Supplemental Material to "Testing the Usability of the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template"

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.

Plain language summary (ENG)

Why did we do the study?

Preregistration means that researchers document and publish a plan of their study that outlines their expectations, reasoning, and statistical methods before they have seen the data. This approach increases objectivity because researchers are not yet influenced by the results at this stage. It also promotes transparency because deviations from the plan can be tracked. Preregistration templates, such as the "Psychological Research Preregistration-Quantitative (PRP-QUANT) Template", assist in the preparation of preregistrations by listing all aspects that should be considered. Researchers can complete the form to create their preregistrations.
As preregistration in psychology is voluntary, it is important to make it as user-friendly and accessible as possible. Therefore, our studies aimed to assess the usability and acceptance of the PRP-QUANT Template among psychological researchers. In addition, we investigated researchers' willingness to use the template in the future and the factors influencing this decision.

What did we do?

We conducted two online studies to inspect the template's usability and researchers' intention to use it.

- In study 1, participants filled in the template items, provided feedback about their impression of the template, and suggested possible improvements. We also asked them about their expectations regarding the benefits and ease of use when using the template, as well as their perceptions of other researchers' opinions of them using preregistration.
- In study 2, we surveyed authors and reviewers who had used this template about their experiences when using it.

What did we find?

We found that the template was easy to use, and users were happy with it. Many participants planned to continue using it in the future, mainly because they believed it to be helpful.

What now?

Our results show that the template is already being well received. However, we also found some areas where it can be improved. These findings can now be used to revise the template.

Allgemeinverständliche Kurzzusammenfassung (DE)

Warum haben wir die Studie durchgeführt?

„Prä-Registrierung“ bedeutet, dass Forschende einen Plan ihrer Studie schreiben und veröffentlichen, bevor sie die Daten gesehen haben, in dem sie ihre Erwartungen, Überlegungen und statistischen Methoden beschreiben. Das erhöht die Objektivität, da die Forschenden in dieser Phase noch nicht von den Ergebnissen beeinflusst sind. Es fördert außerdem die Transparenz, da Abweichungen vom Plan nachvollzogen werden können.
Prä-Registrierungs-Vorlagen wie die „Psychological Research Preregistration-Quantitative (PRP-QUANT)-Vorlage“ helfen bei der Vorbereitung von Prä-Registrierungen, indem sie alle Aspekte auflisten, die dabei berücksichtigt werden sollten. Die Forschenden können das Formular ausfüllen, um ihre Prä-Registrierung zu erstellen.
Da die Prä-Registrierung in der psychologischen Forschung freiwillig ist, ist es wichtig, sie möglichst benutzerfreundlich zu gestalten. Daher war es das Ziel unserer Studien, die Benutzerfreundlichkeit und Akzeptanz der PRP-QUANT-Vorlage bei psychologischen Forschenden zu untersuchen. Außerdem interessierte uns, ob die Forschenden die Vorlage in Zukunft nutzen wollen, und welche Faktoren diese Entscheidung beeinflussen.

Was haben wir gemacht?

Wir haben zwei Online-Studien durchgeführt, um die Benutzerfreundlichkeit der Vorlage und die Absicht der Forschenden, sie zu verwenden, zu untersuchen.

- In Studie 1 füllten die Studienteilnehmer*innen die Vorlage aus, gaben Feedback und schlugen mögliche Verbesserungen vor. Wir fragten sie außerdem, wie sie den Nutzen der Vorlage einschätzten, ob sie glaubten, dass diese einfach zu nutzen sei, und was andere Forschende ihrer Meinung nach darüber denken würden, dass sie prä-registrieren.
- In Studie 2 befragten wir Autor*innen und Gutachter*innen, die diese Vorlage verwendet hatten, zu ihren Erfahrungen damit.

Was haben wir herausgefunden?

Wir fanden heraus, dass die Vorlage einfach zu verwenden war und die Forschenden damit zufrieden waren. Viele Forschende planten, sie auch in Zukunft zu verwenden, vor allem, weil sie sie für hilfreich hielten.

Und jetzt?

Unsere Ergebnisse zeigen, dass die PRP-QUANT-Vorlage bereits gut angenommen wird. Wir haben jedoch auch einige Bereiche gefunden, in denen sie verbessert werden kann. Diese Erkenntnisse können nun genutzt werden, um die Vorlage zu überarbeiten.

Table S1

Research areas

Research area	Overall (answered at least one template item) *N* = 88	Answered at least one overall evaluation item *N* = 63	UTAUT sample *N* = 60
Experimental/cognitive psychology	28.41 (25)	30.16 (19)	31.67 (19)
Educational psychology	19.32 (17)	20.63 (13)	20 (12)
Social psychology	19.32 (17)	15.87 (10)	15 (9)
Other (e.g., health, HCI, engineering, environmental, traffic)	15.91 (14)	14.29 (9)	13.33 (8)
Developmental psychology	11.36 (10)	14.29 (9)	15 (9)
Organizational psychology	11.36 (10)	11.11 (7)	11.67 (7)
Clinical psychology	10.23 (9)	9.52 (6)	10 (6)
Research methods	9.09 (8)	11.11 (7)	10 (6)
Neuroscience/neuropsychology	7.95 (7)	7.94 (5)	8.33 (5)
General psychology	6.82 (6)	9.52 (6)	10 (6)
Differential psychology	3.41 (3)	4.76 (3)	5 (3)
Did not respond	3.41 (3)	3.17 (2)	3.33 (2)

Note. The following parameters are displayed: Percentage (Frequency). Multiple options could be selected.

Table S2

Overview of template response fit

Item	*Mean*	*SD*	*Median*	*Range*
T8	2.91	0.42	3	2
I1	2.33	0.62	2	2
I2	2.18	0.4	2	1
I3	2.8	0.63	3	2
I4	2.43	0.79	3	2
M2	2.74	0.61	3	2
M3	2.1	0.72	2	2
M4	2.58	0.61	3	2
M5	2.44	0.73	3	2
M6	2.69	0.6	3	2
M7	2.56	0.51	3	1
M8	2.07	0.59	2	2
M10	2.45	0.69	3	2
M11	2.57	0.53	3	1
M12	2.67	0.5	3	1
M13	1.71	0.95	1	2
M14	2.29	0.95	3	2
AP1	2.53	0.62	3	2
AP2	2.5	0.52	2.5	1
AP3	2.46	0.88	3	2
AP4	2.43	0.65	2.5	2
AP5	1.58	0.51	2	1
AP6	2.29	0.83	2.5	2
AP7	1.85	0.99	1	2
AP8	2.3	0.67	2	2
Overall	2.38			

Note. Parameters were calculated based on the coding of participants' responses' fit to the respective template item (scale: 1 = *fits poorly*, 2 = *fits moderately*, 3 = *fits well*). The options 0 = *not applicable* and -9 = *nonsense answer* were excluded from this calculation. For the overall mean, the mean of means was calculated.

Table S3

Participants' answers to the item "What would you add, change, or remove about the item?"

Item	Comments
T8 *N* = 12	- *Add*: Default answer when there is no conflict (e.g., "All authors declare that they have no conflicts of interest." (*n* = 3) - *Remove*: Item (*n* = 2) - *Add/change/remove*: Nothing (*n* = 6)
T10 *N* = 14	- *Add:* Multiple options for different sub-datasets (*n* = 2) - *Add:* Ask for repository (as it is indicated in the item title) (*n* = 1) - *Add:* Link to repository location (if already existing, e.g., in a project) (*n* = 1) - *Add:* Possibility to justify (*n* = 1) - *Change:* Indicate if all or anonymized data will be shared (*n* = 1) - *Change:* Differentiate between raw or accumulated data (*n* = 1) - *Remove:* Repository (*n* = 1) - *Add/change/remove*: Nothing (*n* = 4)
T11 *N* = 11	- *Add:* Clarify and ask for code (e.g., experimental or analysis, which software) (*n* = 5) - *Add:* Make applicable for coding for qualitative analyses (*n* = 1) - *Add:* Confirmation that code is commented to increase reusability (*n* = 1) - *Add:* Adherence to standards (*n* = 1) - *Change:* "data" to "code" in dropdown menu (*n* = 1) - *Add/change/remove*: Nothing (*n* = 3)
T12 *N* = 11	- *Add:* Explanation of what such a document is and list with examples of existing standard lab practices documents (*n* = 3) - *Add:* Option to indicate there is no such document (*n* = 1) - *Remove:* Item (*n* = 2) - *Add/change/remove*: Nothing (*n* = 2)
I1 *N* = 9	- *Add*: Suggestions how this should be done (*n* = 2) - *Change*: Distinguish between theory and study-specific predictions (*n* = 1) - *Remove*: Item (*n* = 1) - *Add/change/remove*: Nothing (*n* = 4)
I2 *N* = 9	- *Change*: Better distinguishment between this item and the other items of the *introduction* section (especially I1), and between theoretical and practical research questions (*n* = 4) - *Change*: Clarify description (*n* = 2) - *Add/change/remove*: Nothing (*n* = 2)
I3 *N* = 10	- *Add:* Option for indicating that there are no hypotheses (*n* = 1) - *Add:* Prompt to indicate which of the hypotheses are mutually exclusive or not (*n* = 1) - *Change:* Use the plural "hypotheses" in the title and description (*n* = 1) - *Change:* It should not always be required that hypotheses are numbered, as this is not always necessary (*n* = 1) - *Change:* "Hypothesis/aims" instead of only "Hypothesis" for studies that rather provide aims than hypotheses (e.g., meta-analyses) (*n* = 1) - *Change:* Integrate this item with I2 (*n* = 1) - *Add/change/remove*: Nothing (*n* = 2)
I4 *N* = 7	- *Add*: More information about what to register here (*n* = 1) - *Add*: Prompt for indication how multiple comparisons will be handled (*n* = 1) - *Change*: Clarify if one should include the research questions for purely exploratory studies here or in I2 (*n* = 1) - *Remove*: Item (*n* = 1) - *Add/change/remove*: Nothing (*n* = 2)
M1 *N* = 25	- *Add*: Add options "Registration prior to piloting", "Registration after piloting", and "Registration prior to complete data collection" (*n* = 2)

Item	Comments
	- *Add*: Item to described other studies in the same project or future studies (*n* = 1) - *Add*: More information when the item should be answered (*n* = 1) - *Change*: Clearer description and distinguishment between options (e.g., especially considering the terms "creation", "access", and "observation") (*n* = 5) - *Change*: Make it suitable for other research types (e.g., meta-analyses, field studies) (*n* = 2) - *Change*: "Other" option, add example "If study 1 is already conducted, but not published, and study 2 is now being preregistered.", and make it an open text input field (*n* = 2) - *Change*: "Collection" instead of "creation" (*n* = 1) - *Add/change/remove*: Nothing (*n* = 10)
M2 *N* = 13	- *Add*: Examples of what it means to have knowledge of the data or how to remain unaware of results when using pre-existing data (*n* = 3) - *Add*: Possibility to answer for each of multiple studies individually (*n* = 1) - *Add*: Category where raw data has been collected and is now used by various researchers, or where some analyses were already conducted, but now in a new paper, different aspects are analyzed (*n* = 1) - *Add*: Question about whether existing data will be compared to newly collected data (*n* = 1) - *Add*: Filter question "Will you re-analyze existing data in this study" and only if yes, inquire in more detail (*n* = 1) - *Change*: Break down into multiple sub-items, if pre-existing data will be analyzed, to make answering it easier (*n* = 1) - *Change*: Make it suitable for meta-analyses or systematic reviews (*n* = 1) - *Add/change/remove*: Nothing (*n* = 3)
M3 *N* = 10	- *Add*: Aspect "practical reasons" (*n* = 1) - *Add*: Information that references may be given (*n* = 1) - *Add*: Clarify all aspects and add exampled (*n* = 1) - *Change*: Make it easier to answer, e.g., make question shorter (e.g., "Provide (planned) sample sizes and power analysis.") and instead add sub-questions like "If you are doing a multilevel analysis, also provide X, if you use a sequential design, also provide Y") (*n* = 2) - *Change*: Place the part for sequential designs on a separate paragraph for better reading (*n* = 1) - *Remove*: "(e.g., t-tests and correlations, but even descriptively such as with histograms)" (*n* = 1) - *Remove*: Item (*n* = 1)
M4 *N* = 5	- *Add*: "e.g.," in d) because not all points are required in all studies (*n* = 1) - *Add*: Examples of stratification sampling methods (*n* = 1) - *Change*: Make it more suitable for other designs (e.g., field studies) (*n* = 1) - *Change*: Make item shorter (*n* = 1)
M5 *N* = 8	- *Change*: Clarify drop-out (e.g., if experiment crashes, is this participant treated as drop-out) and what to fill in if it is not a longitudinal study (*n* = 2) - *Change*: Link this item to the inclusion/exclusion criteria of M4 (*n* = 1) - *Change*: Rather focus on how attrition is treated (listwise deletion, multiple imputation, unbalanced data allowed?) (*n* = 1) - *Change*: Ask for procedures to handle incomplete data (*n* = 1)
M6 *N* = 8	- *Add*: Filter, since this is oftentimes not applicable (*n* = 3) - *Add*: Term "blinding" (*n* = 2) - *Add*: Option for non-group comparison work (*n* = 1) - *Change:* Clarify what to insert if this does not apply (*n* = 1)
M7 *N* = 9	- *Add*: Examples (*n* = 2) - *Add*: Interrater reliability training (*n* = 1) - *Add*: Check boxes (*n* = 1) - *Change*: Link to other items (e.g., could be combined with M6 or AP3) (*n* = 3)
M8 *N* = 6	- *Add*: Options "(a) intention to treat, (b) per protocol analysis, (c) intention to treat and per protocol analysis" (*n* = 1) - *Change*: Integrate in AP1–3 (*n* = 2) - *Change*: Make "deletion" the default and only answer if one deviates (*n* = 1) - *Change*: FIML is not an imputation procedure (*n* = 1)

Item		Comments
	-	*Remove*: Item (n = 1)
M10 N = 3	-	*Add*: Possibility to answer for each of multiple studies individually (n = 1)
	-	*Change*: Make it clearer that the list in parentheses are examples, not categories one has to choose from (n = 1)
	-	*Change*: Make it more suitable for other sub-disciplines (e.g., include terminology used in personality psychology) (n = 1)
M11 N = 4	-	*Add*: Information on which studies this does not apply to (n = 1)
	-	*Change*: Split this item into multiple items (n = 1)
	-	*Change*: Make it more suitable for other designs (e.g., population representative survey studies) (n = 1)
	-	*Add/change/remove*: Nothing (n = 1)
M12 N = 6	-	*Add*: Allow possibility to specify hypotheses in terms of model comparison (n = 1)
	-	*Add*: Make it easier to connect I3, M12, and AP6, e.g., by displaying them in a table (n = 1)
	-	*Change*: "Operationally defined" as it might be too technical (n = 1)
	-	*Change*: Clarify what to write if multiple hypotheses correspond to different variables (and vice versa) (n = 1)
	-	*Change*: Clarify what "consistent with statistical analysis plans" means (n = 1)
	-	*Change*: Clarify if factors need to be repeated (already defined in other items) (n = 1)
	-	*Change:* Make it more suitable for other designs (e.g., survey research) (n = 1)
	-	*Change:* Make item shorter (n = 1)
	-	*Add/change/remove*: Nothing (n = 1)
M13 N = 5	-	*Add:* Flexible order of items to tailor template to specific study designs (n = 1)
	-	*Change*: Link to other items (redundant to M12, items assessing hypotheses) (n = 2)
	-	*Remove*: Item (as it imposes a large effort without reducing many researcher degrees of freedom) (n = 1)
	-	*Add/change/remove*: Nothing (n = 2)
M14 N = 4	-	*Add*: Inquire how participants are recruited and (if applicable) translation of measures (n = 2)
	-	*Add/change/remove*: Nothing (n = 1)
AP1 N = 10	-	*Add:* Inquire if participants will be excluded list/case wise from the analysis (n = 1)
	-	*Change*: Make it clearer that examples are examples, not a list of all things that need to be answered (n = 1)
	-	*Change*: Ask for cut-off values if the exclusion relates to data (either incorrect or missing) (n = 1)
	-	*Change:* Split into separate items (n = 1)
	-	*Add/change/remove*: Nothing (n = 2)
AP2 N = 11	-	*Add*: "reliability score" and "incorrectly solved trials" as examples (n = 2)
	-	*Add*: Offer recommendations that help researchers select a strategy, together with an "else, please specify" option (n = 1)
	-	*Change*: Clarify wording (n = 1)
	-	*Change*: Make it more suitable for other study designs (i.e., it might be difficult to find statistical outliers in large-scale studies) (n = 1)
	-	*Add/change/remove*: Nothing (n = 1)
AP3 N = 5	-	*Add*: Possibility to submit additional files (e.g., syntax) (n = 1)
	-	*Change*: Combine with M7 (what choices you make about data cleaning/screening/processing are typically made in combination, not at two different times as is implied by Methods and Analysis Plan placement) (n = 1)
	-	*Change*: Make it optional (n = 1)
AP4 N = 3	-	*Change*: Split this item between traditional analyses and measurement development studies (n = 1)
	-	*Change*: Inquire about how one will establish if items are removed from scales and in what cases excluded items will be analyzed separately or skipped completely (n = 1)
	-	*Add/change/remove*: Nothing (n = 1)
AP5 N = 7	-	*Change:* Make it optional, as it is not always relevant (n = 2)
	-	*Change*: Ask for effect sizes elsewhere, not with the descriptive statistics (n = 1)
	-	*Remove*: Item (n = 3)

Item	Comments
AP6 *N* = 4	- *Add*: Mention NHST versus Bayesian approach to analysis, and plans to use confidence intervals for point estimates and/or effect sizes as well as planned visualization (*n* = 1) - *Add*: Possibility to submit additional files (e.g., syntax) (*n* = 1) - *Change*: Let the author decide which information is most relevant (*n* = 1) - *Remove*: Tests for statistical violations, as it is too much to report (*n* = 1)
AP7 *N* = 6	- *Add:* Asking for software is more important, since this oftentimes decides about what effect sizes are used (*n* = 1) - *Change*: Effect sizes are inquired about in multiple items; they should not be covered here again (redundant) (*n* = 3) - *Change*: Make it optional (*n* = 1) - *Remove*: Not necessary to describe how effect sizes are computed (*n* = 1) - *Add/change/remove*: Nothing (*n* = 1)
AP8 *N* = 7	- *Change*: Clarify that other exploratory analyses are allowed afterwards (*n* = 2) - *Remove*: Item, as it is not relevant for preregistering (*n* = 3)

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

As described in the article, in addition to the individual items of the PRP-QUANT Template, a number of web probing items were presented, asking, for example, why participants had chosen an answer, whether they had correctly understood the items' concepts, how they perceived the link between items, whether they could distinguish items from each other, and which concepts were unclear. The results for these are presented below.

When asked to explain why they selected a specific answer, participants' explanations fit the intentions of the items in most of the cases (T10 "Data accessibility statement and planned repository": 93.75% of 16 responses; T11 "Optional: Code availability": 94.12% of 17 responses). Explanations fit slightly less well for M1 "Time point of registration" (78.05% of 41 responses); however, this was not due to incorrect responses (as they accounted for only 2.44% of cases), but because participants expressed more general opinions about preregistration that did not match the web probing question (19.51%).

Understanding of complex terms was good for some items but could be improved for others. For example, I4 "Exploratory research questions" and AP4 "Reliability analysis" were understood well, that is, participants provided appropriate definitions of exploratory analyses (92.31% of 13 responses), explained the codes used in I4 correctly (100% of 13 responses), and provided appropriate cases in which AP4 should be answered (100% of 15 responses). Other items contained terms that could not be encompassed as clearly. For example, for item M2 "Use of pre-existing data", participants were prompted to provide examples of how it can be assured that someone is unaware of the data. Fifty percent of the 26 participants who answered the item could think of examples, while 23.08% could not think of any, and 26.92% gave more general

opinions or did not answer the question. Thus, it might be useful to include some examples in a new version of the PRP-QUANT Template. When asked for definitions of the term “sample sizes (or sample ranges) found at each level of multilevel data” (M3), answers also varied in terms of adequacy, that is, 65% of 20 participants indicated an adequate definition, while the remaining 35% did not provide a sufficient definition or indicated that they were not familiar enough with the term to respond. The item might benefit from emphasising more that multilevel data is a special case that is not relevant to everyone. In addition, participants indicated that power estimations are difficult for multilevel data, which might also be considered. For M11 “Randomization of participants and/or experimental materials”, participants were asked how many different types of randomisations are covered in this item. Of eight responses, 50% gave a correct answer, while the other half gave a general opinion which did not fit the question or answered incorrectly. Therefore, it might be helpful to highlight the different levels more strongly, possibly by numbering them. In addition, this item asked participants to describe their understanding of “constraints on randomization”. Here, 75% of the eight responses contained correct definitions, while the remainder were more general comments (e.g., that participants do not do this kind of research).

Participants found it rather beneficial when related items were referenced within items, for example, I3 “Hypothesis” in M12 “Measured variables, manipulated variables, covariates” (*Mean* = 2.18, *Median* = 3, *SD* = 0.98, *IQR* = 2, *range* = 2, on a scale from -3 = *not beneficial* to 3 = *very beneficial*). Additionally, participants tended to feel that it would be easy to combine the information from such related items throughout the preregistration (*Mean* = -0.27, *Median* = -1, *SD* = 2.05, *IQR* = 3, *range* = 6, on a scale from -3 = *very easy* to 3 = *very difficult*, for the items I3, M12, and AP6).

However, distinguishing between similar items was not a trivial task for the participants. For example, when asked to differentiate M7 “Data cleaning and screening” and AP3 “Data preprocessing”, only around 40% of responses provided a clear differentiation (condition 2: 44.44% of 18 responses; condition 4: 40% of 15 responses). In most of the other cases, participants either indicated that they thought the items were completely different or suggested that the content of the other item should be explicitly excluded because both items were so similar, or that both items should be combined. Differentiating was also not entirely clear for M4 “Participant recruitment, selection, and compensation” and AP1 “Criteria for post-data collection exclusion of participants”. While 64.71% of the 17 responses included an appropriate differentiation, the remaining responses did not or indicated insecurity.

Lastly, perceived understanding and unclear terms were inquired for M3 “Sample size, power and precision”, M4 “Participant recruitment, selection, and compensation”, M8 “How will missing data be handled?”, and AP4 “Reliability analysis”. For M3, participants on average indicated that they understood the item well (*Mean* = 0.79, *Median* = 0.5, *SD* = 1.84, *IQR* = 2.25, *range* = 6, on a scale from -3 = *very poorly understood* to 3 = *very well understood*), but still reported uncertainty concerning the sample sizes (e.g., what is the “relevant” sample size, 14.29% of 14 responses); the power analyses (e.g., what to include if you did not do any power analysis, 14.29%); the term “fixed-N designs” (14.29%); “multilevel data” (7.14%); or they indicated insecurity with the overall item (14.29%). For M4, participants also indicated overall good understanding (*Mean* = 2, *Median* = 2, *SD* = 1.09, *IQR* = 2, *range* = 3), but were unsure concerning the terms “stratification sampling” (30% of 10 responses) and “planned participant characteristics” (10%); and regarding the difference between “b) selection and inclusion/exclusion criteria” and “d) planned participant characteristics” (20%). Additionally, it

was not clear if each point should be answered individually or not (10%). For M8, overall understanding was again good (*Mean* = 1.45, *Median* = 2, *SD* = 1.57, *IQR* = 2, *range* = 5), but the points “c) test of missingness”, (28.57% of 7 responses), “d) imputation procedures” (14.29%), and “e) intention to treat analysis and per protocol analysis” (28.57%) were unclear. Participants were unsure if “missing data” was equal to deleted data (14.29%) and suggested to write out abbreviations. One participant indicated they did not know most of the terms. Lastly, concerning the reliability analyses in AP4, 12 of 15 participants knew all types of reliability mentioned in the item, but three participants did not know the term “Cronbach’s omega”.