Meta-Psychology, 2024, vol 8, MP.2023.4039
https://doi.org/10.15626/MP.2023.4039
Article type: Original Article

Open data: Yes
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: Yes

Edited by: Rickard Carlsson
Reviewed by: van den Akker, O., Topor, M., Rahal, R.
Analysis reproduced by: Lucija Batinović
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/3QARW

Testing the Usability of the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template

Lisa Spitzer[1], Michael Bosnjak[2], and Stefanie Mueller[1]
[1]Leibniz Institute for Psychology (ZPID)
[2]Trier University

The Psychological Research Preregistration-Quantitative (PRP-QUANT) Template provides researchers with a comprehensive list of elements to consider when planning a psychological study. We assessed its usability and researchers' intention to use it. We conducted a usability test (study 1) and surveyed researchers who submitted or reviewed a preregistration created with the template (study 2, authors: $N = 19$, reviewers: $N = 29$) regarding their impression of the template. For the usability test, we recruited participants via the mailing lists of the German Psychological Society, the American Psychological Association, and the British Psychological Society, and social media. Participants answered selected template and web probing items and provided an overall rating ($N = 88$). Based on the Unified Theory of Acceptance and Use of Technology (UTAUT), we expected that the intention to use the template is influenced by performance expectancy (moderated by age), effort expectancy (moderated by age and experience), and social influence (moderated by age, experience, and voluntariness, $N = 60$). The results suggest that the PRP-QUANT Template is suitable for different research areas within psychology, is evaluated as effective, and perceived positively. Performance expectancy and all predictors combined significantly predicted researchers' intention to use the template. A plain language summary in English and German is available in Text S1 in the supplemental material.

Keywords: meta-research, open science, preregistration, reproducibility, replicability

Introduction

In recent years, there has been a growing call for methods and procedures to increase the transparency of research (e.g., see Munafò et al., 2017), one of them being study preregistration (Nosek et al., 2018), in which researchers document and publish their study plan before data have been collected or examined. Preregistrations are time-stamped and published with an independent party (e.g., a repository) so that they can be accessed by others (possibly after an embargo period). This way, preregistration aims to provide transparent documentation of study procedures, clear identification of deviations from preregistered plans, and a clear distinction between confirmatory and exploratory research (Parsons et al., 2022).

Studies suggest that preregistration can help reduce questionable research practices and the rate of false positive findings (Kaplan and Irvin, 2015; Swaen et al., 2001) and, among other open science techniques, might increase the replication rate of research (Protzko et al., 2020). However, despite its benefits, there is also evidence that questions the effectiveness of current preregistration (e.g., as much flexibility remains open and deviations are often not sufficiently disclosed, see Van Den Akker, Bakker, et al., 2023). A recent study also showed that uncertainty about which aspects needed to be included in the preregistration was both a concern of researchers who had not yet preregistered, and a problem experienced by researchers with preregistration experience. Accordingly, better education about preregistration was one of the most common suggestions to increase motivation and reduce obstacles of preregistration (Spitzer and Mueller, 2023).

Preregistration templates can help overcome uncertainty by listing important elements that researchers should address in their preregistration (e.g., hypotheses, study design, data acquisition, and data analysis plan). Nowadays, a variety of templates are available, differing in scope and targeted research type. Besides more universal templates, there are templates specifically focusing on social psychology (van 't Veer and Giner-Sorolla, 2016), fMRI studies (Beyer et al., 2021), replication studies (Brandt et al., 2014), cognitive models (Crüwell and Evans, 2021), or secondary data analyses (Van den Akker et al., 2021). This wide range of options may lead to fragmentation and potential con-

fusion among researchers as to which template should be used. Selecting the right template is no trivial decision and can be particularly challenging for preregistration beginners. A common standard, such as a universal template that covers most of the research areas in psychology, provides an easy starting point and facilitates comparability between preregistrations.

To develop such a standard for psychology and reinforce the importance of preregistration, members of the American Psychological Association (APA), the British Psychological Society (BPS), the German Psychological Society (DGPs), the Center for Open Science (COS, https://www.cos.io/), and the Leibniz Institute for Psychology (ZPID, https://leibniz-psychology.org/en/) formed a Joint Psychological Societies Preregistration Task Force. Together, they developed the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template (Bosnjak et al., 2022), a comprehensive template that aids the preregistration of quantitative studies in psychology.

Testing the Usability of the PRP-QUANT Template

Since preregistration of research in psychology is mainly voluntary, it is essential to provide a good usability to enhance acceptance. There is currently little empirical evaluation of preregistration templates (for an example, see Bakker et al., 2020; Heirene et al., 2021; Van Den Akker, Van Assen, et al., 2023), however, it is reasonable to not simply assume usability but to test it empirically. Thus, in line with the PRP-QUANT Template's goal of becoming increasingly adapted to the needs of the psychological research community (see Bosnjak et al., 2022), the first aim of our studies was to evaluate its usability and identify areas for improvement.

Various definitions of usability exist, largely sharing the same underlying concepts, but highlighting different aspects. A popular definition comes from the International Organization for Standardization (ISO), which measures usability along the dimensions of effectiveness, efficiency, and satisfaction with regard to specific users, objectives and contexts (International Organization for Standardization, 2018). Another definition is provided by Shackel (2009), who defines usability as "the capability to be used by humans easily and effectively" (p. 340). According to Shackel (2009), four aspects of usability should be considered: learnability (i.e., being usable with an appropriate amount of training), flexibility (i.e., allowing adaptation in different tasks and environments), effectiveness (i.e., achieving a required level of performance), and attitude (i.e., satisfaction, considering human costs like tiredness or effort). These partly align with the ISO standard but place more emphasis on learnability and flexibility. As the PRP-QUANT Template is intended to cover a wide range of different psychological sub-disciplines, and therefore the issues of learnability and flexibility are particularly relevant for assessing the template's usability within all of psychology, the aspects defined by Shackel (2009) were used as the basis for this research. Since no training was carried out in our study, the understandability of the template was used as an approximation for learnability.

Specifically, we were interested in the following questions:

A) *Learnability*: Do authors from the various sub-disciplines of psychology understand how to fill in the different items of the template? Do they understand the items in the same way?

B) *Flexibility*: Does the template capture the main points across sub-disciplines, as indicated by researchers from different sub-disciplines?

C) *Effectiveness*: Are the items specific enough (i.e., are researcher degrees of freedom minimized)? Are items answered as expected (i.e., is the information requested in the item provided by researchers in response to it)?

D) *Attitude*: Are users satisfied with using the template? Are costs (e.g., tiredness, personal effort) acceptable? Can the goals of the template (i.e., a detailed mapping of the preregistered study) be achieved with a reasonable amount of effort? Would authors recommend/use the template?

To assess the usability of the PRP-QUANT Template, we conducted an online study in which we asked psychological researchers to think about one of their studies and create a preregistration for that study using the template (see *study 1: simulation trial and intention to use*). Participants did not actually submit their preregistration. Alongside the template items, several web probing questions were presented. We wanted to assess the overall perceived usability (based on the criteria outlined above), as well as participants' comments and suggestions for improving individual items.

In addition, we conducted a survey among researchers who responded to a call for online studies by submitting a preregistration created with the PRP-QUANT Template (see *study 2: survey of preregistration authors and reviewers*). Responding to this call, researchers applied with their preregistrations for funding for their data collection from ZPID's service PsychLab ONLINE. PsychLab aims to encourage preregistration by offering the incentive of free-of-charge data collection for high-quality preregistrations, which addresses another current obstacle to preregistration, i.e., insufficient incentives (Spitzer and Mueller, 2023). The sub-

mitted preregistrations were evaluated by external peer-reviewers. After the peer-reviews were completed, the authors of preregistrations (i.e., the applicants) and the reviewers were surveyed about using the PRP-QUANT Template for writing and reviewing, respectively.

Assessing the Intention to Use the PRP-QUANT Template

In addition to exploring the usability of the preregistration template, we also wanted to find out whether psychological researchers plan to use the template in the future to create their preregistrations. To investigate this, the theoretical framework of the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003, 2016) was used. This theory postulates that performance expectancy (i.e., the belief that using the system will help achieve performance gains), effort expectancy (i.e., the degree of ease associated with using the system), and social influence (i.e., the perception that important others believe one should use the new system) predict people's intention to use a new system. According to the UTAUT, intention, combined with facilitating conditions (i.e., the belief that an organizational and technical infrastructure exists which supports using the system), is a predictor of actual behavior. Therefore, it is of interest to investigate the intention to use the template in more detail, as this might be an estimator of how likely psychological researchers will use it in the future to create their preregistrations.

To examine the intention to use the PRP-QUANT Template, we asked the participants of study 1 to answer various UTAUT items (see *study 1: simulation trial and intention to use*). Then, we computed a moderated multiple regression model. Based on the UTAUT and the effects described by Venkatesh et al. (2003), we had the following predictions, which are also displayed in Figure 1:

1) Performance expectancy is a positive predictor for the intention to use the template.

2) Effort expectancy is a positive predictor for the intention to use the template.

3) Social influence is a positive predictor for the intention to use the template.

4) Age negatively moderates the effect of performance expectancy on intention, as it has been shown that extrinsic rewards may be more important for younger persons.

5) Age positively moderates the effect of effort expectancy on intention, since older persons have more difficulties in processing complex stimuli and attention allocation.

6) Age positively moderates the effect of social influence on intention, since older persons might place more importance on social influences and affiliation.

7) Experience (operationalized as the academic experience, i.e., participants' academic group) negatively moderates the effect of effort expectancy on intention, as prior experience would serve as a facilitator for using the new system.

8) Experience negatively moderates the effect of social influence on intention, since it has been shown that the salience of social influences decreases with experience.

9) Voluntariness (i.e., the extent to which researchers feel they can decide whether or not to preregister) negatively moderates the effect of social influence on intention, as social influence is less important in settings where the decision to use the system is completely voluntary.

Study 1: Simulation Trial and Intention to Use

In study 1, we examined the usability of the PRP-QUANT Template by asking psychological researchers to think of one of their studies and complete selected parts of the template. We also presented web probing questions and examined the researchers' intention to use the template in the future.

Methods

This study was preregistered (https://doi.org/10.23668/psycharchives.4636). It was conducted as preregistered, except for the deviations summarized and justified in the section *Deviations From the Preregistration*.

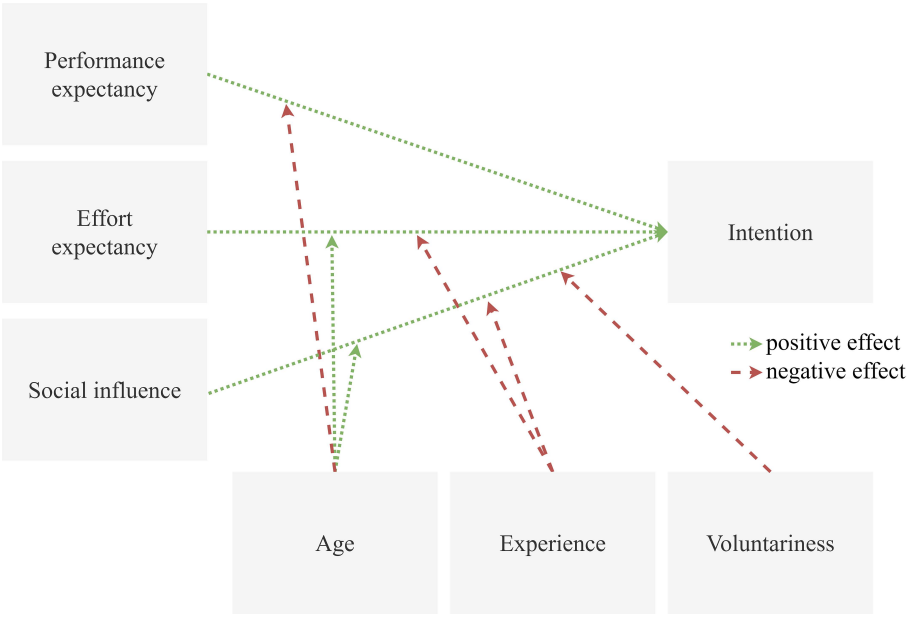
Participants

Participants were invited via the mailing lists of all research-oriented APA (https://www.apadivisions.org/), BPS (https://www.bps.org.uk/member-networks), and DGPs (https://www.dgps.de/fachgruppen) divisions. A reminder was sent a few weeks after the initial invitation. Furthermore, the survey was advertised on social media (Facebook and Twitter). Participants were not compensated.

Of the 2668 persons that clicked on the study link, 314 provided informed consent and started the main body of the study. Nine participants who indicated that they were not researchers or that their research did not fall within the scope of psychology were screened out at the beginning of the study since we specifically aimed to collect data from psychological researchers. Of the remaining participants, 88 subsequently worked on the template items and were thus included in the descriptive reports, as this was the core part of the study

Figure 1

Hypotheses based on the UTAUT (Venkatesh et al., 2003, 2016)



(34.09% male, 57.95% female, 2.27% other, 5.68% preferred not to answer; $Mean_{age}$ = 37.19; 86.36% from Europe, 11.36% from North America, 2.27% did not respond; 13.64% native English speakers). We were able to collect data from all targeted academic groups, that is, 32.95% of participants were PhD students, 37.5% were postdocs, 26.14% were professors, 2.27% indicated "other", and 1.14% did not respond. Additionally, all the considered research areas were present, with experimental/cognitive, educational, and social psychology being the most prominent (see Table S1 in the supplemental material). More than half of the participants indicated having preregistered before (56.82%). Of the participants with preregistration experience (n =50), 16% had preregistered one study, 14% had preregistered two studies, 16% three studies, 8% four studies, 12% five studies, and 34% more than five studies.

For answering the template and web probing items, participants were randomly assigned to one of four conditions. Depending on their condition, they were requested to fill out only a subsection of the template: 1) title and introduction, 2) overall methods, sampling procedure, and data collection, 3) overall methods, conditions, and design, or 4) analysis plans. This aimed to reduce the burden placed on each participant. Twenty-three participants were in condition 1, 29 in condition 2, 19 in condition 3, and 17 in condition 4 (overall: N = 88).

Of all participants included in the descriptive reports, 60 answered all items relevant for the UTAUT model and were therefore included in the hypotheses tests (36.67% male, 56.67% female, 3.33% other, 3.33% preferred not to answer; $Mean_{age}$ = 36.12; 33.33% PhD students, 41.67% postdocs, and 25% professors; 88.33% from Europe, 10% from North America, 1.67% did not respond; 13.33% native English speakers).

Data were collected between March 1, 2021, and April 24, 2021. As preregistered, data collection was stopped one month after the initial invitation was sent to the last contacted division. Originally, a sample size of N = 89 was targeted to be able to detect effects of R^2 = 25% with $\alpha = \beta = .05$ for the UTAUT regression model including 12 predictors, which was determined by an a priori power analysis (see preregistration: https://doi.org/10.23668/psycharchives.4636). This sample size was not reached within the set timeframe, but the effect for the overall model was quite large ($R^2_{adjusted}$ = 42.79%) and could thus also be detected with the achieved sample size. However, since only one of the individual predictors was significant, we conducted a sensitivity analysis to determine how informative the tests of the individual predictors were (see Figure 2). The upper bound of the curve (f^2 = 0.149) represents the effect size of our significant predictor (performance ex-

pectancy), for which a high power of 90% was achieved. However, when inspecting the power for the predictor with the second-highest beta value (social influence x age, $f^2 = 0.066$), power drops to 62.4%. To reach a power of 80%, the effect size would have needed to be at least $f^2 = 0.106$. The results for the individual predictors should therefore be interpreted with caution.

Material and Measures

The online survey was created using the software SoSci Survey (Leiner, 2019) and was supplied via www.soscisurvey.de. It was presented in English. In line with the two aims of this project, the study items focused on assessing the usability of the PRP-QUANT Template and measuring the UTAUT variables (see online materials for a complete list of items: https://doi.org/10.23668/psycharchives.12959).

As a first measure of usability, the template's effectiveness was inspected. For this purpose, participants were asked to answer the items of the PRP-QUANT Template as if they were preparing a real preregistration, thinking about a study they were currently planning or conducting (or, if no current project was available, a previous study). They first provided a brief description of their study and answered items about its status and whether they planned to preregister it. They were then asked to complete the individual template items. Effectiveness was measured by coding participants' responses to the template items in terms of their fit with what was asked in the item (see section *Data Analysis and Pre-Processing*) and by having participants rate the perceived importance of all items.

To gain a deeper insight into the participants' interaction with the template items, as well as collect participants' suggestions for improvements for all items, several web probing questions were displayed alongside the template items (derived from Behr et al., 2017), probing for category-selection (e.g., for items such as T11 "Code availability" which required selecting an option, participants were asked to elaborate why they selected the respective category), comprehension (e.g., asking participants for the meaning of terms or paraphrasing, requesting participants to rate how well they understood the item, or to differentiate template items from related items), or elaboration (e.g., asking participants for examples). Some of the web probing items were displayed for all template items (i.e., rating the perceived importance of the item and an open-ended question asking what participants would add, change, or remove about the item), while others were specific to individual items.

Meanwhile, participants' attitudes regarding the template, as well as learnability and flexibility, were assessed using various rating items that were displayed after participants had finished working on the template. These items inquired about, for example, participants' satisfaction with using the template (*attitude*), how well they understood it (*learnability*), or how well it covered the most important aspects of their research (*flexibility*).

The items used for the usability test were open text input items, single- or multiple-choice items, and rating items with varying scales (see online materials: https://doi.org/10.23668/psycharchives.12959).

To measure the UTAUT variables, the following scales were assessed: a *performance expectancy* scale (i.e., five items measuring participants' expected performance when using the template), an *effort expectancy* scale (i.e., five items inquiring about the expected effort when using the template, where higher scores were associated with lower expected effort), a *social influences* scale (i.e., a scale of five items examining the perceived social pressure to use preregistration), and a *voluntariness* scale (i.e., perceived control over the behavior). The *intention* to use the template (dependent variable) was measured with one variable, as were *age* and *experience* (operationalized as the participants' academic group). In addition, *facilitating conditions* (i.e., a scale of five items) were measured. The latter scale was not part of our hypotheses tests, as it is assumed to influence people's actual behavior, not their intention (Venkatesh et al., 2003). However, this scale was still assessed since it might provide insights into factors that might help foster the practice of preregistration. The UTAUT scales were measured on a seven-point rating scale with 1 = *Disagree* to 7 = *Agree* (Venkatesh et al., 2003).

All UTAUT items were adapted from Venkatesh et al. (2003). The other items were developed based on face validity and revised in consultation with the members of the Preregistration Task Force that developed the PRP-QUANT Template (Bosnjak et al., 2022). Additionally, before data collection, a pre-test was conducted with four participants (two PhD students, one postdoc, and one professor), and its results were used to further improve the items (e.g., by increasing their comprehensibility).

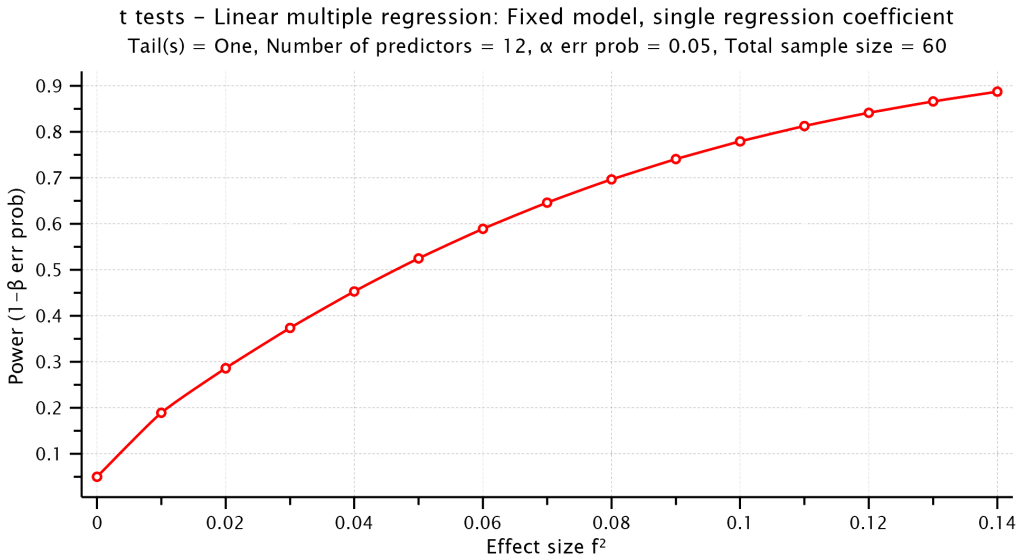
Procedure

Participants received the study link via their respective society's mailing list or social media (see section *Participants*). After the welcome page, participant information was presented, and participants were required to provide informed consent to proceed. They were informed of the study objectives (i.e., evaluation of the PRP-QUANT Template).

At the beginning of the study, the participants provided information about their sociodemographic data

Figure 2

Sensitivity curves for hypothesis tests of individual predictors



Note. Conducted with G*Power (Faul et al., 2007, 2009).

and general use of preregistration. The study then focused successively on the usability of the template and measurement of the UTAUT items. Some items of the study were only shown to participants who had preregistered before (see online materials: https://doi.org/10.23668/psycharchives.12959). Before any PRP-QUANT Template or web probing items were displayed, all participants were shown the entire template. They opened the template in a table format in a new browser tab by clicking on a link provided in the study and were asked to look at the entire template to obtain a general impression. A control question regarding the content of a template item had to be answered correctly to proceed. Participants were asked to keep the template open in the additional tab so that they could refer to it throughout the study.

No items of the study were mandatory besides the filter question at the beginning, which inquired whether the participants worked in psychological research (see section *Participants*). However, for participants' study descriptions and the UTAUT questionnaire, participants who did not respond were asked to confirm their choice to ensure that gaps were not created inadvertently. Additionally, if participants did not respond to the template items, they were asked to provide a reason for this (i.e., whether they thought the item was optional, made a mistake, did not know what to answer, did not like the item, if the item did not fit their research, or they could provide other reasons via open text input). This question itself was not mandatory.

On average, it took participants approximately 31 minutes to complete the study (SD = 12 min, *range* = 54 min, times adjusted for interruptions). The procedure was approved by the ethics committee of Trier University, Germany (approval number: 27/2020). An example screen recording of condition 1 of the procedure and a PDF of the questionnaire for all conditions are available online (https://doi.org/10.23668/psycharchives.12959).

Data Analysis and Pre-Processing

We used R (Version 4.2.2, R Core Team, 2021) and the R-packages *corrplot* (version 0.92, Wei and Simko, 2021), *lm.beta* (version 1.7-2, Behrendt, 2014), *olsrr* (version 0.5.3, Hebbali, 2020), *psych* (version 2.2.9, Revelle, 2021), *RColorBrewer* (version 1.1-3, Neuwirth, 2022), *readxl* (version 1.4.2, Wickham and Bryan, 2022), *Rmisc* (version 1.5.1, Hope, 2022), *tidyverse* (version 2.0.0, Wickham et al., 2019), and *writexl* (version 1.4.2, Ooms, 2021) for all analyses. All analysis scripts and anonymized data (including metadata about variables and values) are publicly accessible online (data: https://doi.org/10.23668/psycharchives.12915; code: https://doi.org/10.23668/psycharchives.14244).

Data were pre-processed by recoding responses from

multiple-choice questions (originally: 1 = *not selected* and 2 = *selected*; new: 0 = *not selected* and 1 = *selected*) and turning single-choice items into factors. The polarity of negatively poled scale items was reversed. All UTAUT items as well as some other items of the web probing and overall evaluation were recoded from "1 to 7" to "-3 to 3", yielding a middle category which has absolute meaning (i.e., 0 = neutral opinion, neither agreement nor disagreement). As specified in the preregistration, empty data and nonsense responses (e.g., random key pressing) were excluded (i.e., 24 of 397 responses for the template items; 27 of 206 responses for the "What would you add, change, or remove . . . ?" items; 1 of 131 responses for the open text input items; and 12 of 268 responses given to the web probing items).

Coding of open text input. Participants' responses to the template items and other open text input items were coded for the analysis. Three coders were involved in this process, but coding was split between coders itemwise so that only one individual coded all responses for one item.

For the participants' responses to the template items, it was coded if the given response matched what was requested in the item (0 = *not applicable*, 1 = *fits poorly*, 2 = *fits moderately*, 3 = *fits well*, -9 = *nonsense answer*). For this, a coding scheme was used, which was developed and published prior to data collection alongside the preregistration (https://doi.org/10.23668/psycharchives.4636). To improve the prespecified coding scheme and represent as many potential responses as possible, the template responses of 25% of participants per condition were randomly selected and coded, while the coding scheme was revised in the process (e.g., the coding categories 0 = *not applicable* and -9 = *nonsense answer* were added). Subsequently, the improved coding scheme was applied to the remaining datasets. The final coding scheme is available online (https://doi.org/10.23668/psycharchives.12959).

Next, open web probing questions and other open text input items were evaluated by coding common themes. Responses were shuffled and the coder read the first 10% of the shuffled responses. They identified common themes mentioned by the participants, which were then transferred to new columns in a coding sheet. Then, it was coded for all other responses if the theme was mentioned (= 1) or not mentioned (= 0). If new relevant topics appeared to the coder that they had not coded before, these were added as categories as the coding continued and were coded later. For the item "What would you add, change, or remove about the item?", common themes were categorized into 1) things to add, 2) things to change, and 3) things to remove.

When asked for definitions, explanations, or examples, a different coding was implemented. For definitions, it was coded if the term was correctly described in the response (= 1) or not (= 0), and for examples it was coded if the examples fit the requested term (= 1) or not (= 0). All coded, anonymized comments are published alongside the data (https://doi.org/10.23668/psycharchives.12915).

Quality check of UTAUT scale items. For the items of the UTAUT scales and the overall evaluation, floor and ceiling effects were inspected, that is, items for which ≥ 90% of participants selected the lowest or highest category. No floor or ceiling effects were found for the overall sample, nor the sample used for the UTAUT analyses. Furthermore, considering only the data of participants included in the hypotheses tests, the reliability of the UTAUT scales was inspected. The reliability analyses showed high to excellent reliability for the performance expectancy (α = .87) and effort expectancy scales (α = .9), adequate reliability for the social influence scale (α = .76), and moderate reliability for the voluntariness scale (α = .62).

Deviations From the Preregistration

All deviations from the preregistered plan are displayed in Table 1 below. For each deviation, a justification is provided.

Results

Usability of the PRP-QUANT Template

Satisfaction, Perceived Effectiveness, Fit to Research Area, and Comprehensiveness. Participants' responses concerning the overall evaluation of the template are displayed in Figure 3. On average, they were rather satisfied with using the template (*Mean* = 0.72, *Median* = 1, *SD* = 1.54, *IQR* = 2, *range* = 6, on a scale from -3 = *very dissatisfied* to 3 = *very satisfied*, see Figure 3A). They rated it as being effective for helping them create a preregistration (*Mean* = 1.18, *Median* = 1, *SD* = 1.41, *IQR* = 1.25, *range* = 6, on a scale from -3 = *very ineffective* to 3 = *very effective*, see Figure 3B). Compared to their favorite preregistration template, the PRP-QUANT Template convinced the participants to about the same extent (*Mean* = 0.04, *Median* = 0, *SD* = 1.51, *IQR* = 2, *range* = 6, on a scale from -3 = *less* to 3 = *more*, see Figure 3C). When asked how likely they would use the template in the future to create their preregistrations, participants indicated an average probability of 61.47% (*Median* = 68, *SD* = 28.51, *IQR* = 33.75, *range* = 100). Additionally, they indicated an average probability of 64.67%

Table 1

Deviations from the preregistration

Section	Description and justification
Recruitment	Reminder emails were sent later than anticipated (not after one week, but after five weeks for the DGPs, and two and a half weeks for the BPS). For the APA, no reminder was sent, instead the study was also advertised in their newsletter.
Participants	Originally, it was planned to include all participants that started the main body of the study in the descriptive reports. However, since many participants dropped out before starting to work on the template items, and these are the core part of the study, we decided to report all descriptive reports for this sub-sample (n = 88).
Pre-processing	In addition to the preregistered pre-processing steps, further quality checks were conducted, but did not result in any modifications in item inclusion. Specifically, reliability as well as floor and ceiling effects were inspected (i.e., it was checked for items of the UTAUT scales and overall evaluation, if ≥ 90% of participants answered the lowest/highest category). Reliability analyses showed moderate to excellent reliability, no items needed to be excluded. No floor or ceiling effects were found.
UTAUT	Since the assumption tests showed a high multicollinearity due to the interaction terms, for the hypotheses test the UTAUT scales were centered instead of recoding them from "1 to 7" to "-3 to 3". In the UTAUT sample, for the "academic group" variable, the option "other" was excluded (n = 1) because it holds no information for the regression model (heterogeneous group). For the scales, instead of displaying means and standard deviations, these were displayed in a plot showing the mean and confidence interval, for easier inspection. It was not clearly defined a priori that one-sided tests would be used for the regression weights, however, since directional hypotheses were tested, this was implemented. This had no impact on the results.
Coding of open comments	For the web probing, it was originally planned to code common themes for questions of the type "how is this item different from another item". However, it makes more sense to code whether the reported differences were perceived correctly (= 1) or incorrectly (= 0).

(*Median* = 72, *SD* = 30.25, *IQR* = 40, *range* = 100) for recommending the template to a colleague.

The PRP-QUANT Template fit quite well to the participants' research areas (*Mean* = 1.22, *Median* = 2, *SD* = 1.38, *IQR* = 1.25, *range* = 5, on a scale from -3 = *not at all* to 3 = *very well*, see Figure 3D) and covered the most important aspects of their studies (*Mean* = 1.65, *Median* = 2, *SD* = 1.34, *IQR* = 2, *range* = 5, on a scale from -3 = *not at all* to 3 = *very well*, see Figure 3E). Additionally, the participants understood the template well (*Mean* = 1.58, *Median* = 2, *SD* = 1.11, *IQR* = 1, *range* = 5, on a scale from -3 = *not at all* to 3 = *very well*, see Figure 3F).

When asked to rate the template's overall balance between comprehensiveness and parsimony, they rated it as rather comprehensive (*Mean* = 0.68, *Median* = 1, *SD* = 1.08, *IQR* = 1, *range* = 6, on a scale from -3 = *too parsimonious* to 3 = *too comprehensive*, see Figure 3G). Moreover, the assessment of the template's complexity varied widely (see Figure 3H), but on average, the participants did not find it unnecessarily complex (*Mean* = −0.08, *Median* = 0, *SD* = 1.99, *IQR* = 3, *range* = 6, on a scale from -3 = *disagree* to 3 = *agree*). Instead, they found it moderately easy to use (*Mean* = 0.35, *Median* = 0.5, *SD* = 1.71, *IQR* = 3, *range* = 6, on a scale from -3 = *disagree* to 3 = *agree*, see Figure 3I), although

Figure 3

Rating of overall satisfaction, perceived effectiveness, fit to research area, and comprehensiveness in percent

Note. All items were rated on a seven-point scale. Percentages are based on all responses to each item (A: $N = 62$; B: $N = 63$; C: $N = 50$; D – I: $N = 63$). Only percentages above 5% are labelled.

it should be noted that 30% of participants still reported a lower level of ease of use.

Participants' Suggestions for Improving the Template. Participants were invited to provide open text

input suggestions to improve the template. Only the themes mentioned more than once are included here, but all coded comments can be inspected online (https://doi.org/10.23668/psycharchives.12915). Of the 37 participants who responded to this item, 13.51% complimented the comprehensiveness of the template and found it to be a good guide for preregistration beginners and early career researchers. However, 32.43% pointed out that the template was very long and specific and that it might be beneficial to reduce its complexity. Correspondingly, 10.81% suggested providing a shorter basic template, where you specify the study type at the beginning and then get more specific items matching your study type. Additionally, 10.81% of the participants indicated that some items seemed redundant and that it would be helpful if the instructions provided additional information to clarify their differences. The participants also provided suggestions for the practical implementation of the template: for example, 8.11% suggested offering it in different formats (which is currently already the case, see https://doi.org/10.23668/psycharchives.4584) and to provide predefined options which can be adjusted if you deviate from them. Another 5.41% wished for example answers that could be used as support when filling in the questionnaire. Furthermore, 16.22% of the participants indicated that, while the template fits best with confirmatory and experimental studies, items for other research types might be added. Paying attention to interdisciplinarity was also suggested by participants in the general comment section at the end of the study (8.7% of 23 responses), while most responses were praise to the template (34.78%) or more general comments about preregistration or the study.

Individual Template Items and Web Probing. Next, the participants' responses and comments regarding the individual template items were inspected. Overall, 88 participants worked on the template items (see section *Participants*), whose responses were consequently used for the analysis of the individual template items and web probing questions. Of these, 21.59% had just started planning the study, 35.23% had planned the study in detail, 13.64% were currently conducting their study, 28.41% had already completed their study, and 1.14% did not indicate their study status. Moreover, 29.55% planned to preregister their study, 19.32% were currently working on the preregistration, 21.59% had already preregistered, 28.41% did not (plan to) preregister the study, and 1.14% did not indicate the preregistration status.

Overall Good Fit of Responses. Of all responses provided for the template items, 48.61% fit well with the item, 27.96% fit moderately, 11.34% fit poorly, 6.05% indicated that the item was not applicable to the participants' studies, and 6.05% were nonsense answers. Figure 4 provides an overview of the response fit for all template items presented that required an open-ended response from participants. Inspecting the plot reveals that for some items, answers were primarily well-fitting (e.g., item T8 "Conflict of Interest Statement", M2 "Use of pre-existing data"), and that most items showed a moderate to good response fit. However, a higher proportion of poor answers were given for the items M13 "Study Materials", M14 "Study Procedures", AP3 "Data preprocessing", AP5 "Descriptive statistics", AP6 "Statistical models" and AP7 "Inference criteria" (i.e., $\geq$ 20% of responses were coded as "poor"). The means, standard deviations, medians, and ranges for each item's fit are displayed in Table S2 in the supplemental material.

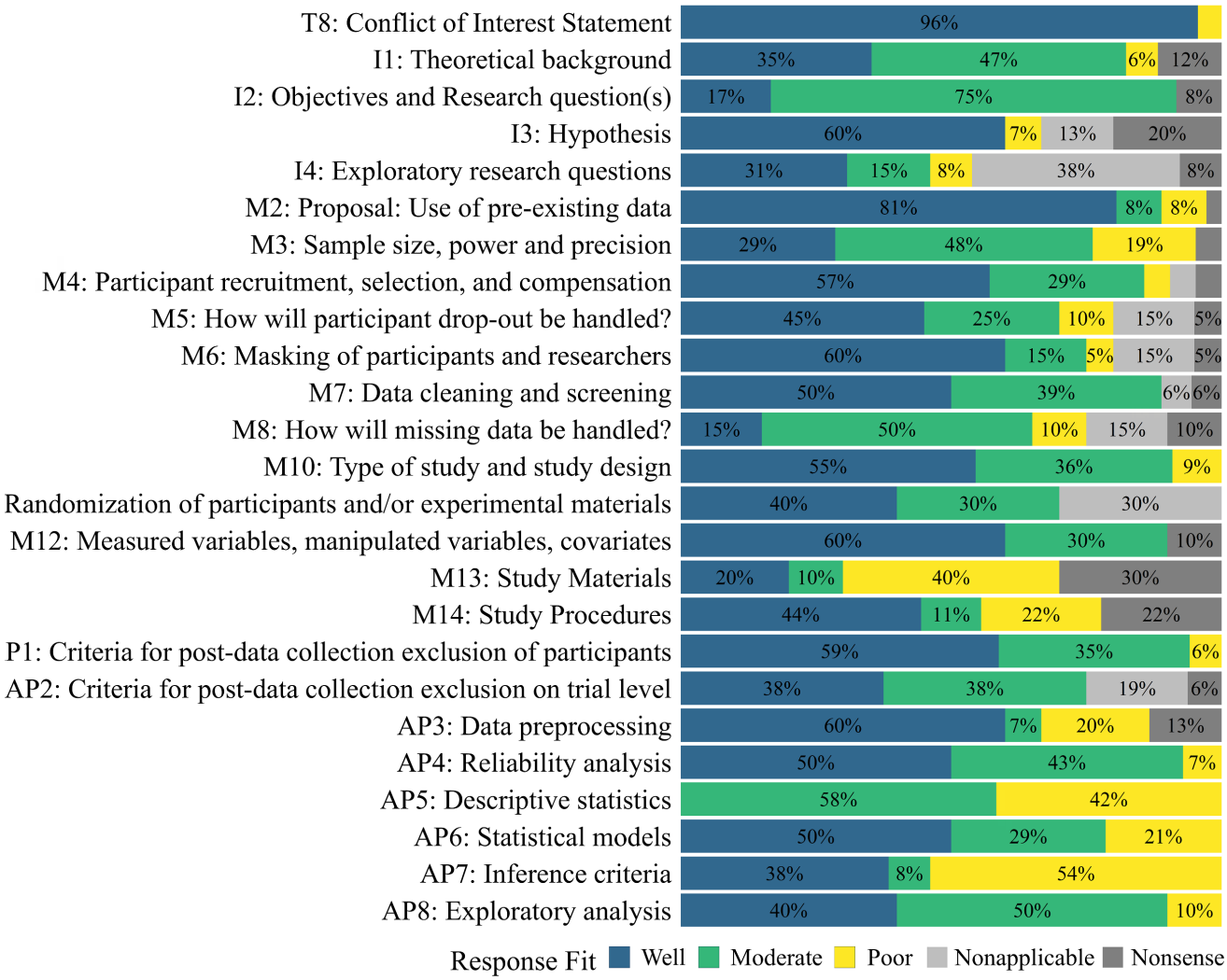
Whenever participants did not answer the template items, they were prompted to provide reasons for doing so (see section *Procedure*). Of the 44 responses to these prompts, 50% indicated that the participants did not know what to answer, 11.36% of the participants said that it was a mistake, 6.82% thought the item was optional, 4.55% did not like the item, 2.27% said that it did not apply to their research, and 25% gave other reasons, most of which aligned with the given options.

Template Items Perceived as Important for Preregistration. Participants felt that most of the items in the PRP-QUANT Template were important for preregistering their studies (see Figure 5). The items rated most important were AP6 "Statistical models", M12 "Measured variables, manipulated variables, covariates", M14 "Study Procedures", M10 "Type of study and study design", M13 "Study Materials", and M1 "Time point of registration" (i.e., their mean was above 2 on a scale from -3 = *not important at all* to 3 = *very important*). Most other items were also rated as important (i.e., their mean was above 0, for most items above 1). The item I4 "Exploratory research questions" was rated as neither important nor unimportant. Still, most participants felt that including exploratory research questions and analyses in the preregistration was appropriate (i.e., for research questions, 56.25% indicated "definitely yes" or "maybe yes", and 57.14% did so for analyses). The items AP4 "Reliability analysis", T12 "Optional: Standard lab practices", and AP5 "Descriptive statistics" were rated as relatively unimportant (i.e., their mean was below 0). This makes sense in that reliability analyses are not applicable to all studies, providing standard lab practices is optional (and only three of the 20 participants who answered even had a standard lab practices document), and descriptive statistics have no direct impact on hypotheses testing.

Participants' Suggestions for Individual Items and

Figure 4

Fit of the participants' responses to the PRP-QUANT Template items



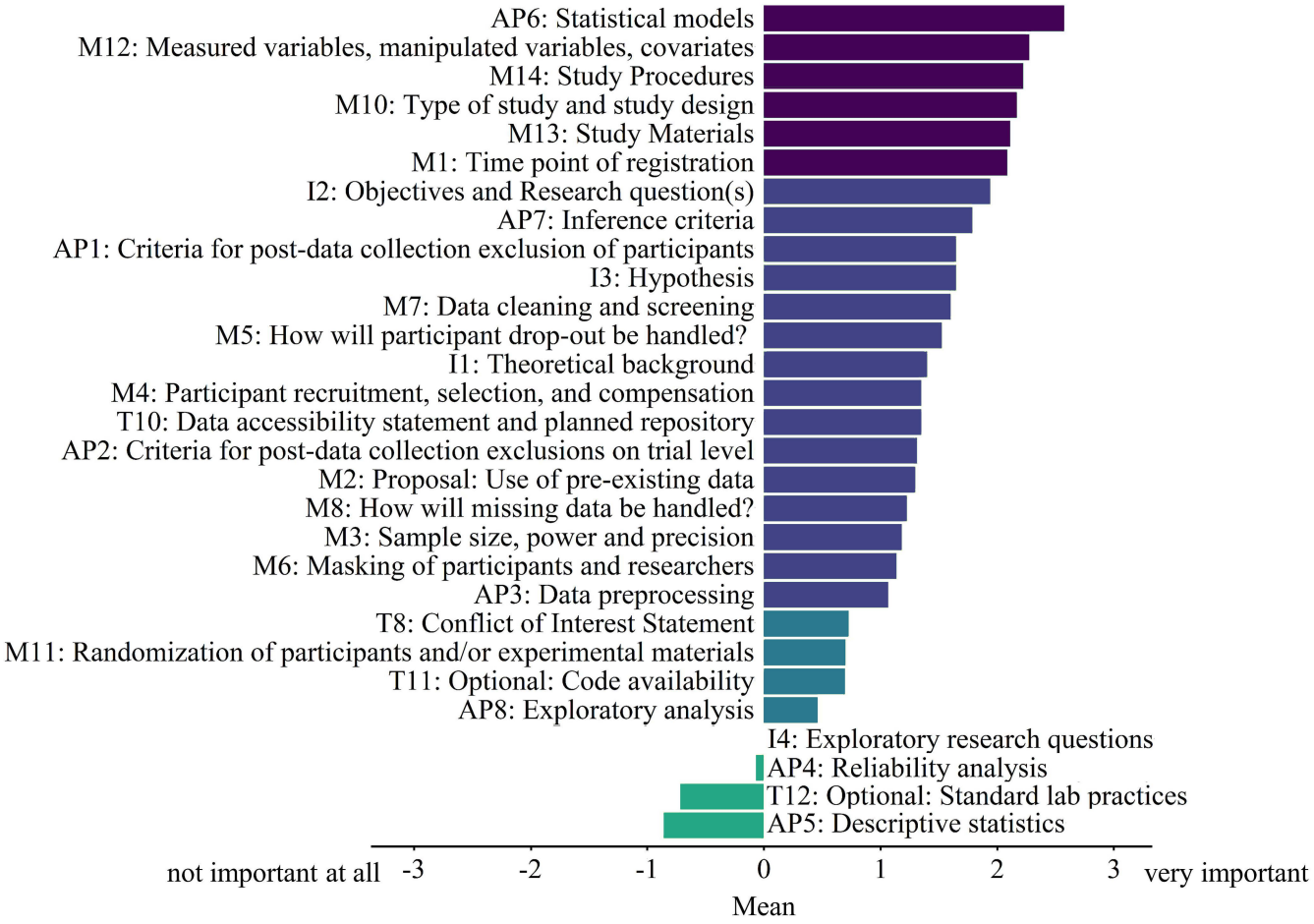
Web Probing. For each template item that the participants worked on, they were asked what they would add, change, or remove. The participants offered a variety of comments and suggestions, which are summarized in Table S3 in the supplemental material. Additionally, participants responded to several other web probing items (see online materials for a complete list: https://doi.org/10.23668/psycharchives.12959) which queried, for example, why they had selected an answer, whether they correctly understood the concepts underlying the items and which were unclear, how they perceived the link between items, and whether they could distinguish items from others. These are presented in detail in Text S2 in the supplemental material.

Significant Prediction of Intention by Performance Expectancy and All Predictors Combined

To investigate participants' intention to use the PRP-QUANT Template in the future, the UTAUT items were analyzed. For each participant, the mean scores were computed for all UTAUT scales (i.e., performance expectancy, effort expectancy, social influence, facilitating conditions, and voluntariness). The means and confidence intervals for all scales are displayed in Figure 6.

Figure 5

Importance rating of PRP-QUANT Template items



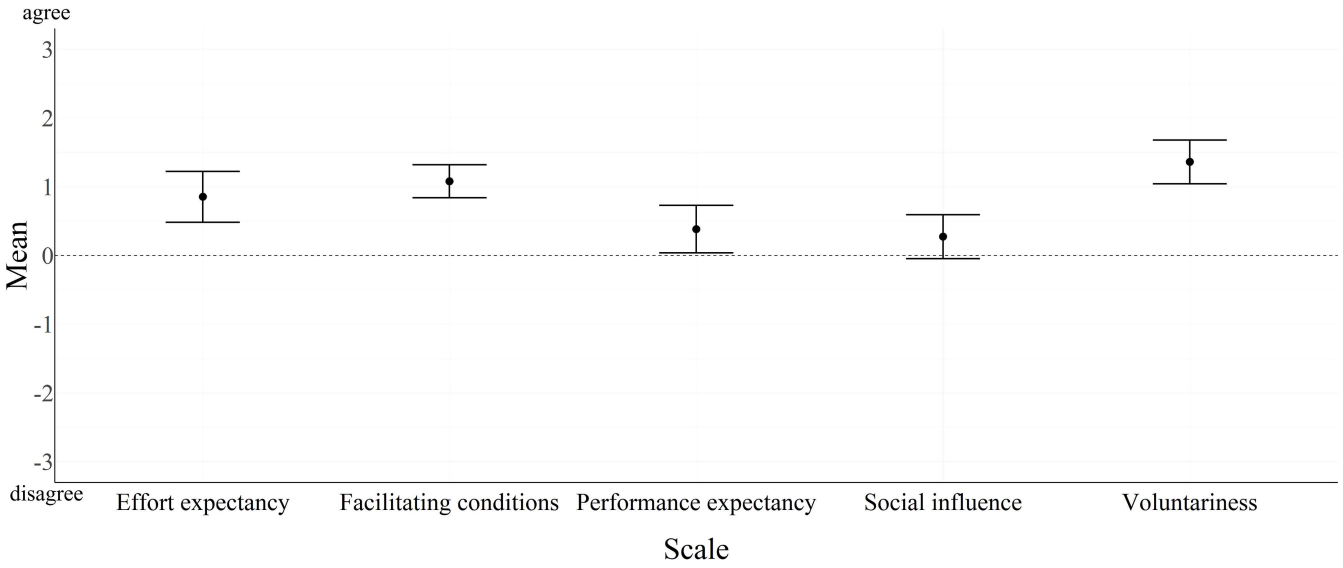
We expected that the intention to use the template in the future is predicted by performance expectancy (moderated by age), effort expectancy (moderated by age and experience, i.e., academic group), and social influence (moderated by age, experience, i.e., academic group, and voluntariness of use, see Figure 1). To test these hypotheses, a moderated multiple regression model was computed, which is a method that has been frequently used to test the UTAUT (see Williams et al., 2015).

Behavioral intention (i.e., the answer to the item "How likely would you use the template in the future to create your preregistrations?") was included as the dependent variable. The following predictors were included in the model: 1) the score on the performance expectancy scale, 2) performance expectancy × age, 3) the score on the effort expectancy scale, 4) effort expectancy × age, 5) effort expectancy × experience (i.e., academic group), 6) the score on the social influence scale, 7) social influence × age, 8) social influence × experience, 9) social influence × voluntariness, 10) age, 11) experience, and 12) voluntariness. The significance of the overall model, as well as of individual predictors and moderating effects, was evaluated at $\alpha = .05$. Because of our directional hypotheses, the regression weights were tested in a one-tailed fashion.

Before computing the moderated regression model, its assumptions were tested: linearity, uncorrelated predictors, independence and normality of residuals, and homogeneity of variance. The assumption tests showed high multicollinearity among the predictors, as judged based on the variance inflation factors (VIF > 10 for seven predictor/interaction terms). Because of this, all predictors except for experience (nominal variable) were centered, which drastically reduced multicollinearity (VIF < 10 for all except one predictor).

Figure 6

UTAUT Scales



Note. Scales ranged from 1 = *Disagree* to 7 = *Agree* and were recoded to "-3 to 3" (however, these were centered for the hypotheses tests, see below). The parameters were calculated based on the sample used for the UTAUT hypotheses tests (i.e., participants who responded to all items used in these analyses, N = 60). Error bars indicate 95% confidence intervals. Higher effort expectancy scores are associated with lower expected effort. Facilitating conditions were considered descriptively but were not included in the hypotheses tests.

As expected, the predictors of the UTAUT combined were able to significantly predict researchers' intention to use the PRP-QUANT Template in the future, $F(15, 44) = 3.94$, $p < .001$, $R^2_{adjusted} = 42.79\%$. However, of the individual predictors, only performance expectancy was a significant predictor for intention, $t(44) = 2.28$, $p_{one-sided} = .014$, $\beta = .36$. The interaction of performance expectancy and age, as well as all other effects, were not significant (all $ps_{one-sided} > .05$). As a sensitivity analysis, we re-ran the analyses using the preregistered model with un-centered predictors. Here, performance expectancy was non-significant, $t(44) = 1.58$, $p_{one-sided} = .06$, $\beta = .87$.

Study 2: Survey of Preregistration Authors and Reviewers

In study 2, we surveyed psychological researchers who had used the PRP-QUANT Template to create a preregistration and submitted it when applying for free-of-charge data collection in ZPID's call for online studies. In addition, call reviewers were asked how they felt about reviewing preregistrations based on the PRP-QUANT Template. Instead of being constructed based on theoretical considerations, these surveys were designed to provide a quick exploration of the participants' experiences with the template and processes within the call. Thus, while some of the items of study 2 related to other aspects of usability measured in study 1, they mostly focused on participants' attitudes regarding the template.

Methods

This study was not preregistered as it was conducted on a short notice. It does not include hypotheses tests. Instead, we report the survey results descriptively.

Participants

Twenty-eight preregistrations reached the review stage of the call, which covered a variety of different research areas (i.e., social, organizational, personality, clinical, experimental, and developmental psychology, psychology of climate change, media and technology, neuropsychology, meta-science, and misinformation). After this stage, all submitting authors (i.e., mostly the first author) were invited to participate in the author survey. Nineteen authors participated in this study. Data were collected between May 16, 2022, and May 29, 2022. After the initial invitation, two reminders were sent to all potential participants.

Meanwhile, the 44 researchers who reviewed an accepted proposal were invited to participate in the reviewer survey. Twenty-nine participated. Their data were collected between March 11, 2022, and March 26, 2022, with one reminder sent to them. Participants of both surveys were not compensated.

Material and Measures

Both surveys were created using Google Forms (https://docs.google.com/forms). The author survey consisted of 21 items and the reviewer survey consisted of 17 items. There were five different sections of items in the author survey: 1) items about the participants' previous and future use of preregistration and PsychLab, 2) items regarding their experiences with using the structured template compared to a continuous format (i.e., a normal report) and 3) their experiences during the review process, 4) questions about their general opinion of preregistration, and 5) additional comments. The shorter reviewer survey had three different sections of items: 1) items asking participants to compare their experiences with reviewing this structured format versus a continuous text, 2) questions concerning their general impression of reviewing preregistrations instead of full manuscripts, and 3) additional comments. Some of the items related more generally to the call's processes and participants' general impression of preregistration, however, we only present those responses related to the usability of the template in this article.

Of the various aspects of usability measured in study 1 (Shackel, 2009), attitudes were the focus of study 2. Authors were asked whether they felt that the structure of the template facilitated creating their preregistration, whether it made them think of details that were important for planning their study (which can also be seen as an indicator of flexibility), whether the template helped them include all relevant information, and whether they felt that the items should be reduced. Similarly, reviewers were asked whether they felt that the structure of the template facilitated their evaluation, helped them find the information more easily, focus their attention on the relevant sections, and assess the completeness of the information, or whether, in contrast, the template hindered their reading flow or contained too many elements irrelevant to the review. Participants were also asked whether they would have preferred to prepare/review the proposal in this structured preregistration format or in a normal report format. As a further aspect, learnability was measured by asking the authors if they had difficulty understanding what they were supposed to fill in for some items.

Most items were 7-point rating items with 1 = *strongly disagree* and 7 = *strongly agree*. In addition, participants were given the opportunity to express their opinions in several open text input items (e.g., "Anything else you would like to add about the template? Please comment here"). An overview of the items in the author and reviewer surveys is available online (https://doi.org/10.23668/psycharchives.12959).

Procedure

The surveys took approximately five minutes to complete. Participants were invited via personal email, and data were collected anonymously. The processing of data was explained in writing at the beginning of the survey. The survey was conducted after the decision for or against funding of the participants' studies had been made.

In both surveys, participants successively completed the different item sections (for authors: use of preregistration and PsychLab, comparison of structured template versus continuous text, review process, general opinion of preregistration, and additional comments; for reviewers: comparison of structured template versus continuous text, comparison of reviewing preregistrations versus complete manuscripts, and additional comments), which were each displayed on a new page. None of the items were mandatory.

Data Analysis and Pre-Processing

Again, R (Version 4.2.2, R Core Team, 2021) was used for the analysis. Data of the rating items were pre-processed similarly to study 1, that is, they were recoded from "1 to 7" to "-3 to 3" to facilitate interpretation. Scales that deviated from this format were not recoded (these are labelled accordingly below). For all rating items, the mean, median, standard deviation, interquartile range, and range were calculated. Percentages were computed to examine preregistration experience and intention. Open-ended comments were manually reviewed and summarized. As for study 1, all data and analysis scripts are available online (data: https://doi.org/10.23668/psycharchives.12915; code: https://doi.org/10.23668/psycharchives.14244).

Results

Author Responses

Of all authors who participated in the survey, 31.58% had preregistered for the first time (i.e., 68.42% had previous preregistration experience). Overall, 94.74% of participants intended to preregister in the future (i.e., their scores were above 3.5 on a scale from 0 = *very unlikely* to 7 = *very likely*), of which 73.68% selected *very likely*.

Overall, the authors rated the PRP-QUANT Template favorably. They mostly agreed that it facilitated creating their preregistration (*Mean* = 1.63, *Median* = 2, *SD* = 1.01, *IQR* = 1, *range* = 3), that it made them think of details that were important for planning their study (*Mean* = 1.79, *Median* = 2, *SD* = 0.92, *IQR* = 1, *range* = 3), and that it helped them include all relevant information (*Mean* = 2.06, *Median* = 2, *SD* = 0.87, *IQR* = 0.75, *range* = 3). They mostly disagreed with the statements that the template items should be reduced (*Mean* = −0.47, *Median* = −1, *SD* = 1.84, *IQR* = 3, *range* = 6), that they (i.e., the authors) had difficulty understanding what they were supposed to fill in on some items (*Mean* = −1.16, *Median* = −2, *SD* = 1.83, *IQR* = 3.5, *range* = 5), and that they would have preferred to write the proposal in a normal report format (continuous text) rather than a structured format (*Mean* = −1.37, *Median* = −2, *SD* = 1.64, *IQR* = 2, *range* = 5).

In the open comments, most participants expressed satisfaction with using the template. Some suggestions for improvements were made, each voiced by one participant, respectively. For example, it was suggested to provide a front page with a link to all different subsections to make it easier for authors and reviewers to navigate the document, shorten the template and reduce redundancies, revise the structure of the Word template, and query the abstract in one item rather than subdividing it.

Reviewer Feedback

Reviewers also perceived various advantages of the PRP-QUANT Template. They described that the structure of the template facilitated their evaluation (*Mean* = 1.07, *Median* = 1, *SD* = 1.25, *IQR* = 2, *range* = 4), that it helped them find the information (*Mean* = 1.03, *Median* = 1, *SD* = 1.4, *IQR* = 2, *range* = 4), that the structure of the template helped them focus their attention on the relevant sections (*Mean* = 1.21, *Median* = 1, *SD* = 1.29, *IQR* = 2, *range* = 5), and that it helped them evaluate the completeness of the information (*Mean* = 1.14, *Median* = 2, *SD* = 1.46, *IQR* = 2, *range* = 5). Correspondingly, they would not have preferred to review the proposal in a normal report format of continuous text rather than in a structured format (*Mean* = −0.83, *Median* = −1, *SD* = 1.56, *IQR* = 2, *range* = 5), they did not feel that the layout of the template hindered their reading flow (*Mean* = −0.9, *Median* = −1, *SD* = 1.8, *IQR* = 3, *range* = 5), or that the template contained too many elements irrelevant to reviewing the proposal (*Mean* = −1.1, *Median* = −2, *SD* = 1.59, *IQR* = 2, *range* = 5).

In the open text field, reviewers commented that they found the template generally helpful and provided some suggestions for improvements. For example, it was suggested to add an item about scientific and thematic relevance, or the possibility of including scripts and results from data analysis (e.g., power analyses). This is already possible by using the PRP-QUANT Template in R Markdown or JupyterLab. Furthermore, it was commented that the template could be more concise, and that the items A2 “Objectives and Research questions” and I2 “Objectives and Research question(s)” seemed redundant (however, since A2 is part of the abstract, these items ask for different depths of information). One person recommended dividing the template into two sections to facilitate reviewing: Authors could elaborate in the first section everything relevant to reviewing and then give all the relevant technical information in the second section.

Discussion

We conducted two studies to evaluate the usability of the PRP-QUANT Template and identify areas for improvement. Furthermore, we wanted to find out whether psychological researchers plan to use the template in the future to create their preregistrations and examine which variables might be important for this intention formation.

Usability of the Template Rated High, With Suggestions for Improvements

We assessed the usability of the PRP-QUANT Template in study 1 by conducting a simulation trial in which we asked psychological researchers to think of one of their studies to complete selected parts of the template, and in study 2 by surveying authors and reviewers of preregistrations that were part of a call for online studies. For this evaluation, we referred to the four aspects of usability defined by Shackel (2009): learnability, flexibility, effectiveness, and attitude.

In both study 1 and study 2, participants indicated that they understood the template well and that it covered the most important aspects of their studies. This suggests that both the learnability and flexibility of the template are adequate and that the template seems to capture the main points of various sub-disciplines. Still, it seems worthwhile to simplify the template to enhance the perceived ease of use. For example, some participants struggled to understand complex terms and differentiate between similar items, which could be improved, for instance, by providing examples. Table S3 in the supplemental material provides an overview of suggestions to improve the various template items, sampled from our participants.

Moreover, various indicators point to the effectiveness of the PRP-QUANT Template. More than three-quarters of participants' responses to the template items matched the requested information moderately or well, with nearly half of the responses fitting well. Participants also felt that most of the template items were important for preregistering their studies, with the most highly rated items being AP6 "Statistical models", M12 "Measured variables, manipulated variables, covariates", M14 "Study Procedures", M10 "Type of study and study design", M13 "Study Materials", and M1 "Time point of registration". Nevertheless, there was a higher proportion of poor answers for the items M13 "Study Materials", M14 "Study Procedures" AP3 "Data preprocessing", AP5 "Descriptive statistics", AP6 "Statistical models" and AP7 "Inference criteria". These items could be prime candidates for revision. However, it must be noted that these are also items that require very elaborate responses. Since study 1 imposed quite high demands on the participants while no compensation was given, it may be assumed that this could be the lower end of the scale of possible response quality. Nevertheless, the participants had various suggestions on how these and other items could be improved (see Table S3), which could be included in a new version of the template.

Participants' attitudes towards the PRP-QUANT Template were also rather positive and they indicated an average probability of over 60% that they would use it in the future to create their preregistrations or recommend it to a colleague. However, while the template was not considered unnecessarily complex on average, there was still a high proportion of participants who found it rather complex and commented on its length on several occasions. Accordingly, many of the participants' suggestions were aimed at requesting the information in a more condensed form, which could possibly be considered in a new version of the template.

In summary, although being based on relatively small samples, our results suggest a good usability of the PRP-QUANT Template and point to a number of possible further improvements. The next step is to implement these in a new version of the template. For this purpose, we organized the hackathon "Community Revision of the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template" at the SIPS 2023 conference (June 2023, Padova, Italy) where the template was revised based on our study findings and the hackathon participants' suggestions. Additionally, we are currently conducting a study to investigate the PRP-QUANT Template's ability to restrict researcher degrees of freedom (Spitzer and Mueller, 2024), which received in-principle-acceptance by PCI RR (Lakens, 2024). Based on our usability test, the feedback received during the hackathon, and this additional study, an updated version of the template will be published in PsychArchives in the future.

Prediction of Intention Primarily Through Expected Performance Gains

Besides evaluating the usability of the PRP-QUANT Template, study 1 examined researchers' intention to use it in the future, as well as possible influences on that intention. Based on the UTAUT (Venkatesh et al., 2003, 2016), we expected that the intention to use the template is influenced by performance expectancy (moderated by age), effort expectancy (moderated by age and experience), and social influence (moderated by age, experience, and voluntariness of use).

Our results show that participants' average intention to use the template in the future to create their preregistrations was rather high (61.47%). Descriptively, all UTAUT variables indicated a positive perception of the template, that is, all scale means were above zero, where zero indicated a neutral opinion and positive values indicated a positive opinion. This suggests that participants tended to believe that using the template would help them attain gains in performance (*performance expectancy*) and that the template would be easy to use (*effort expectancy*). Additionally, participants felt that, while preregistration is voluntary (*voluntariness*), others would approve of them preregistering (*social influence*) and that organizational and technical infrastructures exist that support the preregistration process (*facilitating conditions*).

As hypothesized, all predictors of the UTAUT combined were able to significantly predict researchers' intention to use the PRP-QUANT Template in the future. However, of the individual predictors, only performance expectancy significantly predicted intention. This suggests that the expectation that the template will be useful for one's own research has the strongest influence on whether researchers plan to use it. Highlighting this benefit could therefore help raise researchers' awareness and adoption of the template in the future. For example, an explanatory text could be presented alongside a new version of the template, emphasizing that it is worth investing the effort in this detailed template, as it enables precise planning and helps decrease the workload during the later stages of the study (i.e., analysis and reporting).

Limitations

The implementation of study 1 had some limitations. Responding to the template represented a considerable amount of effort for the participants, for which they

were not compensated. It may be assumed that the quality of the responses reported here represents the lower bound, as researchers would likely put much more effort into creating their actual preregistrations. This suggests that the fit of the answers might be even better in a real deployment of the template. In the future, this could be tested by examining the preregistrations of the authors we surveyed in study 2, as they created their preregistrations to apply for a high external incentive and therefore likely spent more time on their preregistrations. However, this does not undermine the participants' suggestions for improvements that can be used to revise the template.

The format and method of responding to the items in study 1 were also constrained. We presented participants with the PRP-QUANT Template in a table format and queried the template items one after another in our online questionnaire, where participants could not skip back and forth. Again, it can be assumed that satisfaction with the template would probably be even higher if researchers could freely choose between all available formats (e.g., table, text, online form, R Markdown, and Jupyter Notebook) and be able to switch flexibly between items. In addition, if they were using the template to preregister a study outside the present usability test, they would probably invest much more time and would not have to complete responding to the items in one session, which would likely further improve their opinion of the template and the quality of their responses. In line with this assumption, the template was evaluated very positively in study 2, both by authors and reviewers.

In addition to the limitations of our chosen format, our studies faced some sample restrictions. Although participants were drawn from diverse sub-disciplines within psychology, we deliberately excluded fields outside of psychology due to the focused nature of the PRP-QUANT Template, which specifically aids in preregistering quantitative studies in psychology. Furthermore, given our recruitment strategy through the mailing lists of the APA, BPS, DGPs, and social media, our participants came from North America and Europe. Thus, our results may not be generalizable to researchers from other countries. Additionally, it should be noted that the group size per condition was relatively small. However, this does also not limit the usefulness of the participants' suggestions for revising the template.

Regarding our hypotheses tests, it must also be noted that our a priori power analysis was calculated with respect to the overall model, not the individual predictors. It could be that other predictors besides performance expectancy are important for predicting the intention to use the template but were not detectable with our sample size. In addition, there was multicollinearity in our model, which we could improve by centering, but not eliminate completely.

Lastly, on a more general note, there has been some challenge to the validity of the UTAUT in a recent meta-analysis (Blut et al., 2021). Nevertheless, the robustness of the UTAUT and its main effects has been repeatedly validated by research (e.g., see Jadil et al., 2021; Khechine et al., 2016; Venkatesh et al., 2016). The authors of the meta-analysis argue that there is not one specification of UTAUT that applies to all contexts, but that the ability of the theory to predict the use of a new system is context dependent. Besides low power, this might have also contributed to some of the predictors not being significant in our model.

Future Research

Our studies examined the usability of the PRP-QUANT Template and identified its strengths and areas for improvement. These can now be used to create an empirically founded revision.

In the future, usability studies should be used to continually adapt the template to the needs of the community. Other templates could also benefit from such usability assessments. For this purpose, our study could be repeated for new versions of the PRP-QUANT Template or other templates. Their results could then be used to gradually revise the templates, similar to our approach of conducting a community revision hackathon.

Additionally, studies could not only focus on the templates themselves, but also on preregistrations created with the respective templates, following the approach of Bakker et al. (2020), Heirene et al. (2021), or van den Akker, Bakker, et al. (2023). This would allow an empirical investigation of the effectiveness of the templates in reducing researchers' degrees of freedom.

Conclusion

In two studies, we identified both strengths and areas for improvement in the PRP-QUANT Template. We obtained insights into learnability, flexibility, effectiveness, and attitudes, as well as participants' comments and suggestions regarding the template. These can now serve as the basis for an empirically informed revision. Moreover, we demonstrated that performance expectancy, as well as all variables of the UTAUT combined, significantly predicted psychological researchers' intention to use the template in the future. Overall, participants were likely to use the template or recommend it to a colleague, which indicates that the template is being well received.

Author Note

We have presented the results of these studies at the SIPS 2023 conference (June 2023, Padova, Italy) as part of the hackathon "Community Revision of the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template". Based on our study findings and the feedback received during the hackathon, an updated version of the template was published in PsychArchives https://doi.org/10.23668/psycharchives.15193.

Author Contact

Lisa Spitzer: 0000-0002-4925-7291
Michael Bosnjak: 0000-0002-1431-8461
Stefanie Mueller: 0000-0002-3611-6190

Corresponding author: Lisa Spitzer, ls@leibniz-psychology.org

Author Contributions

Lisa Spitzer: conceptualization, data curation (study 1), formal analysis, investigation, methodology, software, visualization, writing – original draft; Michael Bosnjak: conceptualization, methodology, supervision, writing – review & editing; Stefanie Mueller: conceptualization, data curation (study 2), investigation, methodology, software, project administration, supervision, writing – review & editing.

Acknowledgments

We would like to thank the members of the Joint Psychological Societies Preregistration Task Force for their feedback on the items of study 1, our student assistants for coding, and T. Dauber and J. Pauquet for proofreading.

Data Accessibility Statement

All anonymized data (including meta-data about variables and values and coded comments for the qualitative analyses) as well as all analysis scripts and experimental materials are publicly accessible in the digital research repository PsychArchives: Data: https://doi.org/10.23668/psycharchives.12915. Code: https://doi.org/10.23668/psycharchives.14244. Experimental materials: https://doi.org/10.23668/psycharchives.12959.

Conflict of Interest and Funding

All authors declare no conflicts of interest. Michael Bosnjak and Stefanie Mueller were authors of the taskforce that created the PRP-QUANT Template but have no financial interest in the results of the presented studies. We received no funding for these studies.

Ethical Approval

APA ethical standards were followed in the conduct of our studies. Study 1 was approved by the ethics committee of Trier University, Germany (approval number: 27/2020), and participants provided informed consent before starting the main body of the study.

Study 2 involved a brief anonymous survey conducted among participants of our service PsychLab. The processing of data was explained in writing at the beginning of the survey. The survey took about five minutes to complete and was conducted after the decision for or against funding of the participants' studies had been made. Formal ethical approval was not obtained.

Supplemental Material

A supplemental material file (PDF) is provided, which includes the following:

Text S1. Plain language summary. A plain language summary is provided to increase the accessibility of these research findings to a wider audience. It is available in English and German.

Table S1. Research areas. Research topics indicated by the sample of study 1 are displayed in a table.

Table S2. Overview of template response fit. This table shows the fit of responses given to the template items in study 1.

Table S3. What would participants add, change, or remove about the template items. A table summarises all comments on what participants would add, change, or remove about the template items.

Text S2. Web probing. The results of the web probing items are reported here.

Open Science Practices

This article earned the Preregistration+, Open Data, Open Materials, and Open Code badge for preregistering the hypothesis and analysis before data collection, and for making the data, materials, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Bakker, M., Veldkamp, C. L. S., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., Soderberg, C. K., Mellor, D., & Wicherts, J. M. (2020). Ensuring the quality and specificity of preregistrations. *PLOS Biology*, *18*(12), e3000937. https://doi.org/10.1371/journal.pbio.3000937

Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2017). Web probing – implementing probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions. *Mannheim, GESIS – Leibniz-Institute for the Social Sciences (GESIS – Survey Guidelines)*. https://doi.org/https://doi.org/10.15465/gesis-sg_en_023

Behrendt, S. (2014). *Lm.beta: Add standardized regression coefficients to linear-model-objects* [R package version 1.7-2]. https://CRAN.R-project.org/package=lm.beta

Beyer, F., Flannery, J., Gau, R., Janssen, L., Schaare, L., Hartmann, H., Nilsonne, G., Martin, S., Khalil, A., Lipp, I., Puhlmann, L., Heinrichs, H., Mohamed, A., Herholz, P., Sicorello, M., & Panagoulas, E. (2021). *A fMRI pre-registration template*. PsychArchives. https://doi.org/10.23668/PSYCHARCHIVES.5121

Blut, M., Chong, A., Tsiga, Z., & Venkatesh, V. (2021). Meta-Analysis of the Unified Theory of Acceptance and Use of Technology (UTAUT): Challenging its Validity and Charting a Research Agenda in the Red Ocean. *Journal of the Association for Information Systems, Forthcoming*. https://ssrn.com/abstract=3834872

Bosnjak, M., Fiebach, C. J., Mellor, D., Mueller, S., O'Connor, D. B., Oswald, F. L., & Sokol, R. I. (2022). A template for preregistration of quantitative research in psychology: Report of the joint psychological societies preregistration task force. *American Psychologist*, *77*(4), 602–615. https://doi.org/10.1037/amp0000879

Brandt, M. J., IJzerman, 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

Crüwell, S., & Evans, N. J. (2021). Preregistration in diverse contexts: A preregistration template for the application of cognitive models. *Royal Society Open Science*, *8*(10), 210155. https://doi.org/10.1098/rsos.210155

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Hebbali, A. (2020). *Olsrr: Tools for building ols regression models* [R package version 0.5.3]. https://CRAN.R-project.org/package=olsrr

Heirene, R., LaPlante, D., Louderback, E. R., Keen, B., Bakker, M., Serafimovska, A., & Gainsbury, S. M. (2021). *Preregistration specificity & adherence: A review of preregistered gambling studies & cross-disciplinary comparison* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/nj4es

Hope, R. M. (2022). *Rmisc: Ryan miscellaneous* [R package version 1.5.1]. https://CRAN.R-project.org/package=Rmisc

International Organization for Standardization. (2018). *ISO 9241-11:2018: Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. https://www.iso.org/standard/63500.html

Jadil, Y., Rana, N. P., & Dwivedi, Y. K. (2021). A meta-analysis of the UTAUT model in the mobile banking literature: The moderating role of sample size and culture. *Journal of Business Research*, *132*, 354–372. https://doi.org/10.1016/j.jbusres.2021.04.052

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. *PLoS ONE*, *10*(8), e0132382. https://doi.org/10.1371/journal.pone.0132382

Khechine, H., Lakhal, S., & Ndjambou, P. (2016). A meta-analysis of the UTAUT model: Eleven years later: A meta-analysis of the UTAUT model: Eleven years later. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, *33*(2), 138–152. https://doi.org/10.1002/cjas.1381

Lakens, D. (2024). Examining the restrictiveness of the PRP-QUANT Template. *Peer Community in Registered Reports*. https://doi.org/https://rr.peercommunityin.org/articles/rec?id=480

Leiner, D. J. (2019). SoSci Survey [Computer software]. https://www.soscisurvey.de/

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Neuwirth, E. (2022). *Rcolorbrewer: Colorbrewer palettes* [R package version 1.1-3]. https://CRAN.R-project.org/package=RColorBrewer

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Ooms, J. (2021). *Writexl: Export data frames to excel 'xlsx' format* [R package version 1.4.2]. https://CRAN.R-project.org/package=writexl

Parsons, S., Azevedo, F., Elsherif, M. M., Guay, S., Shahim, O. N., Govaart, G. H., Norris, E., O'Mahony, A., Parker, A. J., Todorovic, A., Pennington, C. R., Garcia-Pelegrin, E., Lazić, A., Robertson, O., Middleton, S. L., Valentini, B., McCuaig, J., Baker, B. J., Collins, E., … Aczel, B. (2022). A community-sourced glossary of open scholarship terms. *Nature Human Behaviour*, *6*(3), 312–318. https://doi.org/10.1038/s41562-021-01269-4

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., Buttrick, N., DeBell, M., Ebersole, C. R., Lundmark, S., MacInnis, B., O'Donnell, M., Perfecto, H., Pustejovsky, J. E., Roeder, S. S., Walleczek, J., & Schooler, J. (2020). *High Replicability of Newly-Discovered Social-behavioral Findings is Achievable* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/n2a9x

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research* [R package version 2.2.9]. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psych

Shackel, B. (2009). Usability – Context, framework, definition, design and evaluation. *Interacting with Computers*, *21*(5-6), 339–346. https://doi.org/10.1016/j.intcom.2009.04.007

Spitzer, L., & Mueller, S. (2023). Registered report: Survey on attitudes and experiences regarding preregistration in psychological research. *PLOS ONE*, *18*(3), e0281086. https://doi.org/10.1371/journal.pone.0281086

Spitzer, L., & Mueller, S. (2024). Stage 1 Registered Report: Restriction of researcher degrees of freedom through the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template. https://doi.org/https://doi.org/10.23668/PSYCHARCHIVES.14119

Swaen, G. G., Teggeler, O., & van Amelsvoort, L. G. (2001). False positive outcomes and design characteristics in occupational cancer epidemiology studies. *International journal of epidemiology*, *30*(5), 948–954. https://doi.org/10.1093/ije/30.5.948

Van Den Akker, O. R., Bakker, M., Van Assen, M. A. L. M., Pennington, C. R., Verweij, L., Elsherif, M. M., Claesen, A., Gaillard, S. D. M., Yeung, S. K., Frankenberger, J.-L., Krautter, K., Cockcroft, J. P., Kreuer, K. S., Evans, T. R., Heppel, F., Schoch, S. F., Korbmacher, M., Yamada, Y., Albayrak-Aydemir, N., … Wicherts, J. M. (2023). *The effectiveness of preregistration in psychology: Assessing preregistration strictness and preregistration-study consistency* (preprint). MetaArXiv. https://doi.org/10.31222/osf.io/h8xjw

Van Den Akker, O. R., Van Assen, M. A. L. M., Bakker, M., Elsherif, M., Wong, T. K., & Wicherts, J. M. (2023). Preregistration in practice: A comparison of preregistered and non-preregistered studies in psychology. *Behavior Research Methods*, *56*(6), 5424–5433. https://doi.org/10.3758/s13428-023-02277-0

Van den Akker, O. R., Weston, S., Campbell, L., Chopik, B., Damian, R., Davis-Kean, P., Hall, A., Kosie, J., Kruse, E., Olsen, J., Ritchie, S., Valentine, K., Van 't Veer, A., & Bakker, M. (2021). Preregistration of secondary data analysis: A template and tutorial. *Meta-Psychology*, *5*. https://doi.org/10.15626/MP.2020.2625

van 't Veer, A. E., & Giner-Sorolla, R. (2016). Preregistration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y., & Xu, X. (2016). Unified Theory of Acceptance and Use of Technology:

A Synthesis and the Road Ahead. *Journal of the Association for Information Systems*, *17*(5), 328–376. https://ssrn.com/abstract=2800121

Wei, T., & Simko, V. (2021). *R package 'corrplot': Visualization of a correlation matrix* [(Version 0.92)]. https://github.com/taiyun/corrplot

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2022). *Readxl: Read excel files* [R package version 1.4.2]. https://CRAN.R-project.org/package=readxl

Williams, M. D., Rana, N. P., & Dwivedi, Y. K. (2015). The unified theory of acceptance and use of technology (UTAUT): A literature review. *Journal of Enterprise Information Management*, *28*(3), 443–488. https://doi.org/10.1108/JEIM-09-2014-0088