Piloting Practices Across Psychological Sub-Disciplines

by the Pilot Reporting Task Force[+]

September 2024
Technical Report

[+]List of authors and affiliations in alphabetical order
Agata Bochynska, University of Oslo, Norway
Caroline Zygar-Hoffmann, LMU Munich and Charlotte Fresenius Hochschule, Germany
Hannah Lönneker, University of Tübingen, Germany
Isaac Handley-Miner, Boston College, USA
Jade Pickering, Gorilla Experiment Builder, UK
Jamie Adams, Gorilla Experiment Builder, UK
Johanna Tomczak, Gorilla Experiment Builder, UK
Kai Li Chung, University of Nottingham Malaysia, Malaysia
Mary Beth Neff, University of Oslo, Norway

Data Availability

All survey materials and quantitative data are publicly accessible on OSF (https://osf.io/zg2n5/). Qualitative data are under restricted access due to re-identification risks; please contact the authors about access.

Funding

This survey was supported by the Berkeley Initiative for Transparency in the Social Sciences (BITSS) Fund, managed by the Center for Effective Global Action (CEGA), through a Catalyst grant awarded to IHM.

Conflict of Interest

All authors declare that they have no conflicts of interest.

Contact

Website: https://pilotreportingtf.github.io/
Email: pilotreportingtf@gmail.com

Reference

Pilot Reporting Task Force. (2024). *Piloting practices across psychological sub-disciplines*. OSF. https://doi.org/10.17605/OSF.IO/3QDY2

Disclaimer

This technical report presents a summary overview of the data collected. It has not gone through formal peer review. While we have taken steps to ensure accuracy and rigor, we encourage readers to view these findings as part of an ongoing discussion rather than as definitive conclusions.

CONTENTS

EXECUTIVE SUMMARY

Piloting is the pre-testing of a method ahead of planned data collection and plays a vital role in psychological research. However, little guidance exists on how to effectively design, conduct, and report pilot studies. We surveyed 135 psychology researchers about their piloting practices, including decision-making processes and reporting behaviors.

Most of the researchers who participated in the survey pilot the majority of their studies and decide upfront whether a study qualifies as a pilot. Despite this consensus, the survey results generally suggest that psychology researchers take diverse approaches to planning, running, and reporting pilot studies. While a large majority believe that specific details—such as sample size, procedures, and differences from the final study—should be reported, participants did not consistently include this information in their publications.

Overall, this survey highlights the diversity of piloting practices and, more importantly, their influence on the research process. While many researchers do not consistently report their pilot studies, they agree on the importance of including basic pilot study information. Based on these findings, we will develop pilot reporting template(s) to help researchers share their piloting practices more effectively. Additionally, we plan to conduct a meta-assessment of current reporting standards and promote pilot study transparency through an upcoming perspective article. For up-to-date information on these projects, see https://pilotreportingtf.github.io/projects/.

BACKGROUND

Piloting—conducting preliminary studies to refine research designs, timelines, resources, procedures, or instruments—is common in psychological science, yet practices vary widely both within and across subfields. Some researchers publish pilot studies on interventions (e.g., Ersser et al., 2012), experimental methods (e.g., Buie & Croft, 2023; Hamburger & Knauff, 2019), or psychometric scale development (e.g., Chen et al., 2022; Sarkeshikian et al., 2018). Meanwhile, others highlight the risks of relying on pilot data, such as biasing sample size estimates based on pilot data (Albers & Lakens, 2018).

Despite its influence on research outcomes, pilot study information is rarely reported. There is currently no clear empirical basis to estimate how often piloting efforts go unreported or how this affects subsequent research procedures. This lack of transparency and comparability may stem from inconsistent journal policies, with some requiring statements on piloting (e.g., Collabra Registered Reports, PCI Registered Reports) while most do not.

Given that there is no common approach to piloting, we first sought to understand the different ways psychological researchers conduct pilot studies (or any preliminary study). We surveyed researchers about their piloting practices, attitudes toward reporting pilots, and perceived barriers to doing so. The findings from this survey will guide two subsequent projects aimed at facilitating the reporting of pilot studies: a meta-assessment of current piloting reporting practices and template(s) for pilot study reporting.

RESEARCH OBJECTIVES

The primary objective was to understand the diversity of piloting practices in psychology. To achieve this, we surveyed psychological researchers on their definitions, implementations, and reporting of pilot studies.

We aimed to estimate the prevalence of piloting and its reporting, and to gather insights into common practices, such as the number of piloting phases before the main study and the average sample sizes used in pilot studies. Additionally, we sought to understand the rationale behind both conducting and not conducting pilot studies, as well as the reasons for reporting or not reporting these practices/findings.

METHOD

Participants were taken through an informed consent process then asked open and closed questions about their piloting practices. The survey took approximately 15–20 minutes to complete. Specific questions are outlined in more detail in the results overview below. See our OSF repository for the full questionnaire and datasets (https://osf.io/zg2n5/).

- **Pilot Study**
  In order to assess the wording of questions and to check the questionnaire logic, the study was piloted with a sample of 32 participants (see the OSF repository for previous versions). In addition to the actual questionnaire about piloting, these participants also gave feedback about the clarity and length of the survey, which was considered or implemented for the final version.

- **Ethics**
  The study was hosted on Gorilla Experiment Builder (gorilla.sc), and as such is fully compliant with GDPR and the British Psychological Society guidelines, as identifying data, demographic information and performance data are all stored separately. Data were anonymous[1] and any identifying information (e.g., emails for prize distribution) was stored separately from responses. Participants were able to withdraw at any moment and had access to the researchers' contact details at the beginning and end of the study. Analysis of the de-identified data was approved by the Boston College Institutional Review Board (Protocol #25.038.01e).

- **Participants**
  The study link was shared online via social media platforms (X, Bluesky, and Mastodon) and emailed to colleagues and faculty/society listservs. Participants who consented were entered for a prize draw to win one of 20 x $50 USD gift vouchers.

[1] Note that quantitative data could be fully anonymized (e.g., as mentioned here by storing email addresses separately from other responses) while qualitative responses might sometimes carry a risk of potential re-identification from content. Thus, only quantitative data from the current survey are openly available without sharing restrictions.

RESULTS OVERVIEW

We recruited 319 participants (see Appendix for demographic information). Prior to conducting any analyses, we excluded n = 184 due to suspected bot responses. While we did not have a pre-specified pattern to identify bot behavior, we used responses to the open-ended questions to distinguish bot vs. genuine data (see Simone et al., 2023 for more discussion on bot detection strategies). Examples of suspected bot responses include irrelevant comments in the survey feedback question, such as descriptions of maintaining/piloting airplanes or vague phrases like “Psychological Academic Practice.”

Both the suspected bot (n = 184) and actual participant (n = 135) responses have been uploaded to the OSF repository. Closed questions were analyzed descriptively, while open questions were coded using a bottom-up approach, where themes emerged from the data rather than being predefined. In the following sections, we present the results from the post-exclusion participant data, organized by the following themes:

- **Prevalence And Type Of Piloting**: This section provides descriptive data on the types of piloting used by psychological researchers, including broad categories of the types of approaches used, sample sizes, and qualitative descriptions of their piloting practices.

- **Decision-Making Process Behind Piloting:** This section details the decision-making processes involved in researchers’ piloting practices, such as how they classify a study as a pilot, how they evaluate pilot studies, and what occurs when a study does not progress beyond the piloting stage.

- **Pilot Reporting (Current And Future Possibilities):** This section outlines how participants currently report or share their pilot study findings. It also includes qualitative data on what researchers value in pilot studies and any challenges they anticipate with reporting these findings in the future.

PREVALENCE AND TYPE OF PILOTING

We began the survey by providing a broad definition of piloting and then asked participants to provide any synonyms they might have in their field for preliminary testing.

> This survey is about piloting. Piloting is broadly the practice of conducting preliminary studies or trial runs to refine research designs, procedures, or instruments.
>
> Conventions vary across different fields and subdisciplines. To capture all the ways people use or refer to these preliminary studies (e.g. pilot studies) please list all terms or keywords (i.e. possible search terms) used in your field for any studies/checks conducted in advance of final data collection

We also asked participants to briefly describe what they believe differentiates a preliminary study from a final study.

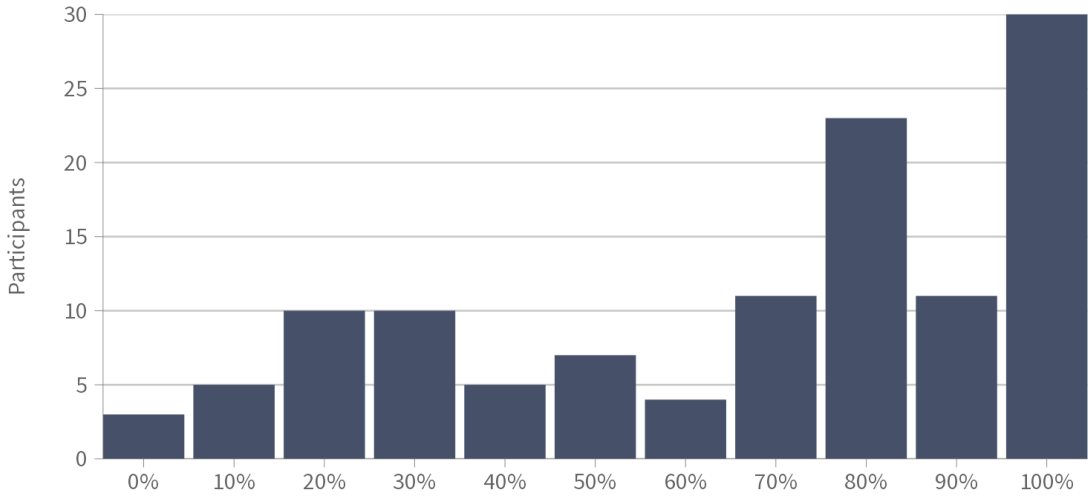
For terms and keywords (n = 259), the most commonly supplied term was **pilot(ing)**, appearing 92 times, though participants varied as to whether they referred to it as a pilot 'study', 'test', or 'trial'. The other more frequent terms were, **preliminary study** (n = 38), **pre-test** (n = 12), and **exploratory study** (n = 10). A word cloud of the terms is presented below.

When asked whether they run pilot studies for their own research, 119 said 'Yes' while 16 said 'No'. We then asked participants what characteristics distinguish a pilot study from the final study. Below are the categories from the open data:

- **Sample size:** smaller (n = 48); depends on method (n = 1)
- **Sample characteristics:** convenient, accessible (n = 8); not naïve (n = 5)
- **Stimulus/ trial number:** smaller (n = 6); larger (n = 3); depends (n = 2)
- **Duration:** shorter (n = 2); longer (n = 1); depends (n = 1)
- **Purpose:** for testing/verifying/checking design / instruments / task instructions (n = 77); proof of concept (n = 1); determine viability (n = 16); testing research question / hypotheses (n = 13); to gain feedback (n = 20)
- **Data not included / analyzed in the final sample** (n = 25)
- **Modifiable** (n = 24)
- **Exploratory / incomplete / not finalized** (n = 14)
- **Not pre-registered** (n = 6)
- **Not categorized** (n = 3)

Of the participants who used piloting in their research (n = 119), we asked follow-up questions about how often they run these preliminary studies and what variety of piloting procedures they use. While the responses varied, n = 30 participants reported they pilot 100% of their studies (Plot 1).

Plot 1. **"Think about the non-pilot studies you have run, what percentage of those studies did you pilot first?"**



We then asked how often they engage in four categories of piloting (from Never to Always). Below are the categories and definitions we provided participants as well as their response ratings.

- **Assessing a method's effect:** e.g., effect size estimation, manipulation check
- **Developing a method:** e.g., creating new experimental methodologies or procedures, refining survey questions
- **Evaluating a method's feasibility:** e.g., testing the practicality of data collection procedures, assessing participant comprehension
- **Technical Pilot Studies:** e.g., for checking programming, technical, or administrative protocol issues

Plot 2. **"Please rate how often you engage in each type of piloting for your research:"**

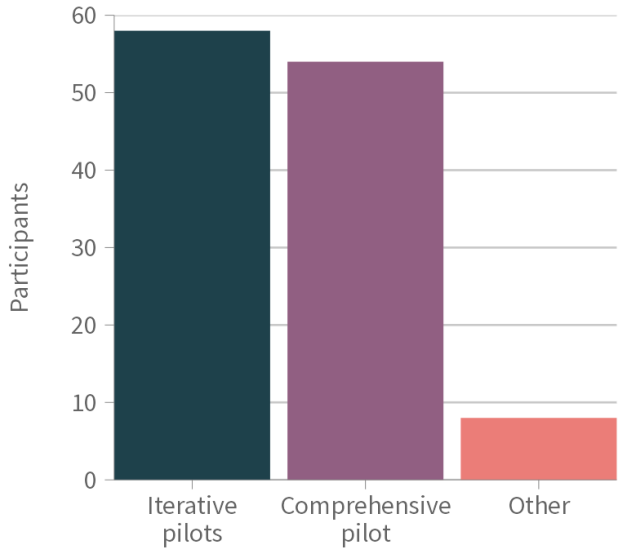
Although all types of piloting were relatively common, piloting to "assess a method's effect" had fewer 'always' responses (Plot 2).

After these four categories, we also asked whether participants ever used piloting to select which particular 'stimuli, manipulation, or measure' to use in the final study and whether there were any 'other' types of piloting we may have missed.

Seventy-seven percent of participants reported using piloting to select study measurements. Although few 'other' types of piloting were mentioned, some responses included assessing 'adaptability,' such as determining whether a method could be adapted for a specific age range or language. Others cited using piloting to refine recruitment strategies. Additionally, a few participants used this question to clarify what they believe piloting does not encompass, rather than suggesting alternative practices.

After addressing the more specific questions on piloting practices, we invited participants to take a broader perspective on piloting as a whole for the remaining questions.

Plot 3. **"How would you describe your piloting process before running the final version of a study?"**

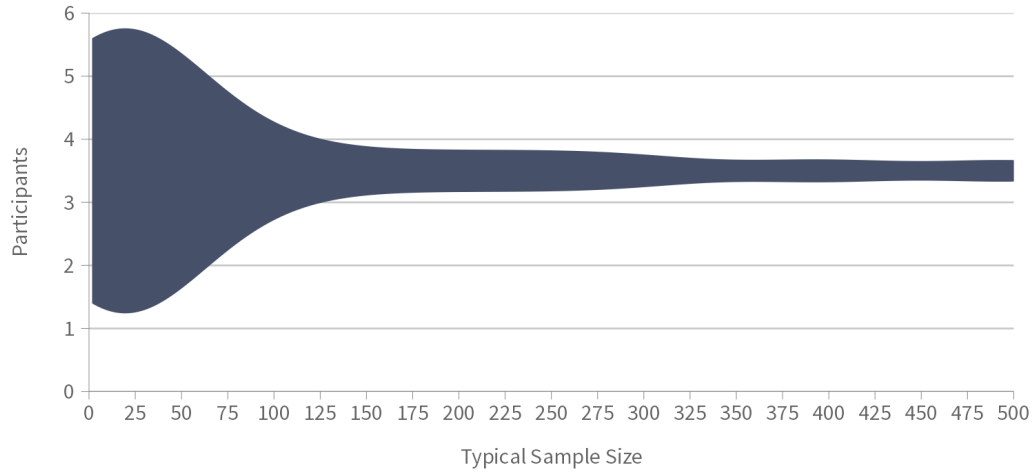


We asked participants whether they typically pilot using a single comprehensive study or an iterative approach. Both types of piloting were common, and very few 'other' processes were mentioned (Plot 3). Those who reported iterative piloting practices said they tend to run 2-3 pilots before the main study.

We also asked for an approximate sample size of a typical pilot study. Note for this question, we instructed participants to disregard 'technical piloting' from their responses.

Participants' pilot sample sizes varied, with most respondents clustering around sample sizes of fewer than 25 and up to 75, though a few reported using as many as 500 subjects (Plot 4).

Plot 4. **"Excluding technical pilots, what would you estimate is the typical sample size for your pilot studies (if it is a range, please put in the average)?"**



For these ranges, it is important to note the variety of subfields reflected in this research (see Appendix for these demographics). For further context we also asked participants about the average sample size for final studies (M = 234, range: 2–5000) as well as their general approach to determining the sample size for their pilot studies.

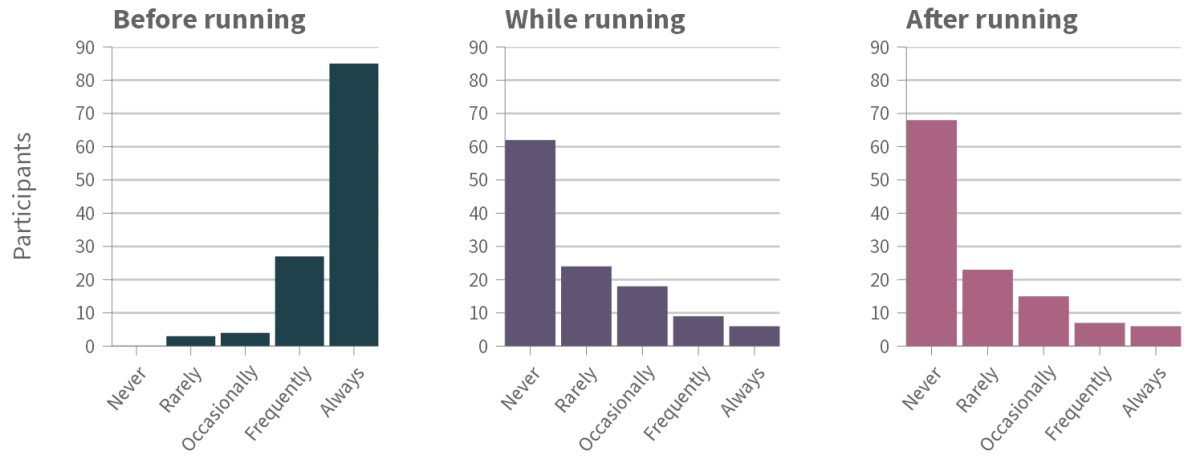
Below are the categories from the open responses (from 119 participants) to the question: "**How is this (pilot) sample size typically determined?"**

- Convenience / practicality (n = 40)
- Ballpark figure / estimation / rule of thumb (n = 25)
- Available resources / feasibility (n = 28)
- Until saturation (n = 16)
- Previous / usual practice (n = 14)
- Power calculation (n = 12)
- Not categorized (n = 6)

DECISION-MAKING PROCESS BEHIND PILOTING

We first asked participants at what point they decide a study is a pilot study. We then asked how they decided about these piloting phases, including how they decided when to stop a pilot study and progress to the main data collection.

Plot 5. **"How often do you decide that a study is a pilot study…"**



In general, participants reported that most pilots were considered a pilot before data collection began (Plot 5).

Below are the categories from the open responses (from 116 participants) to the question: **How do you determine whether a study will need to undergo a piloting phase?**

- New stimuli / materials / method (n = 49)
- Always pilot (n = 42)
- Check suitability (n = 8)
- New sample (n = 7)
- Cost of study (n = 6)
- Previous findings (n = 6)
- Complex design (n = 6)
- Obtain estimation of effect (n = 4)
- Intuition / experience (n = 6)
- New theory (n = 2)
- Check validity (n = 2)
- Not categorized (n = 6)

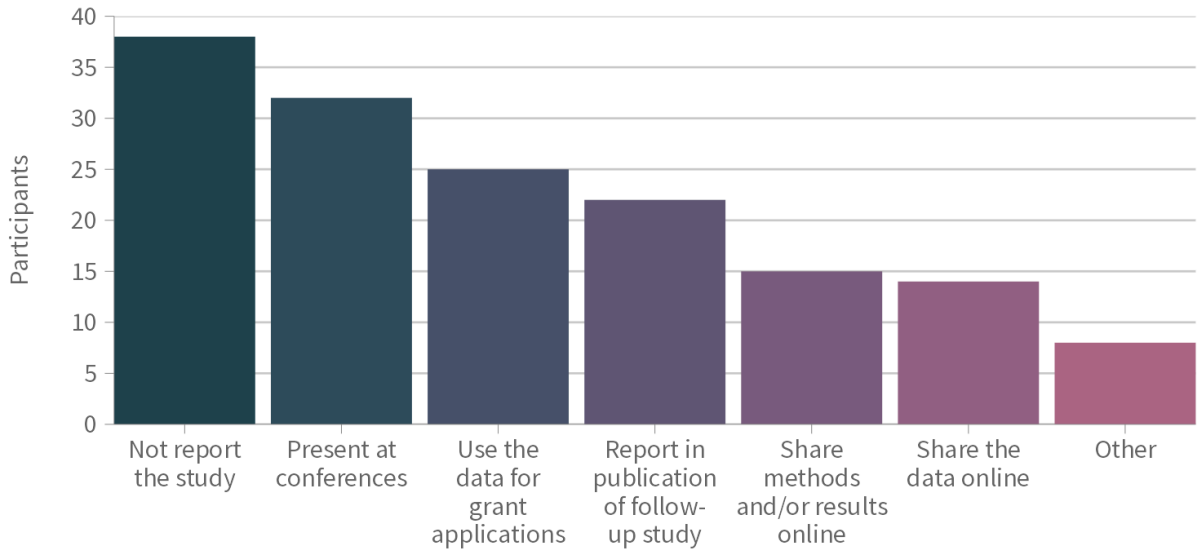
When participants were asked **"How do you decide when to stop piloting a study and run the final version?"** common themes were 'method working' and 'findings as expected'. See below for all themes from the open response data (from 113 participants)

- Method working (n = 46)
- Findings as expected (n = 26)
- Predetermined sample size (n = 16)
- Issues are resolved (n = 15)
- Participants' understand (n = 15)
- Resources (n = 12)
- Data quality (n = 8)
- Just one pilot (n = 6)
- Not categorized (n = 3)

Lastly, when asked whether they had a study that did not make it beyond the piloting stage about half of the participants said that they had (n = 58) and half had not (n = 61).

For those who reported that they *had* abandoned studies during the piloting phase, we asked a follow-up question about what they usually do with those data. Most reported that they do not do anything with the study while some also present the data at a conference or use it in a grant application (Plot 6).

Plot 6. **"What do you usually do with these abandoned studies? (Select all that apply)"**



Note. Participants could check multiple answers for this question

PILOT REPORTING (CURRENT AND FUTURE POSSIBILITIES)

We first asked participants what information, if any, they include about pilot studies in their publications. There were a variety of responses. However, of those that had published, most responded that they did not report pilot study information or only included it in some of their publications (Plot 7). For those who indicated they included pilot study details, we asked them to elaborate on what they typically report; see the Appendix for these data.

We then asked participants what information they think *should* be published. Participants were given a drop-down list of responses, and the majority of participants selected at least one aspect of the piloting process. Very few responded that no information should be reported (Table 1).

Plot 7. **"How often do you include information about pilot studies in your publications (including any supplementary information)?"**

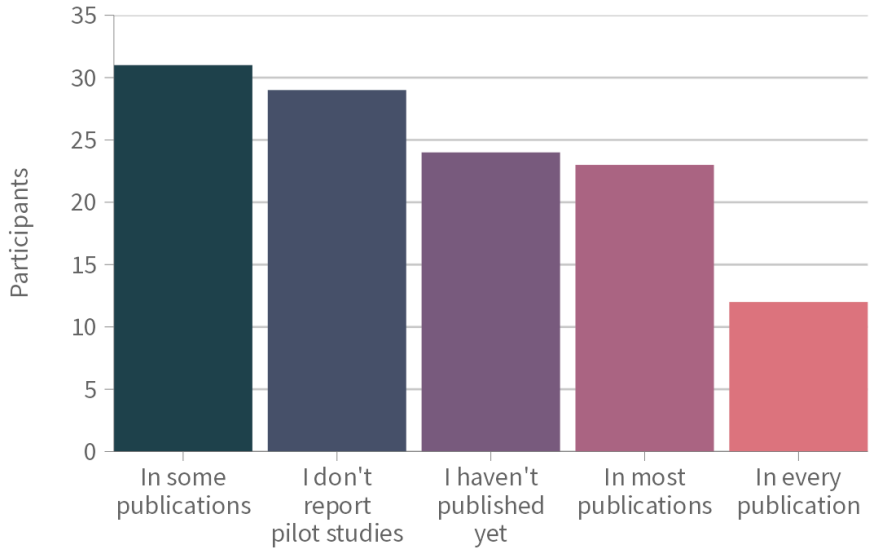


Table 1. **"What (if any) information about pilot studies do you think would be helpful to include in publications? (Select all that apply)"**

Information type	*n*
The fact that a pilot study was run	104
Sample size of the pilot study	88
Rationale for differences between the pilot study and the final study	84
Description of differences between the pilot study and the final study	82
Procedure of the pilot study	82
Conclusions from the pilot study for the final study	63
Quantitative results of the pilot study	52
Qualitative results of the pilot study	47
Other (miscellaneous responses)	15
I do not think any information about pilot studies should be reported	6

For those that selected any information that should be reported (n = 119), we asked a follow-up open question: **Why do you think this information about pilot studies should be reported in publications?**

- Better assessment and understanding of final publications (n = 55)
- Transparency / ethics (n = 36)
- Knowledge advancement / efficiency of research (n = 23)
- Replication (n = 11)
- Uncovering publication bias (n = 5)
- Training opportunity / giving a realistic picture of what research entails (n = 5)
- Accountability (n = 2)
- Not categorized (n = 15)

Even though participants were asked why they should report, there were some responses speaking to why some should *not* (n = 6), which we combined with the question below.

For those that responded that **they do not think that information about pilot studies should be reported** we asked a follow-up question about why that is. Response themes for these reasons are below (n = 15 total responses):

- No obvious reasons (n = 5)
- Not interesting, if negligible conclusions drawn or only technical pilot (n = 2)
- Not relevant, only final version counts (n = 2)
- Word limits / journal requirements (n = 2)
- Low trustworthiness of pilot data (n = 4)
- Negative reactions (n = 2)
- Capacity / time (n = 1)
- Not categorized (n = 3)

In the remaining questions, we asked participants if they would ever combine pilot data with final data, how they would perceive an article if the author reported making changes to the original study based on pilot findings, and what challenges they believe researchers might encounter when reporting pilot study information.

When asked whether they ever combine pilot data with final study data in their reporting, over half (n = 64) said they would *not*. However, responses were split between those who said yes (n = 25) and those who answered 'other' (with other = "it depends", n = 30).

In terms of their perceptions, most participants were either neutral (n = 53) or indicated they would view the article more favorably (n = 52). However, a number of participants provided 'other' responses (n = 28), and only two participants said they would view the article less favorably.

'Other' responses varied but often revolved around how transparently the author(s) reported any changes. Within these other responses, participants also tended to endorse changes that impacted the overall quality of the procedure rather than those aimed at enhancing a specific effect/outcome.

Some of these 'other' themes also emerged in our final question about potential barriers to reporting pilot studies, see below for the complete themes:

What challenges could researchers face when including information about pilot studies in publications? (n = 116)

- Negative reactions (e.g., about how the pilot was conducted, or how robust the final version is) (n = 52)
- Word limits (n = 32)
- Finding the right amount of detail (e.g., to not overwhelm or confuse the reader, or extend the length of the manuscript unnecessarily) (n = 23)
- I don't know / none (n = 15)
- Lack of examples / standardization / guidelines (also on what constitutes a pilot) (n = 8)
- Poor pilot documentation (n = 6)
- Takes time / effort / complicates things (n = 6)
- Funder / journal requirements (other than word count) (n = 2)
- Ethics / IRB (n = 1)
- Not categorized (n = 1)

SUMMARY

135 psychology researchers (after exclusions) completed the survey on piloting practices, 119 of whom have run pilot studies in the past. These 119 researchers answered questions related to three broad topics:

1. the prevalence and type of piloting they carry out in their research
2. their decision-making process behind piloting
3. their practices and attitudes towards reporting pilot studies

Most questions received a wide range of responses. However, the questions with more consistent responses suggested that:

- Most researchers pilot a majority of their studies
- Most researchers determine whether a study is a pilot before they run it
- Most researchers think that some information about pilot studies should be reported in papers (especially: the fact that a pilot study was run, the sample size of the pilot study, the procedure of the pilot study, and a description and rationale for differences between the pilot study and the final study), but few consistently include this information in their own reporting

Some questions that received particularly variable responses included:

- Whether researchers typically pilot using a single comprehensive study or an iterative approach
- The average sample size researchers use in pilots, and how researchers determine this sample size
- Whether researchers have studies that did not make it beyond the piloting stage, and what researchers do with such data
- Whether researchers would ever combine pilot data and final study data
- How often researchers report information about their pilot studies in their publications

In general, the survey results suggest that, although the practice of piloting is common, psychology researchers take diverse approaches to planning, running, and reporting pilot studies.

LIMITATIONS

One of the limitations of the survey is the participating sample. Psychology researchers were recruited to take a "survey about pilot studies" via social media (X, Bluesky), professional societies, and academic department listservs. This could have caused the sample to be unrepresentative of all psychology researchers for two main reasons: 1) Researchers who do not run pilot studies or are uninterested in the process of piloting may have been less likely to take the survey, and researchers who are more interested in methodology may have been oversampled; 2) Sharing this study via social media and society and listservs meant that people in the existing social media, departmental, or societal networks of the authors were more likely to see the advertisement for the survey. Although the authors come from a variety of countries, career stages, and subdisciplines within psychology, it is unlikely that researchers were recruited equally across geographic locations, career stages, and subdisciplines (see Appendix for demographic data of participants).

CONCLUSIONS

Methodological choices informed by piloting can impact research outcomes. Indeed, our results indicate that researchers frequently use pilots for developing a method, evaluating a method's feasibility, assessing a method's effect, and checking for technical issues. Our results also suggest that piloting is a highly variable practice, with researchers employing many different methods for planning, running, and reporting pilots. Yet, most researchers indicated that they do not always report their pilot studies in publications. At the same time, most researchers in our survey seemed to agree that at least basic information about pilots should be reported in publications (e.g., sample size, procedure, differences from the final study). However, there is little guidance for reporting pilots (e.g., the APA Publication Manual is silent on the topic). This calls for templates, tools, or guidelines to assist researchers with reporting details from pilot studies. Collectively, these results point to a need for more transparency in piloting practices.

REFERENCES

1. Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

2. Buie, H., & Croft, A. (2023). The Social Media Sexist Content (SMSC) Database: A Database of Content and Comments for Research Use. *Collabra: Psychology*, *9*(1), 71341. https://doi.org/10.1525/collabra.71341

3. Chen, M., Nah, Y.-H., Waschl, N., Poon, K., & Chen, P. (2022). Developing and Piloting a Computerized Adaptive Test for a Culturally Appropriate Measure of Adaptive Behavior. *Journal of Psychoeducational Assessment*, *40*(2), 238–254. https://doi.org/10.1177/07342829211047005

4. Ersser, S. j., Cowdell, F. c., Nicholls, P. g., Latter, S. m., & Healy, E. (2012). A pilot randomized controlled trial to examine the feasibility and efficacy of an educational nursing intervention to improve self-management practices in patients with mild-moderate psoriasis. *Journal of the European Academy of Dermatology and Venereology*, *26*(6), 738–745. https://doi.org/10.1111/j.1468-3083.2011.04158.x

5. Hamburger, K., & Knauff, M. (2019). Odors Can Serve as Landmarks in Human Wayfinding. *Cognitive Science*, *43*(11), e12798. https://doi.org/10.1111/cogs.12798

6. Sarkeshikian, A. H., Tabatabaee, A.-M., & Doaee, M. T. (2018). Unidimensionality and construct validity of the Self-Regulating Capacity in Vocabulary Learning (SRCvoc) in Iranian EFL context: Item-level responses versus item parcels. *Psychology of Language and Communication*, *22*(1), 21–38. https://doi.org/10.2478/plc-2018-0002

7. Simone, M., Cascalheira, C. J., & Pierce, B. G. (2023). A quasi-experimental study examining the efficacy of multimodal bot screening tools and recommendations to preserve data integrity in online psychological research. *American Psychologist*. Advance online publication. https://doi.org/10.1037/amp0001183

APPENDIX

PSYCHOLOGY SUBFIELDS

Psych Subfield	*n*	*proportion*
Applied Behaviour Analysis	11	0.08
Biological Psychology	19	0.13
Child Psychology	21	0.15
Clinical Psychology	11	0.08
Cognition and Perception	31	0.22
Cognitive Psychology	64	0.45
Community Psychology	7	0.05
Counseling Psychology	11	0.08
Developmental Psychology	58	0.41
Experimental Analysis of Behaviour	16	0.11
Geopsychology	3	0.02
Health Psychology	6	0.04
Human Factors Psychology	3	0.02
Industrial and Organizational Psychology	5	0.04
Multicultural Psychology	8	0.06
Pain Management	4	0.03
Personality and Social Contexts	16	0.11
Quantitative Psychology	18	0.13
Social Psychology	27	0.19
Theory and Philosophy	3	0.02
Transpersonal Psychology	1	0.01

"Other" Subfield	*n*
Psycholinguistics	6
Educational Psychology	8
Comparative Cognition	1
Consumer Psychology	1
Environmental Psychology	1
Forensic Psychology	1
Metascience	2
Indigenous Psychology	1
Differential Psychology	2
Neuroscience	2
Social Science Broadly	3

Note. Participants could select more than one option

RESEARCH PRACTICES

Research Practice	*n*
Quantitative mostly	101
Qualitative mostly	8
Mixed methods	25
NA (I'm not currently practicing research)	1

GEOGRAPHIC LOCATIONS

Location	*n*	*proportion*
Asia	9	0.06
Africa	4	0.03
North America	67	0.47
South America	8	0.06
Europe	52	0.37
Oceania	8	0.06
Antarctica	0	0.00

JOB TITLES

Job Title	*n*
Industry / Research scientist	5
Lab Manager / Research Technician	8
Lecturer / Professor (including assistant, associate, senior, etc.)	41
Postdoc / Researcher	33
PhD Student	37
Master's Student	11
Bachelor's / Undergraduate Student	5

CURRENT PILOT STUDY REPORTING

If a participant indicated that they reported any pilot study information, they were asked to elaborate how in a subsequent open-text question. Responses to this question (n = 66) were categorized under the broader themes of “Length” and “Content.” Below, we present those themes and their sub-themes:

Length / Format of reporting:

- Publication as short report / case report / conference report (n = 3)
- Detailed report in final publications (n = 10)
- Brief report / mention or footnote in final publications (n = 35)
- Length not specified, but report in main text of final publications (n = 20)
- In supplemental materials (n = 12)
- Publication of materials and data of the pilot study (n = 1)
- (Sometimes) no report of the pilot (n = 4)
- ‘Varies’ (n = 28)

Content of report:

- Purpose / target of piloting (n = 18)
- Procedure (n = 7)
- Sample size (n = 12)
- Sample characteristics (n = 2)
- Results and discussion of consequences/conclusions (n = 11)
- Results without discussion (n = 3)
- Only discussion of consequences of pilot results (n = 10)
- Full reporting (typically includes all of the points above) (n = 12)

Content not specified (n = 21)