a venue that asks for original-author comments) and say that any comment will be part of the report. Include the citation of the original study and a link to, or attachment of, your replication report.

Tone

Keep emails short, specific and friendly. Name the study and the materials you need, say why the study matters, and thank the authors. A replication is not an accusation.

The Replication Recipe (Brandt et al., 2014)

The app contains all 36 questions under “Replication Recipe” in the sidebar. They are grouped into six sections: the nature of the effect, designing the replication study, documenting differences between the original and replication study, analysis and replication evaluation, registering the replication attempt, and reporting the replication. The official template is on OSF.

- [Replication Recipe template on OSF](#)

8. Key resources

Core readings (the evidence base)

- Röseler, L., Wallrich, L., et al. (2025). Handbook for Reproduction and Replication Studies (v1.0). Framework for Open and Reproducible Research Training (FORRT). <https://doi.org/10.5281/zenodo.16990115>
- Brandt, M. J., Iljerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Bonett, D. G. (2021). Design and analysis of replication studies. *Organizational Research Methods*, 24(3), 513–529. <https://doi.org/10.1177/1094428120911088>
- Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, 21(1), 1–12. <https://doi.org/10.1037/met0000051>
- Irvine, E. (2021). The role of replication studies in theory building. *Perspectives on Psychological Science*, 16(4), 844–853. <https://doi.org/10.1177/1745691620970558>
- Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist*, 77(4), 576–588. <https://doi.org/10.1037/amp0001006>
- Isager, P. M., van Aert, R. C. M., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., Krueger, J. I., Perugini, M., Ropovik, I., van 't Veer, A. E., Vranka, M., & Lakens, D. (2023). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods*, 28(2), 438–451. <https://doi.org/10.1037/met0000438>
- Obenauer, W. G. (2024). Designing, executing, and publishing replication research. *Journal of Management Scientific Reports*. <https://doi.org/10.1177/27550311241232661>
- Schwab, A., Aguinis, H., Bamberger, P., Hodgkinson, G. P., Shapiro, D. L., Starbuck, W. H., & Tsui, A. S. (2023). How replication studies can improve doctoral student education. *Journal of Management Scientific Reports*, 1(1), 18–41. <https://doi.org/10.1177/27550311231156880>

Further reading

Selected from the ARIM (Alliance for Research on Replication in Management) “How-to” list. These are not in the coach’s evidence base, but they are recommended reading.

- Simons (2014). The value of direct replication. <https://doi.org/10.1177/1745691613514755>
- Nosek, Ebersole, DeHaven & Mellor (2018). The preregistration revolution. <https://doi.org/10.1073/pnas.1708274114>
- Köhler & Cortina (2023). Constructive replication, reproducibility, and generalizability. <https://doi.org/10.1177/27550311231176016>
- Bettis, Helfat & Shaver (2016). The necessity, logic, and forms of replication. <https://doi.org/10.1002/smj.2580>
- Block & Kuckertz (2018). Seven principles of effective replication studies. <https://doi.org/10.1007/s11301-018-0149-3>
- Aguinis & Solarino (2019). Transparency and replicability in qualitative research. <https://doi.org/10.1002/smj.3015>
- Briker & Gerpott (2023). Publishing Registered Reports in management. <https://doi.org/10.1177/10944281231210309>

- Brodeur et al. (2023). Making reproducibility research more systematic. <https://doi.org/10.1038/d41586-023-02997-5>
- Hedges & Schauer (2019). More than one replication study is needed for unambiguous tests of replication. <https://doi.org/10.3102/1076998619852953>
- ARIM How-to resources (full list) <https://www.arimweb.org/resources/how-to>

Tools, databases and registries

Resource	Use it for
FORRT Replication Database (FReD)	Finding existing replications of your target (Stage 1); logging your results (Stage 10)
FORRT Replication Hub	The handbook and other replication resources
PubPeer	Post-publication discussion of your target study (Stage 3)
Open Science Framework (OSF)	Sharing materials, data and code; preregistration
AsPredicted	Short-form preregistration (Stage 8)
PCI Registered Reports	Peer review of Registered Reports (Stage 8)
G*Power	Power analysis (Stage 6)
TOSTER (R package)	Equivalence tests (Stages 6 and 9)
MBESS (R package)	Precision (AIPE) planning and effect-size CIs
BITSS ACRE guide	Computational reproducibility (Stage 4)
Institute for Replication	Reproduction and replication projects in the social sciences
APA JARS replication reporting table	What to report in a replication paper (Stage 10)
Journal of Management Scientific Reports	A management journal that publishes replications

9. Research participation and privacy

Replication Navigator is being evaluated for a methods paper. On the start screen you can tick a box to share your anonymous usage data. It is optional, and you can use every feature without it.

	If you agree	If you do not agree
What is logged	A random ID, messages to and from the coach, checklist ticks, ratings, citations you open or flag, survey answers, and your exported plan	Nothing
Where your work is saved	In your browser	In your browser
Messages to the coach	Processed by Anthropic's API to generate replies	Processed by Anthropic's API to generate replies
An uploaded paper	Sent to Anthropic's API once to extract key details; the PDF is not stored. The extracted summary is kept in your browser.	Same

- We never ask for your name or email. Your data are linked only to a random ID created by your browser.
- Do not type personal or confidential information into worksheets or the coach.
- You can change your choice at any time in Settings.

Questions or feedback: *[name, email]*.

10. Troubleshooting and FAQ

Problem	What to do
The coach takes a long time	The first reply can take up to a minute while the server wakes up. After that, text starts appearing within a few seconds.
A coach reply stops mid-way	Click “Continue” under the reply.
Uploading the paper fails	Check it is a PDF under 15 MB. Scanned papers without selectable text may not work. You can always fill in the details by hand.
“Rate limit reached”	Wait a few minutes. There is a limit on coach requests per hour for each session.
A citation is struck through	The coach cited a card that does not exist. Treat that statement with caution and flag it if you can.
My work disappeared	Check that you are on the same device and browser. If you exported, use Settings → Import plan.
I want to continue on another device	Settings → Download backup, then on the other device Settings → Import plan.
The page looks broken	Reload the page. Your work is kept in the browser.

Frequently asked questions

Is it free? Do I need an account?

Yes, it is free and open access, with no account. If the shared coach reaches its daily limit, you can add your own Anthropic API key in Settings.

Does the coach write my preregistration for me?

No. It helps you think, reviews your drafts and can draft sections on request, but you own every decision. Always check its suggestions.

Why does the coach sometimes say “the evidence base doesn’t cover this”?

Because it is only allowed to give methodological advice from its 322 evidence cards. That is a feature, not a bug. Ask a supervisor or methods expert.

Is this only for quantitative studies?

The current evidence base is mostly about quantitative replication. Qualitative replication is on the list for future versions (see further reading in section 8).

How to cite

Replication Navigator is open source. Cite it as: Shea, K. (2026). Replication Navigator: an evidence-grounded guide and AI coach for planning replication studies (v0.1). <https://github.com/kristinaraevignacyber/replication-navigator>