INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024
https://doi.org/10.1017/pds.2024.14

Replication studies in engineering design – a feasibility study

Jonas Rode [1], Ingo Jonuschies [1,✉], Sven Matthiesen [2] and Kilian Gericke [1]

[1] University of Rostock, Germany, [2] Karlsruhe Institute of Technology, Germany

✉ ingo.jonuschies@uni-rostock.de

Abstract

This paper examines the replicability of studies in design research triggered by the replication crisis in psychology. It highlights the importance of replicating studies to ensure the robustness of research results and examines whether the description in a publication is sufficient to replicate. Therefore, the publication of a reference study was analysed and a replication study was conducted. The design of the replication study appears similar to the reference study, but the results differ. Possible reasons for the differences and implications for replication studies are discussed.

Keywords: research methodologies and methods, design methods, publication bias, evaluation, reliability

1. Introduction

The identification of approaches and strategies of designers during problem solving is an important goal of design research. Such studies enable the development of new tools and methods, with the aim to support practitioners and to facilitate the training of the next generation of engineers and designers (Blessing and Chakrabarti, 2009; Ahmed *et al.*, 2021). However, academically developed methods tend to be notably complex, which can lead companies to develop their own tools rather than make use of design research (Wallace, 2011). Moreover, companies tend to adopt methods that are already established in other companies rather than using entirely new methods. Therefore, there is a need to develop and evaluate methods, which are more specific to the industrial sector (Gericke *et al.*, 2022). To attain high-quality evaluation, the study designs employed should fulfil the quality criteria of objectivity, reliability, and validity (Hussy *et al.*, 2013). This paper focuses on the significance of reliability, specifically the consistency and robustness of study designs in the context of design method evaluation.

1.1. Motivation

The importance of reliable study designs is shown by the ongoing crisis of confidence in psychology, which was triggered by numerous events in the early 2010s (Pashler and Wagenmakers, 2012).
One factor contributing to the crisis is the publishing of false-positive results due to flaws in study design or even the widespread use of questionable research methods (Ioannidis, 2005; John *et al.*, 2012) such as p-hacking and HARKing, as pointed out by Ulrich *et al.* (2016). Along the hierarchy of scientific research, an increased trend of publications with positive outcomes can be observed. Due to the high proportion of non-significant results that are not published, there is a distortion of truth resulting in the publication bias (Fanelli, 2010). In order to counterbalance the excessive number of false positive results that are published, it is necessary to publish true negative results (Nissen *et al.*, 2016). The reasons for

non-publication of non-significant data can be attributed to journals, authors, and readers (Ferguson and Heene, 2012; Curtis and Abernethy, 2015).

One important element of the current crisis of confidence in scientific research is the "replication crisis", characterised by the non-replicability of studies, which has led to a debate in various scientific disciplines about the reliability of results (Open Science Collaboration, 2015; Klein *et al.*, 2018; Klein *et al.*, 2022). The difficulty lies in the fact that replications are frequently deemed unoriginal, making them less likely to be conducted. More importantly, replication studies that are realised by different authors rarely confirm the findings of the original study (Makel *et al.*, 2012). Most of the replication studies that supported the original study were conducted within the same research group (Ryan and A Tipu, 2022). The non-replicability of studies can have several causes. Non-replicability of a result can indicate a falsification of the original study's findings or highlight previously unconsidered factors influencing the result (DFG, 2017). Furthermore, another possible cause of non-replicability of results can be due to incomplete documentation of the study, despite proper planning and execution. As a consequence, replication studies may not reflect the original design and thus influence the result.

The ongoing replication crisis is no longer confined to psychology, other fields have also acknowledged its relevance. These disciplines encompass the social sciences (Stroebe, 2016; Pridemore *et al.*, 2018), medicine (Errington *et al.*, 2014; Kelter, 2020), humanities (Bak, 2016; Peels and Bouter, 2018), neuroscience (Pavlov *et al.*, 2021) and various others. In the field of engineering, only computer science (Cockburn *et al.*, 2020; Rougier *et al.*, 2017) has addressed this issue so far.

As many studies in design research are based on or inspired by research methods and study designs from the social sciences, it cannot be excluded that there is an unrecognised replication crisis in design research. It is therefore necessary to analyse research practices in the design community and develop measures to improve them. When applying the findings from psychology to design research, it becomes evident that there is still a lack of essential knowledge on replicating studies robustly.

It is uncertain whether descriptions of study designs in design research are sufficient to ensure replicability. The work presented in this paper is motivated by the research question: Is it possible to replicate a study based on its publication? Therefore, this paper presents the results of an initial replication study.

2. Research approach

A common information situation between the original and the replication process is an essential key to replication success. One aim was to create a common information environment in the feasibility study in the form of the original publication and the stimuli. Furthermore, it is important to distinguish between exact and close replication (Chen *et al.*, 2021). An exact replication uses the same protocols, instructions and the same information conditions as the original study, only the replicators and subjects are different. The aim is to test whether the findings of the reference study are true and replicable. A close replication, on the other hand, deliberately uses different protocols. These are follow-up studies that can provide us with information on how general and robust the original results are. The study presented here aimed to create an exact replication and followed best practice (Chen *et al.*, 2021), as far as the common information environment allowed. We investigate the question of whether the common information situation in form of a publication and the stimuli are sufficient for an exact replication.

The replication study presented here refers to a study conducted by Matthiesen and Nelius (2018) which is named 'reference study' in the following. A basic condition for selection is the robustness of the reference study. According to Cash (Cash *et al.*, 2023), this would be the case if the study follows the criteria of the "systematic assessment framework". This is given in the reference study. The further selection of the reference study was based on the following criteria:

- It ought to be a recent study investigating a technical issue related to human behaviour in design.
- The research question seems stable over time, reducing the likelihood of failing to replicate because of environmental changes (Stroebe, 2016).
- The study investigates a hypothesis-based research question to enhance result comparability.

Moreover, choosing this study offered the possibility of requesting missing information from the authors if necessary. However, contact was limited to avoid bias of the replication.

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

2.1. Research question of the reference study

With a special focus on the identification and solution of problems, the reference study aims to investigate the analytical approach of designers during their analysis of a technical system. The reference study has two main components: a functional analysis aimed at gaining a comprehensive understanding of the technical system and its target function, and a synthesis-driven analysis designed to expand upon previous knowledge and evaluate the disparities between the target and actual functions to identify new approaches in the design process. The hypotheses examined in the reference study are:

H1: *„A structured verbalization of one's own understanding of a technical system helps to identify gaps and mistakes in the own analysis.“* (Matthiesen and Nelius, 2018)

H2: „*Verifying own assumptions about the function and behaviour of a technical system improves the quality of the analysis.*“ (Matthiesen and Nelius, 2018)

2.2. Methods of data acquisition

Eye Tracking

Many commonly used empirical data collection methods capture participants' explicit knowledge. However, their tacit and implicit knowledge is often either not recorded or not sufficiently recorded (Ahmed, 2007). Nevertheless, understanding this tacit and silent knowledge is essential to comprehend a designer's decision-making and thought processes. Therefore Eye tracking (ET) is used to gain access to these kinds of knowledge (Matthiesen *et al.*, 2013). The ET tracks the participants' gaze so that the source of new insights can be traced. This is mainly to support the evaluator (Matthiesen and Nelius, 2018).

Concurrent Think Aloud

Concurrent Think Aloud (CTA) is then used, so that participants verbalise their thoughts (Matthiesen and Nelius, 2018). This enables participants to express their cognitive processes and insights while simultaneously working on a task (Kelley et al., 2015). This realistic reflection of participants' cognitive processes can be achieved without altering their natural environment (Ericsson and Simon, 1993). This allows identification of patterns and strategies in problem solving (Kelley *et al.*, 2015). Moreover, CTA aims to reveal implicit and tacit knowledge (Matthiesen and Nelius, 2018).

Interview-like presentation of results

Further data is obtained from the presentation of the results. The participants are supposed to perceive the supervisor of the study as a "work colleague" and verbally communicate the findings and solution approaches they have gained. This method is similar to an interview and thus examines the explicit knowledge of the participants (Ahmed, 2007).

Questionnaire

Questionnaires are a popular research method for standardised quantitative data collection (Roopa and Rani, 2012). The aim is to achieve better comparability of the participants (Matthiesen and Nelius, 2018). Other studies also show that combining the methods used is useful for answering empirical questions in engineering (Du and MacDonald, 2014; Bi *et al.*, 2015).

3. Transfer of the study design

3.1. Procedure

The design of the replication study presented here is based on an analysis of the reference study. The objective is to achieve a high degree of consistency with the reference study design. During the analysis of the reference study, various forms of adaptations were necessary. These can be distinguished into implicit and explicit adaptations. Some details of the replication study had to be defined without

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

knowledge of the implementation in the reference study, as the corresponding details were not published. The decisions made in this way are based on assumptions and are referred to as implicit adaptations. Explicit adaptations were made deliberately, although the necessary information was available in the reference study, such as the change in sample size or hardware used. The sample size was decreased due to time constraints in the replication study. The hardware had to be adjusted based on the available equipment. It differs from the equipment used in the reference study but provides the same functionality.

The reference study is structured into three distinct sections (I: introduction and training task, II: task 1 functional analysis, III: task 2 synthesis-driven analysis) and follows a detailed timetable. The participants receive a laptop on which a presentation is prepared, accompanying the participants throughout the entire study to reduce the impact of the study supervisor and to ensure a high degree of objectivity (Matthiesen and Nelius, 2018). The replication followed this procedure, although slight implicit adaptations were made to the detailed schedule and presentation.

The introduction consisted of a welcome, a presentation of the used data collection methods, as well as the calibration of the eye tracking device and a training task on the use of CTA, as described in the reference study.

Subsequently, two tasks, which have an identical structure followed. After a brief reading period for the participants, the self-editing phase starts, which is restricted in time and involves working on the tasks with provided aids. Following the completion of the self-education phase, the participants reported their results to the study supervisor in the form of an interview-like presentation. Subsequently, the aids were removed and the participants completed a questionnaire regarding the recently completed task. The same procedure was used for the replication. However, implicit adaptations had to be made to the content of the questionnaires.

The first task (T1) consists of a functional analysis of a commercially available lawn sprinkler. The lawn sprinkler model and manufacturer have not been specified in detail. Alongside the original lawn sprinkler, the participants are given a manipulated version of the sprinkler which can be easily dismantled. The information provided in the reference study is not sufficient to determine whether the identical sprinkler could be implemented. Both the original and the manipulated sprinkler used in the study were available by contacting the authors, eliminating the need to make assumptions. Without this contact, an implicit adaptation would have been necessary. Furthermore, the participants have been provided with the basic function of the sprinkler. Some implicit adaptation had to be made concerning the level of detail and the way in which this information was provided. As in the reference study, the participants of the replication study were given 10 minutes to complete the task.

The second task (T2) involves synthesis-driven analysis. Participants were asked to redesign an existing system due to a failure of a part of a security mechanism. This task also required implicit adaptations to the supporting tools. However, unlike the previous task, the authors have only partially provided them. During the replication, the bolt-setting device and the broken part were provided as 3D-printed replicas. In addition, the participants were provided with high-resolution photographs. The 3D model and technical drawing remain identical to those utilised in the reference study. As with T1, certain implicit adaptations were necessary for the level of detail and quantity of information regarding the functionality of the device. Videos which were also used in the reference study were provided by the authors. Utilising the provided information, the nail detection system was identified as a crucial component of the bolt setting device, and thus the task was designed accordingly. The time limit of 20 minutes for T2 was not changed.

3.2. Participants

A total of 17 participants took part in the replication study. The participants were divided into a "less experienced" (novices) and "very experienced" (experts) group based on their experience, so that possible differences depending on the expertise can be revealed.

9 of the 17 participants were novices studying a technical subject mechanical engineering or comparable) and were at least in their fifth semester. The data of one participant cannot be used due to technical issues and is therefore not included in the evaluation. The other 8 participants have at least 2.5 years of professional experience in their field and were therefore assigned to the expert group. The composition of the participants of the reference study is compared to the replication in Table 1.

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

Table 1. Comparison of the participants of the reference study (left) and the replication (right)

Reference study	Replication study
14 novices 5th+ semester (∅ 6.9; SD: 1.9) M: 12, W:2 Mechanical engineering	9 novices 5th+ semester (∅ 9.3; SD: 0.94) Not collected Mechanical engineering or comparable
12 design experts 2.5+ yrs. work experience (∅ 12.4; SD: 8.5) All male	8 design experts 2.5+ yrs. work experience (∅ 14.2; SD: 7.4) Not collected
Σ 26	Σ 17

3.3. Data analysis

In order to successfully answer H1, nine sub-functions of the lawn sprinkler were defined in advance. It remains unclear which sub-functions are defined in the reference study, hence own sub-functions had to be predefined for the replication study. Initially, more than nine functions were identified during this process. Given the focus on the alternating mechanism, the functions most closely associated with this were selected for analysis. The study analysed the respondents' statements to determine whether the predefined functions were mentioned accurately (correct), not mentioned (gap) or misunderstood (error). Furthermore, to address the hypothesis, it was crucial to identify additional insights and corrections provided during the presentation of the results.

In order to address H2, in the reference study the audio recordings were transcribed and then analysed in order to detect verifications of the participants own assumptions. However, in the replication, transcription was omitted and replaced by a direct analysis of the audio data using a specific evaluation software (iMotions). Markers were set according to the same conditions as for transcription in order to detect verifications.

Verification refers to the re-examination of an assumption already made by the participants. The ET-recordings assist the researcher and a second coder in conducting a review. Discrepancies must be settled through consensus. No review was included in the replication. To test the hypothesis, it is necessary to measure the quality of the analysis in addition to the amount of verifications. This is achieved by adding the correctly identified functions from task 1 (max. nine) to the scores obtained in the questionnaire for task 2 (max. nine). A total of 0 to 18 points can be attained. Furthermore, the verifications have been examined in more detail and further subdivided. Qualitative verifications and evaluations have been distinguished. Qualitative verifications such as *"whether there is an electric motor or not; whether two parts touch or not"* are mentioned exemplary in the publication. *"There is a force, yet it lacks the strength to cause damage to the part"* serves as an example for an evaluation. Besides, a distinction is made between verifications which provide evidence supporting or disproving an assumption made.

4. Results

During the replication study, it was found that participants needed a lot of time to develop a functional understanding when working on T2. It was discovered that this time should not have been included in the initially planned 20 minutes, thus resulting in the participants having less time for actual analysis than stated in the reference study. Additionally, some participants found the tools provided for T2 to be insufficient. The researchers desired to get the genuine bolt setter and damaged part at hand for the study; however, the original bolt setter was not available.

13 of the 16 participants (81.25%) of the replication study named the majority of the functions correctly, indicating that they developed a good understanding of the technical system. This result is comparable to that of the reference study, which was 75%. The overall outcome of task 1 is illustrated in Figure 2. Despite the presentation of the findings, two participants still had errors in their understanding of the function. During the presentation of the findings, just one of the participants was able to identify a mistake. In addition, three of the participants identified functions that they had not mentioned during the self-editing process and were thus able to close gaps in their functional understanding. Conducting a Pearson correlation test the data shows a weak correlation between the correct mentioned functions

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

and the new insights during verbalisation, but the result is not significant (r=0.38; p=0.147; n=16). Like in the reference study, H1 is not supported and is therefore rejected.

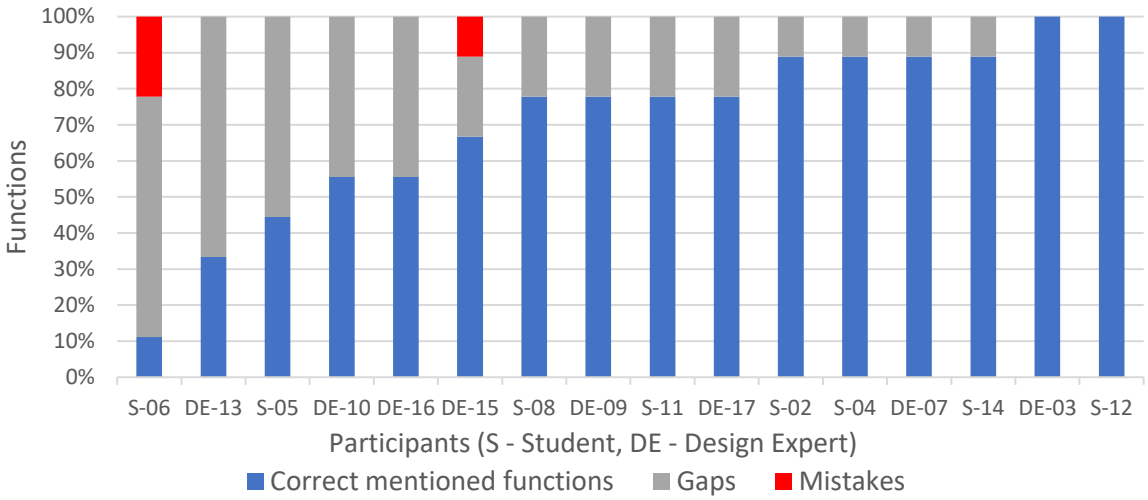

Figure 1. Evaluation of the participants' functional understanding of task 1 (replication)

In the second task, the participants were asked to develop a solution for the bolt setting device so that the nail detection would last the desired amount of time. Out of 16 participants, only five acknowledged that the direction of breakage is 90° to the functional plane and pinpointed recoil as a potential cause, enabling them to develop solutions according to this cause. Participant S-14 identified only the "high loads on the unit" and did not further investigate the system for the exact cause. Among the proposed solutions, DE-03 and DE-07 were especially noteworthy. As the bolt setting device is powered electrically, DE-03 recommended incorporating an electronic sensor and discarding the mechanical nail detection. DE-07 proposed decreasing the mass of the freely oscillating surface of the nail detector by making a hole in it. Furthermore, DE-07 recommends to apply additional clamping force to the component.

Hypothesis 2 examines the correlation between analysis success and the amount of verifications. Figure 3 visually represents both variables for each participant separated in novices and experts as well as participants of the reference and replication study. The group of the replication showed a significantly larger number of verifications compared to the referred group.

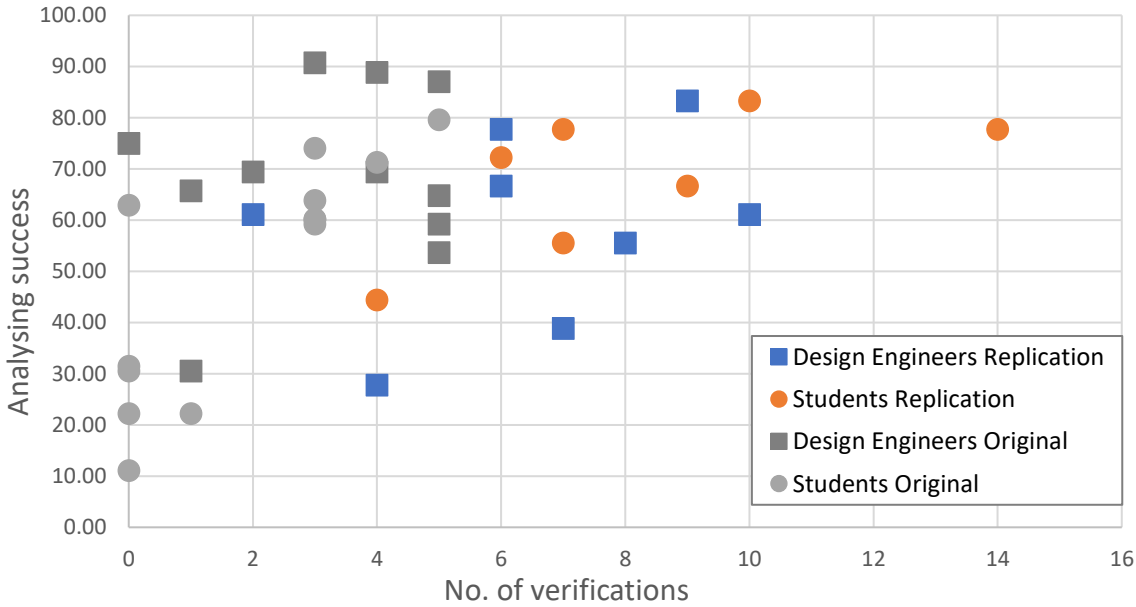

Figure 2. Analysis success and the amount of verifications per participant

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

In the reference study, some participants did not verify any assumptions, whereas in the replication study each participant verified assumptions at least twice. The difference is even more obvious when looking at the maximum number of verifications for each study. The reference study identified a maximum of 5 verifications per participant, compared to 14 verifications in the replication study. In the reference study a significant correlation was identified between analysis success and the amount of verifications ($r=0.63$, $p=0.001$, $n=24$); while also present, this correlation is not significant in the replication study ($r=0.48$, $p=0.06$, $n=16$).

The replication study revealed the identification of 115 verifications, in contrast to the 64 verifications of the reference study. Table 2 presents the classification and distribution of the recorded verifications. When differentiating between qualitative verifications and evaluations, it is noticeable that the number of qualitative verifications is significantly higher. Moreover, it could be observed that verifications that approve previously made assumptions outweighed verifications that disprove them. Similar to the reference study, the majority of verifications did not have a direct impact on understanding. The majority of improvements were achieved by finding evidence that disproved prior assumptions.

Table 2. Classification of verifications

Change of participants' understanding	Total	Qualitative	Evaluations	Approval	Disproval
Improved	25	16	9	13	12
Indifferent	87	64	23	83	3
Worse	3	3	0	4	0
		83	32	100	15
Sum	115	115		115	

The reference study also investigates the correlation between analysis success and the participants' evaluations. The five participants with the highest and lowest score in analysis were selected for the investigation. The results indicate a strong and significant correlation between the measured variables ($r=0.85$, $p=0.002$, $n=10$). However, the same analysis in the replication did not show a significant correlation ($r=0.34$, $p=0.337$, $n=10$).

5. Discussion

This research aims to acquire knowledge on how to conduct a replication based on a given publication. The objective is to determine the feasibility of a replication study based on the information provided in a typical research publication followed by assessing the comparability of the results.

Numerous elements of the design of the reference study used in this investigation could be adopted without making assumptions. It can be hypothesised that the quantity of implied adaptations increases with the level of detail.

The reference study and the replication study seem to be comparable. Therefore, it can be affirmed that conducting a replication study based on a publication is possible if implicit adaptations are made. However, the necessary adaptations, especially implicit adaptations that were necessary may have a concerning effect on the results and have to treated with utterly care.

One example of implicit adaptations that may have had a strong effect on the result is: Both the first and second task were awarded nine points as a basis for analysis success. The predefined functions, which form the basis for assessment of the analysis success, were not described sufficiently in the publication of the reference study. Therefore, the functions had to be redefined under several implicit adaptations for the replication study.

Even though the results of the replication study are different it is not clear whether the deviating results are caused by the implicit adaptations, whether the replication study deviates unknowingly from the reference study or whether the findings of the reference study have been simply falsified.

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

The observed deviations and the ambiguity of the causes call for further studies aimed at an improvement of the replication culture in this community. The replication study presented here clearly shows that different study designs for analysing replicability of study designs are required. Using a published description of a reference study as the basis will not allow to exclude an incomplete description, thus, inaccurate replication, as the main cause for deviating results.
While the study provided important insights for future attempts to replicate empirical studies based on publications, this study has some limitations. The replication study and its results are based on a single coder; no inter-encoder reliability test was performed so far. Moreover, this study was primarily aimed at developing a better understanding of the process of replication not at testing the validity of the results of the reference study. Unlike other replication studies, the stimuli used in the reference study were available. However, the accessibility of the authors during the replication should not be taken as granted.

6. Conclusion

The crisis within psychology demonstrates that studies often cannot be replicated. Nonetheless, it is crucial to know whether a result cannot be verified because the results differ or because the publication of reference study design is too imprecise to investigate the same effect. The objective of this investigation was to ascertain whether it is possible to conduct a replication based on a referred study's publication. Our findings revealed that while a similar study design can be replicated, the study design and data analysis were not adequately described in the publication, so the results cannot be considered robust results of a replication study.
Even though a study is documented in accordance with the usual standards and quality criteria, there is no guarantee that the study can be replicated without adaptation. As far as the information available from the reference study allowed, we followed best practice for replication studies. We analysed the reference study in detail in order to understand the theory of the original research question, to be able to replicate the research topics, to ensure the same subject group composition, to be able to replicate the statistical analysis and to be able to design the procedure and environmental conditions in the same way (Chen *et al.*, 2021). However, the complexity of studies in engineering design makes replication studies very challenging - usually much more complex than in other disciplines. Thus, for exact replication studies in engineering design, it is (besides detailed documentations of study designs) particularly important to use standardised tests, scales and measurement methods.

References

Ahmed, A., Hurwitz, D., Gestson, S. and Brown, S. (2021), "Differences between Professionals and Students in Their Visual Attention on Multiple Representation Types While Solving an Open-Ended Engineering Design Problem", *Journal of Civil Engineering Education*, Vol. 147 No. 3.

Ahmed, S. (2007), "Empirical research in engineering practice", *J. of Design Research*, Vol. 6 No. 3, p. 359.

Bak, T.H. (2016), "Bilingualism and Executive Function", *Linguistic Approaches to Bilingualism*, Vol. 6 No. 5, pp. 699–717.

Bi, Y., Shergadwala, M., Reid, T. and Panchal, J.H. (2015), "Understanding the Utilization of Information Stimuli in Design Decision Making Using Eye Gaze Data", *Proceedings of the ASME Design Engineering Technical Conference*.

Blessing, L.T. and Chakrabarti, A. (Eds.) (2009), *DRM, a Design Research Methodology*, Springer London, London.

Cash, P., Daalhuizen, J. and Hekkert, P. (2023), "Evaluating the efficacy and effectiveness of design methods: A systematic review and assessment framework", *Design Studies*, Vol. 88, p. 101204.

Chen, R., Chen, Y. and Riyanto, Y.E. (2021), "Best practices in replication: a case study of common information in coordination games", *Experimental Economics*, Vol. 24 No. 1, pp. 2–30.

Cockburn, A., Dragicevic, P., Besançon, L. and Gutwin, C. (2020), "Threats of a replication crisis in empirical computer science", *Communications of the ACM*, Vol. 63 No. 8, pp. 70–79.

Curtis, M.J. and Abernethy, D.R. (2015), "Replication - why we need to publish our findings", *Pharmacology research & perspectives*, Vol. 3 No. 4, e00164.

DFGDFG (2017), "Replizierbarkeit von Forschungsergebnissen - Eine Replizierbarkeit von Forschungsergebnissen. Eine Stellungnahme der Deutschen Forschungsgemeinschaft", available at: https://www.dfg.de/foerderung/info_wissenschaft/2017/info_wissenschaft_17_18/ (accessed 9 October 2023).

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

Du, P. and MacDonald, E.F. (2014), "Eye-Tracking Data Predict Importance of Product Features and Saliency of Size Change", *Journal of Mechanical Design*, Vol. 136 No. 8.
Ericsson, K.A. and Simon, H.A.Ericsson, K. Anders; Simon, Herbert A. (1993), *Protocol analysis: Verbal reports as data*, Revised edition, [publisher not identified], [Place of publication not identified].
Errington, T.M., et al. (2014), "An open investigation of the reproducibility of cancer biology research", *eLife*, Vol. 3.
Fanelli, D. (2010), ""Positive" results increase down the Hierarchy of the Sciences", *PloS one*, Vol. 5 No. 4, e10068.
Ferguson, C.J. and Heene, M. (2012), "A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null", *Perspectives on psychological science a journal of the Association for Psychological Science*, Vol. 7 No. 6, pp. 555–561.
Gericke, K., Eckert, C. and Stacey, M. (2022), "Elements of a design method – a basis for describing and evaluating design methods", *Design Science*, Vol. 8.
Hussy, W., Schreier, M. and Echterhoff, G.Hussy, Walter; Schreier, Margrit; Echterhoff, Gerald (2013), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*, Springer Berlin Heidelberg, Berlin, Heidelberg.
Ioannidis, J.P.A. (2005), "Why most published research findings are false", *PLoS medicine*, Vol. 2 No. 8, e124.
John, L.K., Loewenstein, G. and Prelec, D. (2012), "Measuring the prevalence of questionable research practices with incentives for truth telling", *Psychological science*, Vol. 23 No. 5, pp. 524–532.
Kelley, T.R., Capobianco, B.M. and Kaluf, K.J. (2015), "Concurrent think-aloud protocols to assess elementary design students", *International Journal of Technology and Design Education*, Vol. 25 No. 4, pp. 521–540.
Kelter, R. (2020), "Analysis of Bayesian posterior significance and effect size indices for the two-sample t-test to support reproducible medical research", *BMC medical research methodology*, Vol. 20 No. 1, p. 88.
Klein, R.A., et al. (2018), "Many Labs 2: Investigating Variation in Replicability Across Samples and Settings", *Advances in Methods and Practices in Psychological Science*, Vol. 1 No. 4, pp. 443–490.
Klein, R.A., et al. (2022), "Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement", *Collabra: Psychology*, Vol. 8 No. 1, p. 35271.
Makel, M.C., Plucker, J.A. and Hegarty, B. (2012), "Replications in Psychology Research: How Often Do They Really Occur?", *Perspectives on psychological science a journal of the Association for Psychological Science*, Vol. 7 No. 6, pp. 537–542.
Matthiesen, S., Meboldt, M., Ruckpaul, A. and Mussgnug, M. (2013), "Eye Tracking, a method for engineering design research on engineers's behavior while analyzing technical systems", *ICED 13*.
Matthiesen, S. and Nelius, T. (2018), "Managing Assumptions during Analysis - Study on successful Approaches of Design Engineers", paper presented at NordDesign 2018.
Nissen, S.B., Magidson, T., Gross, K. and Bergstrom, C.T. (2016), "Publication bias and the canonization of false facts", *eLife*, Vol. 5.
Open Science Collaboration (2015), "Estimating the reproducibility of psychological science", *Science (New York, N.Y.)*, Vol. 349 No. 6251, aac4716.
Pashler, H. and Wagenmakers, E.-J. (2012), "Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?", *Perspectives on psychological science a journal of the Association for Psychological Science*, Vol. 7 No. 6, pp. 528–530.
Pavlov, Y.G., et al. (2021), "#EEGManyLabs: Investigating the replicability of influential EEG experiments", *Cortex; a journal devoted to the study of the nervous system and behavior*, Vol. 144, pp. 213–229.
Peels, R. and Bouter, L. (2018), "The possibility and desirability of replication in the humanities", *Palgrave Communications*, Vol. 4 No. 1.
Pridemore, W.A., Makel, M.C. and Plucker, J.A. (2018), "Replication in Criminology and the Social Sciences", *Annual Review of Criminology*, Vol. 1 No. 1, pp. 19–38.
Roopa, S. and Rani, M.S. (2012), "Questionnaire Designing for a Survey", *The Journal of Indian Orthodontic Society*, Vol. 46, pp. 273–277.
Rougier, N.P., et al. (2017), "Sustainable computational science: the ReScience initiative", *PeerJ. Computer science*, Vol. 3, e142.
Ryan, J.C. and A Tipu, S.A. (2022), "Business and management research: Low instances of replication studies and a lack of author independence in replications", *Research Policy*, Vol. 51 No. 1, p. 104408.
Stroebe, W. (2016), "Are most published social psychological findings false?", *Journal of Experimental Social Psychology*, Vol. 66, pp. 134–144.
Ulrich, R., et al. (2016), "Inflation von falsch-positiven Befunden in der psychologischen Forschung", *Psychologische Rundschau*, Vol. 67 No. 3, pp. 163–174.
Wallace, K. (2011), "Transferring Design Methods into Practice", in Birkhofer, H. (Ed.), *The future of design methodology*, Springer, London, New York, pp. 239–248.

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press

https://doi.org/10.1017/pds.2024.14 Published online by Cambridge University Press