

Preregistration

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

Lisa Spitzer¹, Stefanie Mueller¹, Michael Bosnjak¹

¹Leibniz Institute for Psychology (ZPID)

26/02/2021

Keywords: meta-research, open science, preregistration, template, usability

Abstract

Currently, no standard procedure is available for creating preregistrations. Thus, a task force, consisting of members of prominent psychological societies and institutions, was formed to create a consensus template for quantitative research in psychology, the Psychological Research Preregistration-Quantitative (PRP-QUANT) Template (available at <http://dx.doi.org/10.23668/psycharchives.4463>; Preregistration Task Force, 2020). This new preregistration template will be tested with the usability study described in this preregistration.

The descriptive insights into the usability of the template are the main objective for this study. Furthermore, the intention to use the template in the future will be tested by implementing the unified theory of acceptance and use of technology. Based on this theory, it is expected that the behavioral 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).

Participants will be recruited by contacting members of the divisions of the German Psychological Society (DGPs), the American Psychological Association (APA), and the British Psychological Society (BPS). Furthermore, the study will be advertised on social media and via other psychology mailing lists. Only psychological researchers will be included.

The study will be conducted as an online questionnaire. Participants will see the whole template and will be asked to give an overall evaluation. Furthermore, they will be requested to fill out a subset of template items and web probing questions, as well as a questionnaire measuring the variables of the unified theory of acceptance and use of technology.

Theoretical background

Preregistration of studies comprises documenting and publicly sharing planned study details before a study is administered (possibly with an embargo, Nosek, Ebersole, DeHaven & Mellor, 2018). This enables the researcher team as well as others to assess what was planned before the study was conducted, which reduces questionable research practices like selective reporting of significant results and allows a clear differentiation between confirmatory and exploratory analyses (e.g., see Wagenmakers, Wetzels, Borsboom, van der Maas & Kievit, 2012).

Overall, preregistration seems effective in reducing questionable research practices and thereby reducing the rate of false-positive findings in published scientific literature (e.g., see Kaplan & Irvin, 2015, for an example of this in another field). Consequently, it becomes more and more accepted, although it is still not the standard for psychological researchers to preregister their studies (Hardwicke et al., 2020).

Preregistration templates support researchers in creating preregistrations by providing a form which includes the most important information to preregister, that can be filled out. For psychological research, various templates are available that have been created independently from each other, but share some similarities (e.g., Brandt et al., 2014; van't Veer & Giner-Sorolla, 2016). Yet, preregistration still lacks a standard that is supported by psychological associations or journals.

To close this gap, a task force, consisting of members of the German Psychological Society (DGPs), the American Psychological Association (APA), the British Psychological Society (BPS), the Open Science Framework (osf), and the Leibniz Institute for Psychology (ZPID) was formed to create the “Psychological Research Preregistration-Quantitative (PRP-QUANT) Template” (available at <http://dx.doi.org/10.23668/psycharchives.4463>; Preregistration Task Force, 2020) which is tailored to the field of psychology and could be used to create a preregistration to be submitted to a repository and/or a stage 1 manuscript to be submitted as a Registered Report (i.e., a manuscript consisting of an introduction, methods, and data analysis section, which is written and reviewed before data collection and which can gain an in-principle acceptance by a journal, that is the commitment to publish the study after study administration provided that it is conducted and analyzed as outlined).

This template will be evaluated in the study described in this preregistration. Specifically, a usability study will be conducted online that is based on the four aspects of usability as defined by Shackel (2009) and the unified theory of acceptance and use of technology (UTAUT, Venkatesh, Morris, Davis & Davis, 2003; Venkatesh, Thong & Xu, 2016). Furthermore, it will include various web probing elements (see Behr, Meitinger, Braun & Kaczmirek, 2017).

Research questions

As mentioned above, the research questions of this usability study are defined based on the different aspects of usability according to Shackel (2009, questions 1 to 4) and the UTAUT (Venkatesh et al., 2003; Venkatesh et al., 2016, question 5):

1. *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?
2. *Flexibility*: Does the template capture the main points across sub-disciplines, as indicated by researchers of different sub-disciplines?
3. *Effectiveness*: Are the items specific enough (i.e. are df minimized)? Are items answered as expected (i.e., is the information requested in the item actually provided by researchers in response to it)?
4. *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?
5. *Formation of behavioral intention*:
 - a. What are researchers' performance and effort expectations? How are social influences and facilitating conditions perceived?
 - b. Can performance expectations, effort expectations, and social influence (moderated by age, experience, and/or voluntariness, see model below) predict the intention to use the template in the future?

Hypotheses

The usability estimation of the overall template as well as of individual items is the main focus of this study. Yet, there are some predictions regarding influences on participants' intention to use the template in the future. Based on the UTAUT (for an overview of the model's postulations, see Venkatesh et al, 2013), we expect that:

1. The behavioral intention to use our template is influenced by
 - a. Performance expectancy (moderated by age)
 - b. Effort expectancy (moderated by age and experience)
 - c. And social influence (moderated by age, experience, and voluntariness of use)

In contrast to the original model, we do not assume a moderating effect of gender in our context. More detailed information about the specific predictions (e.g., the specific moderating effects) is given in Figure 1.

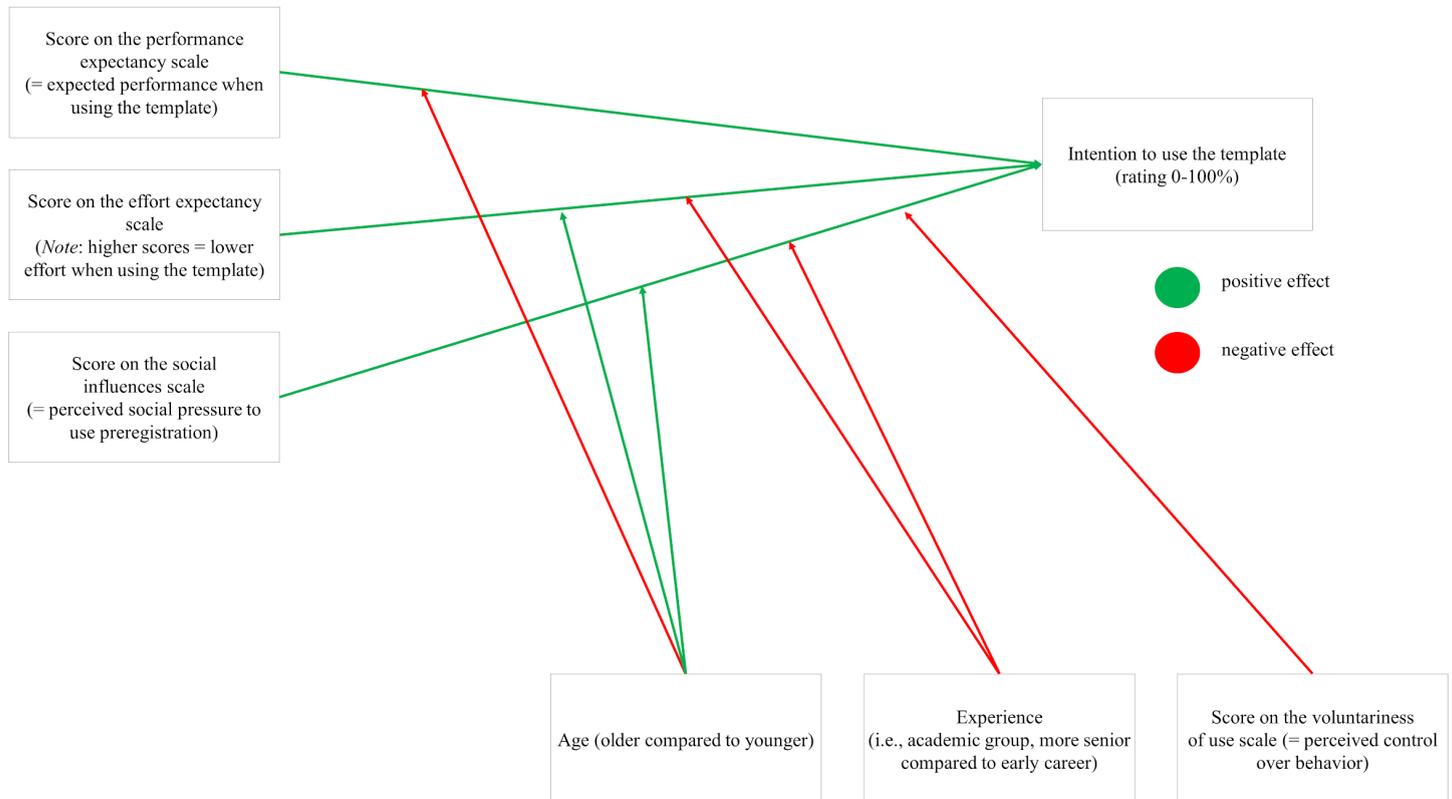


Figure 1. Hypotheses. These hypotheses are based on the UTAUT (see Venkatesh et al., 2003, p. 447). Green arrows indicate a positive effect, red arrows indicate a negative effect of a predictor for the intention to use the template. Moderating variables are displayed below.

Sampling & Data collection procedure

Since our study addresses psychological researchers, participants will be recruited by sharing the study link with members of research-oriented APA (<https://www.apadivisions.org/>), BPS (<https://www.bps.org.uk/members/networks>), and DGPs (<https://www.dgps.de/index.php?id=48&L=1>) divisions (for the recruitment of APA members, all researchers labeled as “quantitative” will be invited). Each of the potential participants will be contacted twice (i.e., with an initial invitation plus a reminder after one week). Furthermore, the survey will be advertised via social media (e.g., Twitter).

The study will be in English and primarily English (recruitment via the APA and BPS divisions) and German (recruitment via the DGPs sections) speaking participants will be

recruited. Thus, the sample will consist primarily of western, English or German speaking participants.

Inclusion/exclusion criteria for our sample are:

- Only data from psychological researchers will be collected. This will be achieved by using a filter question at the beginning of the study for which responding will be mandatory. Specifically, participants will be asked if they are researchers and if their research falls within the scope of psychology. All participants that indicate that they are not researchers or that their focus is not on psychological research, will be redirected to an exit page and thereby excluded before beginning the main study.
- Participants may be members of any psychological subdiscipline. We will ask participants to select their division/research focus from a list of all APA, BPS, or DGPs divisions and common research areas. Participants may also indicate other, not listed psychological research fields by open text input.
- For our descriptive statistics, we will consider all participants that were included based on the criteria above, thus also participants that have not completed all pages or answered all items of our online questionnaire. Yet, we will exclude empty data files and will not consider data that obviously indicate a non-answer (e.g., random key pressing like “geuiwbeuiwg”).
- For the analyses regarding the UTAUT, we will only include participants that answered all UTAUT items of our questionnaire (23 questions) plus the item that inquires about the intention to use the template in the future (since it will serve as dependent variable in our model) and the items asking about the moderating variables (age, academic group).

For the test of the UTAUT model, a power analysis was conducted. As summarized by Venkatesh et al. (2016), the UTAUT explains 77% of the variance in behavioral intention to use a technology in longitudinal field studies. We do expect a rather large effect size, however, we want to be able to also detect smaller effects. Specifically, $R^2=0.25$ was chosen as a moderate effect size of practical significance (see Ferguson, 2009) and thus, we specified a power analysis to determine the sample size in accordance.

F tests - Linear multiple regression: Fixed model, R^2 deviation from zero

Analysis: A priori: Compute required sample size

Input: Effect size $f^2 = 0.3333333$

α err prob = 0.05

Power ($1-\beta$ err prob) = 0.95

Number of predictors = 12

Output: Noncentrality parameter $\lambda = 29.6666637$

Critical F = 1.8818760

Numerator df = 12

Denominator df = 76

Total sample size = 89

Actual power = 0.9509406

This analysis indicated that in our model that includes behavioral intention as dependent variable, and overall 12 predictors and interaction effects (1. performance expectancy, 2. effort expectancy, 3. social influence, 4. performance expectancy x age, 5. effort expectancy x age, 6. effort expectancy x experience, 7. social influence x age, 8. social influence x experience, and 9. social influence x voluntariness, 10. age, 11. experience, 12. voluntariness, a visual presentation of this model is available in Figure 1), we would need data of $N=89$ participants to be able to detect an effect of $R^2=0.25$ with $\alpha=\beta=.05$.¹ Additionally, individual predictors will also be tested, but the power analysis focuses on the overall model. This calculated sample size will be used as minimum sample size, but not as a stopping rule criterion for the data collection. The reason for this is that the descriptive usability assessment is still the main focus of our study, and here no specific sample size is required.

Overall this means that in a first step, as much data as possible will be collected over a timeframe of two weeks (after the last invitation has been sent to a division), with a reminder after one week. If after these two weeks, there are not at least $N=89$ participants who can be included in the UTAUT analyses (thus, who have answered the UTAUT questions, the intention item, and the moderating variable items), data collection will be extended for another two weeks. One month after the last division has been initially contacted, however, data collection will be stopped even if $N=89$ cannot be achieved (see Figure 2). Participants will not be compensated.

¹ Caveat: This sample size might be too small for any exploratory subgroup analyses, which will therefore only be conducted if more data is collected.

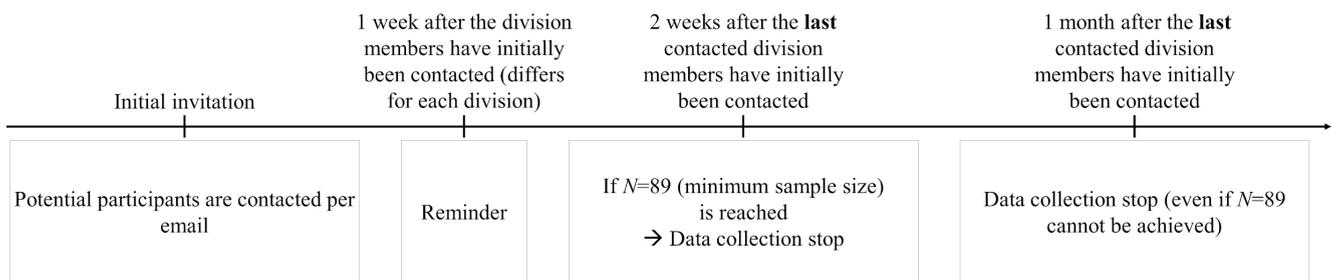


Figure 2. Recruitment procedure. Time schedule for data collection.

Procedure

The study will be applied as an online web probing survey. It will be created with the software SoSci Survey (Leiner, 2019) and supplied via www.soscisurvey.de. Participants will not be blinded, instead they will be informed about the study's objectives (i.e., the evaluation of the new template). Participation will require approximately 30-45 minutes.

The study can be divided into four sections.

1. *General questions* (e.g., socio-demographic questions, general usage of preregistration)
2. *Evaluation of individual items and web probing:*
 - Participants will be asked to fill out the items of the preregistration template, thinking about a study that they are currently planning or conducting
 - Furthermore, various web probing questions will be displayed (see Behr et al., 2017):
 - *Comprehension probing* (e.g. ask for meaning of terms, paraphrasing)
 - *General/elaborative probing* (e.g. ask for examples)
 - *Category probing* (e.g. ask to elaborate on the selected category)
3. *Overall evaluation of the template*
4. *UTAUT items*

The item "How many studies have you preregistered before?" (part of the *general questions*) will serve as a filter question, i.e., some of the items of the general questions and overall evaluation will only be shown to participants who have preregistered before (e.g., what preregistration template they usually prefer).

Importantly, for the second part of the study (*evaluation of individual items and web probing*), participants will be randomly assigned to one of four conditions. Based on their condition, they will fill out only those template items that focus on 1) title and introduction,

2) overall methods, sampling procedure, and data collection, 3) overall methods, conditions, and design, or 4) analysis plans. All participants will see the whole template, but based on their condition, they will only be asked to respond to a subset of items.

A set of web probing questions will be included, some of which will be used for all template items (rating of the perceived importance of the item; and an open question inquiring about what participants would add, change, or remove about the item). Other web probing items will be used specifically for individual template items. For example, participants will be asked to:

- Provide understanding of concepts
- Explain why they selected a specific answer
- Answer different rating questions
- Give examples
- Differentiate template items from related items
- Indicate what exactly they thought about while answering the item
- Indicate how well they understood what information is requested by the item

No items of the study will be mandatory besides the filter question at the beginning of the survey. However, when asking participants to describe the study which they think about while answering the template, as well as for the UTAUT questionnaire, subjects who do not answer them will be asked if they really do not want to answer, to make sure that gaps do not occur accidentally. Furthermore, for the template items (i.e., the responses to the overall task “Think of one of your own studies and fill out the fields as if you were preregistering this study.”) participants who do not respond will be redirected to an extra page, where they will be asked for specific reasons why they did not want to answer this template item. Here, they will have the possibility to indicate whether they think the item is optional, made a mistake, did not know what to answer, don’t like the item, or if the item does not fit their research. Additionally, other reasons can be described via open text input. This question itself will not be mandatory.

A list of all items included into the survey is shared as part of this preregistration (see Table 1 in the Appendix). This procedure has been approved by the ethics committee of Trier University.

Analysis plans

Data will be preprocessed by re-coding responses from multiple choice questions (originally: 1 = “not checked” and 2 = “checked”; new: 0 = “not checked” and 1 = “checked”) and turning single choice items into factors. Polarity of negatively-poled items will be reversed. All UTAUT items as well as some other items of the web probing and

overall evaluation will be re-coded from “1 - 7” to “-3 - +3”, yielding a middle category which has absolute meaning. Open text input will be coded as described below.

The analyses can be grouped based on the four sections of the study.

1) General questions

The socio-demographic items as well as the items inquiring about the general usage of preregistration will be evaluated (e.g., by computing means and distribution parameters like the distribution of participants that have or have not preregistered before, how many preregistrations participants have been created before, or the distribution of psychological sub-disciplines represented in the sample).

2) Evaluation of individual items and web probing

Next, the individual template items (as well as web probing questions) will be inspected to get an insight into how they were perceived. For all items, it will be rated if the given response to the template item matches what was requested in the item (1 = “fits poorly”, 2 = “fits moderately”, 3 = “fits well”). An a priori coding scheme will be used for this coding (see Table 2 in the Appendix). In order to further improve the coding scheme and to represent as many potential responses as possible, the template responses of 25% of subjects per condition of the study will be used. Specifically, 25% of datasets per condition will be randomly selected and coded after data collection, while the coding scheme will be revised in the process. Afterwards, the hereby improved coding scheme will be applied to the remaining datasets. The preliminary coding scheme is part of this preregistration, and the final coding scheme will be published with the study. The mean per item as well as the overall mean of these ratings will be computed.

Then, the web probing questions will be analyzed. As mentioned above, some web probing items will be included for all template items. These will be analyzed first. Particularly, mean and standard deviation will be calculated for the item “How do you perceive the importance of this item for preregistering your study?”. Furthermore, the open text input of the item “What would you add, change, or remove about the item?” will be evaluated for each template item separately by coding common themes: Responses will be shuffled, and coders will read the first 10% of these shuffled responses and identify themes (= things to add/change/remove) mentioned by the participants, which will then be transferred to new columns in a coding sheet. Then, it will be coded for all other responses if the theme was mentioned (= 1) or not (= 0). If new relevant topics appear to the coder that they have not coded before, these are added as categories as the coding continues, and coded later. Furthermore, these common themes will be classified into 1) things to add, 2) things to change, and 3) things to remove, for each template item.

This procedure will also be used to code other open questions of the survey (e.g., when asking participants what they thought about when answering this item, how they would discriminate an item from another similar item, or what terms they found unclear). However, when asked for definitions, explanations or examples, another coding will be implemented, i.e., for definitions, it will be coded if the term is correctly described in the response with 1 = “yes” and 0 = “no”, or for given examples it will be coded if the examples fit the requested term with 1 = “yes” and 0 = “no”.

Means and standard deviations will be computed for individual items across participants. Furthermore, distributions will be assessed (e.g., which percentages of participants selected each of the given options on an item). Selected descriptive statistics will be displayed as plots or frequency tables. The results will also be compared between participants of different psychological sub-disciplines to assess if the template is suitable for different psychological research areas.

3) Overall evaluation of the template

After this, the items concerning the overall evaluation of the template will be analyzed (e.g., how well the template was understood, how well it fits the participants’ research area, see Table 1 in the Appendix) by computing means and standard deviations. Additionally, distributions of answers will be evaluated, and open text inputs will again be coded to assess common themes. Here too, the results will be displayed in tables and plots wherever possible, and will be compared between different sub-disciplines.

4) UTAUT

To investigate if the overall UTAUT can account for the intention to use the preregistration template in the future, and which are the most important influences, the items concerning the UTAUT will be analyzed. For each participant, mean scores will be computed for all scales of the UTAUT questionnaire (i.e., performance expectancy, effort expectancy, social influence, facilitating conditions, voluntariness). As mentioned above, we expect that the behavioral intention to use our template in the future can be 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). The facilitating factors are not part of the model we want to test (since they influence the actual behavior, not the intention), yet we also measure this scale since it might give valuable insights into factors that might help foster the practice of preregistration (thus, this scale will be analyzed descriptively). Our predictions are visualized in Figure 1.

We will compute the mean and standard deviation for every scale across all participants. Furthermore, we will compute a moderated multiple regression model (a model which has been used frequently to test the UTAUT, see Williams, Rana & Dwivedi, 2015),

using the causal structure implied by the UTAUT. In this model, the variables will be included as follows:

- Dependent variable: Behavioral intention (answer to the item “How likely would you use the template in the future to create your preregistrations?”)
- Independent variables:
 - Score on the performance expectancy scale (moderated by age)
 - Score on the effort expectancy scale (moderated by age and experience, i.e., academic group)
 - Score on the social influence scale (moderated by age, experience, i.e., academic group, and the score on the voluntariness of use scale)

The expected effects are displayed in Figure 1. The significance of the overall model as well as of individual predictors and moderating effects will be evaluated based on $\alpha = .05$, and the magnitude of standardized regression weights will be interpreted. Since this is the only computed model, no alpha error correction needs to be done.

Other descriptive statistics might be added as exploratory analyses, for example some items might be reanalyzed separately for participants who have or have not preregistered before.

Existing data

Data collection has not begun.

Conflict of interest

We declare that there is no conflict of interest.

Implications

The results of this usability test will be relayed to the task force that created the template and used to improve the template. We furthermore want to use the results of the UTAUT analyses to promote the template in the best possible way based on evidence and to specifically address the aspects that most strongly influence the intention to use the template (e.g., performance and effort expectancy, or social influence).

The data obtained in this study will be made available in the public repository PsychArchives (<https://www.psycharchives.org/>) where the template is provided and changes to the template will be documented. The data will be shared as unprocessed as possible, yet any information

related to the specific studies described by participants will be omitted. These preprocessing steps will be documented and shared alongside the data set. Additionally, the analysis code will be published. Everything will be published under the license [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).

Planned schedule

Data collection is planned for the first quarter of 2021. Results will be analyzed until fall 2021 and we plan to submit the study for publication to a journal by the end of 2021.

References

- 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/10.15465/gesis-sg_en_023
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, 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>
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. <https://doi.org/10.1037/a0015808>
- Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M., & Ioannidis, J. Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014-2017). MetaArXiv [Preprint]. 2020 [posted 2020 Jan 02; cited 2020 May 26]. Advance online publication. <https://doi.org/10.31222/osf.io/9sz2y>
- 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>
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at <https://www.soscisurvey.de>

- 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>
- Preregistration Task Force. (2020). Preregistration Standards for Psychology - the Psychological Research Preregistration-Quantitative (aka PRP-QUANT) Template. ZPID (Leibniz Institute for Psychology). <https://doi.org/10.23668/PSYCHARCHIVES.4463>
- 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>
- van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration 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.L., & 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. Retrieved from <https://ssrn.com/abstract=2800121>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on psychological science*, *7*(6), 632–638. <https://doi.org/10.1177/1745691612463078>
- Williams, M.D., Rana, N.P. and Dwivedi, Y.K. (2015), "The unified theory of acceptance and use of technology (UTAUT): a literature review", *Journal of Enterprise Information Management*, Vol. 28 No. 3, pp. 443-488. <https://doi.org/10.1108/JEIM-09-2014-0088>