

How to Teach Open Science Principles in the Undergraduate Curriculum - The Hagen  
Cumulative Science Project

Marc Jekel<sup>1,\*</sup>

Susann Fiedler<sup>2,\*</sup>

Ramona Allstadt Torras<sup>3,4</sup>

Dorothee Mischkowski<sup>1</sup>

Angela Dorrough<sup>1</sup>

Andreas Glöckner<sup>1,2</sup>

<sup>1</sup> University of Cologne

<sup>2</sup> Max Planck Institute for Collective Goods

<sup>3</sup> FernUniversität in Hagen

<sup>4</sup> University of Bonn

\* Both authors contributed equally

Author Note

Ramona Allstadt Torras is now at Department of Psychology, University of Cologne.  
Correspondence concerning this article should be addressed to Marc Jekel, Social Cognition  
Center Cologne, University of Cologne, Richard-Strauss-Straße 2, 50931 Köln,  
Germany, E-mail: [marc.jekel@uni-koeln.de](mailto:marc.jekel@uni-koeln.de).

### **Abstract**

The *Hagen Cumulative Science Project* is a large-scale replication project based on students' thesis work. In the project, we aim to (i) teach students to conduct the entire research process for conducting a replication according to open science standards and (ii) contribute to cumulative science by increasing the number of direct replications. We describe the procedural steps of the project from choosing suitable replication studies to guiding students through the process of conducting a replication, and processing results in a meta-analysis. Based on the experience of more than eighty replications, we summarize how such a project can be implemented. We present practical solutions that have been shown to be successful as well as discuss typical obstacles and how they can be solved. We argue that replication projects are beneficial for all groups involved: Students benefit by being guided through a highly structured protocol and making actual contributions to science. Instructors benefit by using time-resources effectively for cumulative science and fulfilling teaching obligations in a meaningful way. The scientific community benefits from the resulting greater number of replications and teaching state-of-the-art methodology. We encourage the use of student thesis-based replication projects for thesis work in academic bachelor and master curricula.

*Keywords:* Open science, Hagen Cumulative Science Project, replication, teaching

## How to Teach Open Science Principles in the Undergraduate Curriculum - The Hagen Cumulative Science Project

Replications are not just an important cornerstone of empirical research (Asendorpf et al., 2013; Simons, 2014), but also an effective tool for learning how to conduct methodologically sound empirical research. In a hands-on fashion, replicators learn about research methods, study designs, statistical analyses, the scientific process, and professional communication skills. Furthermore, they are inspired to reflect on more general values such as transparency and openness in science. In the spirit of recent proposals for new initiatives facilitating the training of students and also contributions to cumulative science (Brandt et al., 2014; Frank & Saxe, 2012; Grahe et al., 2012), we launched the *Hagen Cumulative Science Project* (HCSP) in October 2015—a large scale teaching project aimed at replicating current experimental studies in bachelor and master students' thesis work. By replication we mean that we repeated published studies with sufficient sample sizes and statistical power to determine whether “the [target] finding can be obtained with other random samples drawn from a multidimensional space that captures the most important facets of the research design” (Asendorpf et al., 2013, p. 109). More specifically, we conducted *direct* replications (Simons, 2014) that varied only in terms of the sample of participants (i.e., German-speaking students), time of data collection (i.e., between 2015 and 2019), place of data collection (i.e., web-based) and to a minimal extent study materials (e.g., materials translated into German) and procedures (e.g., smaller monetary incentives in some studies after consulting with the original authors) in comparison to the original studies.

We specifically focused on direct replications of studies published in the journal *Judgment and Decision Making (JDM)* for several reasons: 1) the journal includes topics and methods highly relevant to the expertise of the academic chair supervising the replication projects (chair of cognitive psychology: judgment, decision making, and action), 2) *JDM* is an open access journal providing data sets and supplemental material on the website for each article to easily reproduce the results with the original data (description see below) and 3) *JDM* topics are sufficiently diverse in content and methods to allow more general claims than in extremely narrow journals (e.g., social judgment, morality, interindividual differences, cooperation, heuristics, risky choice). Based on the experience from more than eighty completed student replications in HCSP, we share insights concerning the implementation of replication protocols for students with little prior experience in conducting experimental work. We also discuss potential challenges in the process and possible solutions.

In some universities, bachelor-theses are the final element of an undergraduate degree. Other universities may include optional honors theses that are similar in format. In psychology, thesis projects typically target a research question utilizing different tools of investigation (e.g., writing a literature review, conducting a correlational or experimental study, or using secondary data or analyses). The time frame for conducting the respective research is commonly short (e.g., twelve weeks in the curriculum of the University of Hagen for full-time students) and is stated in most examination regulations to equate to approximately 300-360 working hours.<sup>1</sup> Within this time frame, students who work empirically are typically asked to (1) begin with a literature review in order to develop their research question, (2) develop a research design, (3) prepare the necessary materials for the empirical study, (4) collect data from participants, (5) analyze the data, and (6) write up their thesis. The teaching goal of the thesis is to familiarize students with practical scientific work. Their skills for independently conducting research should be developed and tested. These skills should include the ability to evaluate research questions critically, develop ideas about the type of evidence necessary to answer the research question, and conduct the respective experiment (Manduca, 1997). In addition to these core elements of scientific research, recent evidence concerning the lack of transparency and

---

<sup>1</sup> average over all undergraduate psychology programs in Germany

openness of research (Wicherts et al., 2006) as well as the prevalence of statistical reporting errors in the past (Nuijten et al., 2016) also calls for focusing more strongly on teaching the transparent and comprehensive reporting of research.

Students' comprehension of research practices can be fostered through transfer --- that is, the direct application of learned concepts in an authentic research context. A growing body of students today may learn about the importance of replicability of empirical work in introductory classes in psychology. Experiencing what replicability means in the context of an individual experiment in their thesis would then allow students to transfer this knowledge to a practical problem. This should help students learn about and reflect on the importance of replicability. Arguably, replications are an efficient tool for motivating students to think like future researchers and extend their methodological skills. Furthermore, replications allow students to tackle and solve many typical problems that they are faced with when conducting thesis work.

### **Feeling overwhelmed by generating an interesting and new research question**

When developing their first research question, students oftentimes find it difficult to obtain a comprehensive overview of the relevant research and methods due to the large amount of published papers in a given field. This potential overload of information may often lead to negative feelings of being overwhelmed. As a result, some students may develop research questions that have already been answered but that have been overlooked by the student's incomplete literature review. With a pre-determined study design and statistical analyses of the original article as well as highly structured guidelines on how to proceed, conducting replications contributes to a balanced process between working independently and receiving orientation to successfully conduct a meaningful empirical study. Students can focus on learning skills to conduct a specific study and develop the methodological and statistical expertise needed (Frank & Saxe, 2012). Whether students are able to grasp psychological theory can still be assessed because students are asked to describe the motivation for the research.

**Focusing only on (overly) simple methods and statistics**

Using replications motivates students to not think about research designs in the framework of the statistical procedure with which they feel most comfortable, but rather the appropriate framework for the research question at hand. By re-analyzing the original data under supervision, students will be introduced to statistical tests that could go beyond the methods they learned in class. Although this may be challenging for students at first glance, it ultimately fosters a deeper understanding of statistical tests and statistical modeling in general. The re-analysis also builds up the necessary confidence and expertise for the analysis of the replication data. Additionally, reproducing original analyses as well as replicating the experiment feels like “detective work”, which can be exciting for students (Frank & Saxe, 2012; Janz, 2016).

**Lack of orientation and structure**

Reading scientific articles from the perspective of a scientist planning on replicating the presented work instead of a remote reader and consumer of the research helps students to gain insights into the decision processes of the original authors (Janz, 2016). Furthermore, it allows them to reflect on their own methodological choices more systematically. Experiencing the scientific procedure as an observer and evaluator of the original work as well as an active researcher who is in the process of running the replication teaches the value of reproducible routines as well as the importance of open and transparent documentation (King, 1995; Frank & Saxe, 2012). These insights help students develop a professional routine of making their work accessible to readers with little prior knowledge—a skill valued not only in academia but also in many jobs in industry. Furthermore, the replication process itself allows for a structured schedule with consecutive steps. A replication project thereby lends more orientation and structure, which enables students to manage the process of writing a thesis more effectively.

**Low sense of purpose and responsibility: Limited scientific thinking**

Approximately 90% of student projects are never published (Perlman & McCann, 2005)—they typically do not satisfy the high quality standards in the research field. This is problematic because students might anticipate from the start that their work will never be read, discussed,

or built upon. Thus, they may feel that the sole purpose and goal of their work is to receive their degree. Such conditions exclude students from feeling responsible and being part of the collaborative effort of conducting scientific work. Thus, it may also reduce the tendency for deeper, responsible, and more (self-) critical scientific thinking. In contrast, the experience of responsibility, having a real impact and also receiving the merits of contributing to the body of evidence, can be assumed to change the experience of scientific work substantially. The replication initiative focuses on teaching students to appreciate research and the scientific method, develop their methodological and experimental skills, and motivate critical thinking by doing meaningful work. If the replication attempt adheres to scientific standards, students' theses might even potentially result in their first scientific contribution in the form of a publication.

#### **Misconceiving scientific value: It is not (all) about finding statistically significant results**

Given the strong focus on statistically significant results (i.e., "positive results") in past research and the resulting replicability problems of empirical science (Francis, 2013), one major task of scientific education today is to safeguard the next generation of scientists against this bias. Conducting replications will teach students not to focus on positive results or the evidence of a single study. Rather, it educates students about the value of cumulative science as well as the evaluation of solid methods irrespective of study results (Cetkovic-Cvrlje et al., 2013; Wagge et al., 2018). Additionally, integrating replications into the standard curriculum raises awareness for the importance of replicability and its role in the scientific process (Höffler, 2013; Carsey, 2014).

#### **Procedural steps of the HCSP**

To support students in the process of independently conducting replication studies, we developed a structured teaching schedule to realize HCSP. The teaching schedule for HCSP consists of assignments that build upon each other incrementally (Table 1). We decided to apply a schedule for the following reasons: A schedule with concrete sequential sub-goals helps to break down the thesis project into manageable steps. Sub-goals can further guide students'

behavior at each step towards accomplishing the standards of sound scientific work. Finally, it allows instructors to closely monitor students' progress, intervene if necessary, provide effective feedback (Hattie & Timperley, 2007; Sadler, 1998), and use students' level of performance at each step for evaluating the thesis work.

During the process, students internalize principles of Open Science through practical application. After preparations by advisors (steps 1a and 1b, Table 1), students develop and apply their methodological skills with a focus on replicability by re-analyzing the original data (step 2), conducting a power-analysis (step 3), collecting, adjusting (e.g., translating materials in a different language) or re-creating the materials necessary for a replication (step 4), and professionally communicating with the original authors. Students extend their methodological skills concerning transparency and pre-registration by documenting and publishing study materials and an analysis plan of their targeted replication on the Open Science Framework (step 5). Finally, after collecting data from the university's subject pool (step 6), students learn how to document data and analysis scripts (step 7) as well as summarize and discuss their own results in light of the original evidence (step 8). In the theory section of their theses, students critically reflect on arguments made within the replicability debate (step 9). Within this section, they also evaluate the replicability of their target effect by elaborating on the original effect size, type of effect, sample of participants, and so forth. In addition to these educational goals, a secondary aim of HCSP is to increase the number of direct replications in the field and analyze replication rates and their moderators using meta-analysis (step 10). Furthermore, the firsthand experience enables us to develop tools and assistance for similar projects to lower the threshold for future replication efforts. In the following, we will highlight some aspects listed in Table 1 and provide further explanations.

In step 3, students determine the sample size for the study. Specifically, they calculate the number of participants required to achieve a specific level of statistical power of (e.g.) 80% for detecting an expected effect size. When doing so, they must take into account the study design and properties of the statistical test of the replication study. Based on our experience, however, how the targeted effect size is defined in a power analysis is not a trivial matter, since various possibilities to do so exist. Students may use the observed effect size from the original

study as an estimator of the population effect size in cases where original studies rely on sufficiently large sample sizes. Students may alternatively target at a higher statistical power of (e.g.) 95% or above to compensate for loss of power due to potentially inflated effect sizes in the literature (Simonsohn, 2015). Students may also define effect sizes according to alternative methods such as equivalence testing (Lakens, 2017) or at least rely on useful heuristics (e.g., 2.5 times the original sample size; Simonsohn, 2015) when deciding how many participants to collect. If the necessary sample size is not feasible, students may also consider increasing the alpha level (and pre-register this change from the conventional  $p < .05$  level in OSF) in a compromise power analysis (Faul, Erdfelder, Lang, & Buchner, 2007; Gigerenzer, Krauss, & Vitouch, 2004).

In step 5, students prepare a pre-registration of their replication study. Students use a standard form adapted from the Reproducibility Project (Open Science Collaboration, 2015; cf. Brandt et al., 2014) that can be downloaded from [osf.io/tr6fb](http://osf.io/tr6fb). In the pre-data report, students briefly describe the original research, indicate their central effect (i.e., we followed the protocol of the Reproducibility Project by targeting the central effect in the final study in an article), report a power analysis, and describe their materials, methods, statistical procedure, etc. (see template).

We also created a standardized website for each replication project at OSF, which can be forked from [osf.io/zdskx](http://osf.io/zdskx).<sup>2</sup> The template website consists of folders where students can upload their pre-data report, study materials (e.g., experimental software, questionnaires, etc.), data, and (re-)analyses of the original and replication data. In step 8, students document their results in a post-data report. The post-data report is identical to the pre-data report with an addendum including reports of the statistical analysis and a brief discussion.

Additionally, students upload their raw data, a data sheet explaining the meaning of the variables (see also Arslan, 2019), and the analysis script. We encourage students to use open source software for statistical analyses (e.g., PSPP, JASP, or R) and simple formats (i.e., txt and csv files) for all uploaded files. We also suggest that files follow a common format in terms of

---

<sup>2</sup> Forking a project creates a copy of the projects' tree of folders. Each student can fork the template to have her own copy of the folder structure to start her work (see [http://help.osf.io/m/links\\_forks/l/835698-fork-a-project](http://help.osf.io/m/links_forks/l/835698-fork-a-project) for additional explanations).

content to facilitate the final meta-analysis (step 10). That is, all files ideally consist of a header including the contact address of the student, the reference of the replicated study, and a link to the OSF website of the project. Ideally, analysis scripts begin by loading the raw data, end with printing the inferential statistics of the target effect, and are extensively commented (i.e., each processing step is commented and the version of the software used for the analysis is documented below the header; see also Rouder et al., 2019 for best practices).

For many students, running a replication study is only their first or second experience (e.g., lab practical) with conducting research. Therefore, questions regarding data quality must be considered. To ensure that the replications result in a sound contribution to the research field, instructors must take several measures. For instance, they must ensure that students only replicate feasible studies (step 1a, Table 1). Instructors also need to provide feedback at critical steps of the replication. Instructors may comment and check with students on pre-registration reports to avoid errors in the design and analysis stages of the project before the original authors are informed about the replication (see [osf.io/szb9a/](https://osf.io/szb9a/) for a template) and data is collected. Instructors may also provide feedback on the post-data report to avoid incorrect application and inferences from statistical analyses. We agree with comparable projects reporting that student replications follow a higher degree of transparency and a close-knit net of quality checks than the original work (Wagge et al., 2018) and thus may also result in similar or better data quality. We plan to publish all projects of HCSP in OSF to provide examples and prototypes for instructors and students. We are currently evaluating all projects carefully and rate the quality of each replication attempt concerning how well (often for pragmatic reasons) the quality criteria for a replication could be fulfilled. We plan to publish those evaluations as a report prominently in each project. This might also help other instructors who plan similar projects to assess examples of good replication attempts and detect target studies for which further replications might be worthwhile.

Finally, it should be mentioned that replications are not only a valuable means with which to conduct bachelor theses. Student conducted replication studies can also be extended to be sufficiently complex and demanding to be applicable for master theses as well. We adapted the format to students acquiring a master's degree as follows. In addition to the

replication of the target effect, master students independently derive from theory a moderator of the target effect that they then test in the replication study. To be able to generate this novel moderator hypothesis, students must consider the target effect and the theoretical background more intensively and at a broader and / or deeper theoretical level. This extension allows for an additional learning opportunity that requires a good understanding of theory and the development of relevant hypotheses that allow for efficient theory testing. Adding this additional task to the student project might also be feasible for BA-theses if more time (and ECTS) is allotted to the thesis work in the curriculum. The format may also be adapted for a research methods class: Single steps of the schedule may be practiced (e.g., step 3: how to conduct a power analysis, step 5: how to write and register a pre-data report in OSF, etc.) or the entire schedule may be realized in a two-semester course.

### **Conclusion**

In the Hagen Cumulative Science Project (HCSP), students successfully conducted more than eighty direct replication studies to realize their bachelor- and master-theses.

Within the HCSP framework, students learned key competences of empirical research as documented (e.g.) in their written theses. Specifically, in their intense reading and consideration of the published article and the available data as well as in their attempts to reproduce materials, methods, and analyses, students gained deep insights on real, purposeful scientific work and acquired special competencies in the following skills:

(1) evaluating research questions critically by understanding an original study in detail to prepare its replication,

(2) reflecting whether the applied methods of the original study allow to answer the posed research question, and

(3) obtaining firsthand experience concerning what it takes to conduct and document an empirical study in such a way that other researchers can potentially replicate it.

The format integrates aspects of open science within the individual steps of conducting the empirical project. Students develop concrete skills such as conducting a power analysis and

pre-registration but also gain a broader understanding of the nature of accumulating knowledge in empirical science.

Studies on the learning experience in traditional theses (e.g., literature reviews, empirical studies investigating new research questions) show that students value the autonomy of working independently and in-depth on a self-selected topic, resulting in a strong sense of ownership (Todd, Bannister & Clegg, 2004). However, the challenge of finding a “researchable” research question (Todd et al., 2004), while still developing methodological and writing skills, can frustrate students and lead to studies of poor quality (Grahe et al., 2012). Structured replication projects can overcome those obstacles.

The HCSP is one of few replication projects integrating students in their attempt to validate empirical results. Similar projects have integrated replication studies at different stages in the curriculum such as in introductory methods classes (Frank & Saxe, 2012) or advanced research methods classes (Standing et al., 2014; Wagge et al., 2019). The degrees of freedom in the selection of feasible replication projects allow instructors to adjust the individual workload and level of difficulty to students’ abilities. This is possible, for example, by adjusting the group size of students conducting a replication, pre-selecting studies for replication vs. allowing students to identify suitable studies on their own, and replicating more vs. less challenging replication studies. Students can thus be introduced to replication studies early in the curriculum. Ideally, students learn about the general ideas of transparent and open science (e.g., in an introductory method class) as well as specific tasks such as conducting their first re-analysis (e.g., in an introductory statistics class) and developing their first pre-registration (e.g., in empirical-experimental internships) throughout the course of their studies.

Overall, the framework of the Hagen Cumulative Science Project can be adapted flexibly to different student needs and teaching goals. The format allows instructors to plan, organize, and supervise bachelor- and master-theses in an effective manner. With this structured approach, students are guided through the process of empirical research (in accordance with principles of open science) and are enabled to make a meaningful contribution to cumulative science.

### References

- Arslan, R. C. (2019). How to automatically document data with the codebook package to facilitate data reuse. *Advances in Methods and Practices in Psychological Science*.  
<https://doi.org/10.1177/2515245919838783>
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., . . . Nosek, B. A. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108-119.
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior research methods, 43*, 666-678.
- Bouwmeester, S., Verkoeijen, P. P., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., ... & Evans, A. M. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science, 12*, 527-542.
- 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.
- Carsey, T. M. (2014). Making DA-RT a reality. *PS: Political Science & Politics, 47*(1), 72-77.
- Cetkovic-Cvrlje, M., Ramakrishnan, L., Dasgupta, S., Branam, K., & Subrahmanyam, L. (2013). A multi-disciplinary analysis of intensive undergraduate research. *Council on Undergraduate Research, 33*, 16-22.
- Earp, B. D., & Everett, J. A. (2015). How to fix psychology's replication crisis. *The Chronicle of Higher Education, 62*, B14.
- 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*, 175-191.
- Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology, 57*, 153-169.
- Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science, 7*, 600-604.

- Fuchs, H. M., Jenny, M., & Fiedler, S. (2012). Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science, 7*, 639-642.
- Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 391-408). Thousand Oaks, CA: Sage.
- Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science, 1*, 121-130.
- Grahe, J., Brandt, M. J., Ilzerman, H., Cohoon, J., Peng, C., Detweiler-Bedell, B., & Weisberg, Y. (2016). Collaborative replications and education project (CREP). Retrieved from [osf.io/wfc6u](https://osf.io/wfc6u).
- Grahe, J. E., Reifman, A., Hermann, A. D., Walker, M., Oleson, K. C., Nario-Redmond, M., & Wiebe, R. P. (2012). Harnessing the undiscovered resource of student research projects. *Perspectives on Psychological Science, 7*, 605-607.
- Greenbank, P., & Penketh, C. (2009). Student autonomy and reflections on researching and writing the undergraduate dissertation. *Journal of Further and Higher Education, 33*, 463-472.
- Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., ... & Calvillo, D. P. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*, 546-573.
- Höffler, J. H. (2013) Teaching replication in quantitative empirical economics. In *World Economics Association (WEA), Conference on the Economics Curriculum: Towards a Radical Reformation*, 3–31. Retrieved from <http://curriculumconference2013.weaconferences.net/wp-content/uploads/sites/5/2013/05/WEA-CurriculumConference2013-HOFFLER.pdf>.
- Janz, N. (2016). Bringing the gold standard into the classroom: replication in university teaching. *International Studies Perspectives, 17*, 392-407.

- Joel, S., Eastwick, P. W., & Finkel, E. J. (2018). Open sharing of data on close relationships and other sensitive social psychological topics: Challenges, tools, and future directions. *Advances in Methods and Practices in Psychological Science, 1*, 86-94.
- King, G. (1995). Replication, replication. *Political Science & Politics, 28*, 444-452.
- Kuehberger, A., & Schulte-Mecklenbeck, M. (2018). Selecting target papers for replication. *Behavioral and Brain Sciences, 41*, E139.
- Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*, 355-362.
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science, 1*, 131-144.
- Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods, 48*, 1205-1226.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.
- Perlman, B., & McCann, L. I. (2005). Undergraduate research experiences in psychology: A national study of courses and curricula. *Teaching of Psychology, 32*, 5-14.
- Rouder, J. N., Haaf, J. M., & Snyder, H. K. (2019). Minimizing mistakes in psychological science. *Advances in Methods and Practices in Psychological Science, 1*-9.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*, 76-80.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*, 559-569.
- Standing, L. G., Grenier, M., Lane, E. A., Roberts, M. S., & Sykes, S. J. (2014). Using replication projects in teaching research methods. *Psychology Teaching Review, 20*, 96-104.
- Todd, M., Bannister, P., & Clegg, S. (2004). Independent inquiry and the undergraduate dissertation: perceptions and experiences of final-year social science students. *Assessment & Evaluation in Higher Education, 29*, 335-355.

- Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams Jr, R. B., ... & Bulnes, L. C. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science, 11*, 917-928.
- Wagge, J. R., Baciú, C., Banas, K., Nadler, J. T., Schwarz, S., Weisberg, Y., IJzerman, H., Legate, N., & Grahe, J. (2018). A demonstration of the collaborative replication and education project: Replication attempts of the red-romance effect. Retrieved from <https://osf.io/chax8/download?format=pdf>.
- Wagge, J. R., Brandt, M. J., Lazarevic, L. B., Legate, N., Christopherson, C., Wiggins, B., & Grahe, J. E. (2019). Publishing research with undergraduate students via replication work: The collaborative replications and education project. *Frontiers in Psychology, 10*, 247.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*, 726-728.

## Table

Table 1

*Procedural steps of the Hagen Cumulative Science Project.*

Step	Activity	Goal	Obstacles	Solutions
1a	Identify feasible replication studies	Instructors identify studies for students. Feasible studies are (1) empirical (i.e., correlational or experimental), (2) include a statistically significant effect (if replication success is later defined by statistical significance), (3) use statistical methods introduced as part of the undergraduate curriculum, (4) are not context (e.g., elections) or sample (e.g. clinical) dependent, and (5) do not require specialized technical equipment that is not available to the instructor (e.g., EEG).	Selected studies might not be representative for all studies in a journal. Thus, the average observed replicability rate may be a biased estimate for the general replicability of all studies in the journal (Kuehberger & Schulte-Mecklenbeck, 2018).	If random sampling of studies is not feasible, report how studies were selected and, thus, what constitutes the reference population according to the sampling plan.
1b	Match students with replication	Match students according to their interests with available replication studies.	Some studies are more popular within the student population than others.	Students are matched according to their preferences and by random draw if necessary.

	studies			
1c	Select central finding	Students identify the central statistically significant finding of the study as the replication target. <sup>3</sup>	<p>1. Studies include more than one central finding (i.e., studies reporting a certain correlational pattern with equal theoretical importance).</p> <p>2. Studies report a null-effect as the central finding in the last study.</p>	<p>1. Students discuss with the original authors whether effects are equally important. If no central effect can be identified, students randomly select one of the relevant effects as the respective replication target.</p> <p>2. If the central finding is a null-effect, students choose the central statistically significant finding of the second to last study, etc. Instructors may instead decide to include null-effects and test those in a Bayesian framework.</p>
2	Re-analyze original data	A reanalysis of the original data allows students to obtain a better understanding of the study (i.e., types of variables, manipulations, operationalized hypothesis, etc.) and prepares students to analyze the replication results.	<p>1. Data is unavailable or unusable for re-analysis (e.g., authors did not respond to data requests).</p> <p>2. Results of the re-analysis</p>	<p>1. Instructors may prepare a fictitious dataset that allows students to practice the statistical procedure.</p> <p>2. Students inform original authors and discuss the extent and relevance of the</p>

---

<sup>3</sup> Similar to other replication initiatives (see e.g., OSC, 2015), we focused on the last feasible study in articles with more than one study.

			deviate due to errors within the original analysis.	deviation (i.e., typo vs. alters the conclusion of the study).
3	Do a power analysis	Students estimate planned sample size a-priori (power at least $\geq 80\%$ and $\alpha = 5\%$ ) based on the effect size reported in the article or derived from the reanalysis of the original data if sample sizes in the original studies are sufficiently large. Equivalence testing (Lakens, 2017) or useful heuristics (Simonsohn, 2015) for determining sample sizes may also be used (see main text for details).	The power analyses for the specific design or statistical test (e.g., multilevel regression model) cannot be conducted with standard software for power calculations such as G*Power (Faul et al., 2007).	Instructors provide support in the form of a bootstrapping power analysis using the original data.
4	Implement the study	Students implement the study in an experimental software available to the department (e.g. Psychopy, ProQuest).	<ol style="list-style-type: none"> <li>1. Original materials are unavailable (i.e., authors did not respond to requests for study materials).</li> <li>2. Study materials must be translated.</li> <li>3. Study materials might not be</li> </ol>	<ol style="list-style-type: none"> <li>1. Materials will be created following the description in the method section of the article as closely as possible.</li> <li>2. Translations will be made using back translation by two students and reviewed by the instructor before implementation.</li> <li>3. Add exploratory analysis to test for</li> </ol>

			completely feasible to the participant sample due to cultural differences between the original and replication sample.	potential impact of cultural differences.
5	Pre-register the study	Students prepare a pre-data report (see main text) and upload study materials at (e.g.) Open-Science Framework ( <a href="https://osf.io">osf.io</a> , see Meyer, 2018 for a list of alternative pre-registration platforms). Original authors are informed about the replication attempt and invited to comment on the pre-registration report. Before data collection, the project is pre-registered (i.e., a non-editable copy is saved on OSF). In similar replication attempts, before data collection, IRB approval can be obtained at this point if necessary.	<ol style="list-style-type: none"> <li>1. Copyright restrictions may prevent posting materials openly.</li> <li>2. Feedback from original authors is delayed and project cannot proceed due to uncertainty regarding when feedback will be provided.</li> </ol>	<ol style="list-style-type: none"> <li>1. Students document and post in the pre-registration how the original material was acquired (i.e., contact address of the original author or website).</li> <li>2. Authors are informed that their feedback can only be included when they respond within two weeks. For <i>JDM</i>, we had a positive response rate of approximately 80%.</li> </ol>
6	Collect data	Students collect the targeted sample size for the replication study.	<ol style="list-style-type: none"> <li>1. Students are not able to collect the target sample size within the available time frame</li> </ol>	<ol style="list-style-type: none"> <li>1. Before data collection, instructors and students agree on a deadline for data collection. If the target sample size has not</li> </ol>

to finish their thesis on time.

2. Original authors might conclude that differences between replication and original study have an impact on the target effect (e.g., student vs. general public sample, web vs. lab study, lower vs. original financial incentives).

been reached, students use the data collected thus far for their thesis and discuss this shortcoming critically and openly in their thesis. Simultaneously, data collection will be continued by the instructor and results will be summarized in a post-data report (see main text).

2. Replicators discuss with authors why the differences are expected to moderate the size of the effect (e.g., previous knowledge about the target effect in a psychology student sample might decrease its effect size). If differences cannot be resolved due to practical boundary conditions in the replication lab, identified moderators (e.g., previous knowledge of the effect) might be assessed in the replication study to check whether the effect is indeed moderated.

7 Analyze data Students analyze the data to test whether the original effect can be replicated

-

-

		(including respective assumption tests and manipulation checks).		
8	Document results	After careful re-examination from instructors, students publish data (i.e., code book and raw data; see also Arslan, 2019) and analysis scripts (e.g., on Open Science Framework). Original authors are informed about the results of the replication either by the instructor or students.	Constraints due to anonymization and ethical data sharing must be carefully considered.	Check with local data security official and national guidelines and if necessary, share reduced data set (e.g., without gender or age variables). See also published guidelines on ethical data sharing (e.g., Meyer, 2018; Gilmore, Kennedy, & Adolph, 2018; Joel, Eastwick, & Finkel, 2018; Carsey, 2014).
9	Prepare thesis	Students write their thesis based on the documented results.	-	-
10	Combine results in a meta-analysis	Instructors combine results from all replication studies in a meta-analysis. Overall replication rates may be ideally reported in a publication.	Quality of data and analysis documentation is heterogeneous concerning quality.	Instructors provide a standardized form for documentation and inspect student documents promptly after submission to allow time for contacting students for potential queries.

---