Discovering Empirically-Based Best Practices in Computing Education Through Replication, Reproducibility, and Meta-Analysis Studies

Monica M. McGill
Knox College & csedresearch.org
Galesburg, IL, USA
monica@csedresearch.org

ABSTRACT

Though some empirically-driven best practices in computing education exist, there are legitimate and serious concerns about the dearth of studies that have been replicated and/or reproduced in the sciences, including education science and computing education. Without the empirical evidence that comes from replicated, reproduced or meta-analytic studies to provide further verification that a particular practice is effective, the computing education research community may be unintentionally propagating poor practices driven by false findings derived from individual studies. Propagation of these practices can lead to distrust by practitioners, eroding the relationship between often well-intentioned researchers who want to help inform and shape the practice and those in the classrooms teaching, policymakers, and administrators. Therefore, it is incumbent on us as a community to seriously consider the state of our research practice, the challenges the community faces due to the lack of empirical evidence coming from our published studies, and how the community can have a broader discussion to evolve the field into a stronger practice. This short paper contains some foundational terminology and provides evidence of the lack of replication, reproducibility, and meta-analytic studies in general and in computing education. A summary of potential solutions is also proposed that can be explored in an effort to help frame a larger discussion of this issue with the goal of considering next steps needed to mature our field.

CCS CONCEPTS

• **Social and professional topics → Computing education**; **Computing education programs**; **Computer science education**.

KEYWORDS

Replication, reproducibility, meta-analysis, data synthesis, open science, datasets, primary education, secondary education, transparency, K-12, post-secondary, research

Koli Calling '19, November 21–24, 2019, Koli, Finland

ACM ISBN 978-1-4503-7715-7/19/11.
https://doi.org/10.1145/3364510.3364528

ACM Reference Format:
Monica M. McGill. 2019. Discovering Empirically-Based Best Practices in Computing Education Through Replication, Reproducibility, and Meta-Analysis Studies. In *19th Koli Calling International Conference on Computing Education Research (Koli Calling '19), November 21–24, 2019, Koli, Finland.* ACM, New York, NY, USA, Article 4, 5 pages. https://doi.org/10.1145/3364510.3364528

1 INTRODUCTION

There are legitimate and serious concerns about the dearth of studies that have been replicated and/or reproduced in education science, given that one study is not enough to test a theory in the education and behavioral sciences [30]. In October 2018, the U.S. National Science Foundation (NSF) and the Institute of Education Sciences (IES) jointly released the *Companion Guidelines on Replication & Reproducibility in Education Research*, which provided a thumbnail sketch of the high-level issues with replication and reproducibility, proposed ways to address it, and how the community can work to embed research methods and practices in NSF and IES proposals to ease the crisis [21].

For the purposes of further discussion, the NSF and IES's definitions of these terms are presented and adopted for this paper:

- ***Reproducibility*** is defined as "...the ability to achieve the same findings as another investigator using extant data from a prior study." [21, p.1] That is, reproducibility looks at the data derived from the study and analyzes the data independently. The results can then be compared to the results from the original study to validate the findings.
- ***Replication*** is the process of "...collecting and analyzing data to determine if the new studies (in whole or in part) yield the same findings as a previous study." [21, p. 1]

Given these definitions, conducting either requires a thorough understanding of the original study for which the results are being verified. Reproducibility studies, however, require less time and fewer resources, but can only be conducted if the data is available for further analysis. Replication studies, on the other hand, provide a higher standard and offers more validity to the effects of an intervention–but they also require more extensive time and resources with a whole host of issues that must be addressed. With these complexities come further definitions: direct replication and conceptual replication.

- *Direct replication studies* "...seek to replicate findings from a previous study using the same, or as similar as possible, research methods and procedures as a previous study." [21, p. 2]

- *Conceptual replication studies* "...seek to determine whether similar results are found when certain aspects of a previous study's method and/or procedures are systematically varied." [21, p. 2]

Both forms of replication studies can provide important empirical evidence used in informing best practices–they provide further efficacy and measures of effectiveness. Replicating and validating educational research findings provide assurance that the results of the studies are sound and reliable [9]. They can then give rise to the potential of predicting results through the synthesis of the empirical evidence from the various studies. Conceptual replication studies, by their definition, build upon previous studies to determine through empirical evidence which practices are best under which conditions for which target demographic groups [21].

What does this mean in the context of computer science education research? Does the community face the same or different challenges than the rest of the education science community? And what would efforts entail to move the computing education research community forward with the proposed avenue of addressing these challenges put forth by the NSF and IES and others in education science?

One can consider each of these questions by exploring the current challenges of education science within the context of computing education research and then provide a summary of solutions that have been proposed (and in some cases, recently implemented) within the behavioral and education science community. Given these solutions, this paper proposes a wider discussion of the importance of verifying individual studies, the extent to which our community is engaged in doing so, and what can be done to further move the community to engage in research that provides the capability for deriving best practices from validated empirical evidence.

2 OUR CHALLENGE: LACK OF EMPIRICALLY-DRIVEN BEST PRACTICES

There is no question that already exist some empirically-driven best practices in computing education. These have been primarily driven by the work being performed over the last fifty years in computing education at primarily the post-secondary level, such as pair programming in CS1.

These practices can, of course, be improved upon or replaced through various other methods if empirical evidence supports it. Unlike other fields, though, time-pressures are felt within the field of computing education, and, with the evolution of computing happening so quickly, researchers do not always have the luxury of a decade's time researching an intervention or new teaching method for a given field in computing. Even in the best-case scenario, that at the end of the decade researchers could empirically state that a method of teaching is the "best" for a particular subfield, the subfield may be replaced with or have evolved into a different one.

Further, at the K-12 levels, the situation is more distressing, with a recent study only finding six replication studies in CS education research literature (2009-2018) [10]. Research at the K-12 formal classroom level is more recent and though there is still little research being conducted in comparison to other subjects, the impact that early studies (which may be false) are having on practice can cause more harm than good [23].

Without empirical proof that a particular practice is effective, our community may be unintentionally propagating poor practices driven by false findings [32]. Propagation of these practices can lead to distrust by practitioners, eroding the relationship between often well-intentioned researchers who want to help inform and shape the practice and those in the classrooms teaching. Further leading to distrust is that these false findings can seep into policies, creating standards that are then implemented–only later to be shown to be ineffective [12, 25].

There are many challenges faced by the computing education research community. Three of these challenges that stand in the way of defining empirically-driven best practices within the computing education research community are:

- Lack of replication studies
- Lack of reproducibility studies
- Lack of meta-studies that compare and contrast work

Each are examined here in the context of the recent literature and, more specifically, the computing education research field.

2.1 Lack of Replication

The dearth of replication studies is spread across many fields, including visualization research [15], human-computer interaction studies [13], nutrition and health [26], environmental epidemiology [4], archaeological science [20], social Work [34], sports neuropsychology [28] and so many more. Within the field of behavioral and education sciences, the acknowledgement of the crisis derives from multiple secondary analysis studies. In a 2014 study, a systematic literature review was conducted across the top 100 education journals. The researchers found that only 0.13% of the studies were replication studies and with that, roughly half of these independent attempts at replication were successful [18]. In a 2015 study, a group of 270 psychologists could only replicate about 40% of 100 selected studies, and within that, the replicated studies showed overall effect size to be weaker than what was reported in the original study [7]. In a 2018 study, researchers identified 21 highly-influential social science studies for determining if they could replicate the findings with sample sizes five times greater than the original study [5]. And similarly, the effect sizes of the replicated studies were much weaker (about half) of those reported in the original studies.

The lack of replication studies in education science is not new, but worth exploring in the computing education field. A 2015 ITiCSE Working Group performed a meta-analysis of the use of educational data mining and learning analytics with respect to programming [14]. Findings suggested that a more solid understanding of the influencing variables can be achieved through validation and replication of these studies, with the authors emphasizing the "critical need" for such studies in this field. The group proposes five grand challenges, including the need for a progressive results database, to promote more replication and reproducibility studies, to promote more experimental design (over quasi-experimental), to move empirically-driven best practices back into the classroom, and to help practitioners apply these practices in their field. These challenges present the context of and reason for replication and reproducibility, though leave out the call for meta-analysis, a necessary process for comparing homogeneous studies.

In 2016, two studies were conducted that further brought light the lack of empirical evidence in computing education research, the lack of replication studies, and some reasons behind this [1, 2]. In 2019, Hao et al. published an extensive study to determine the rate of publication in computing education research and discovered that the field has a replication rate of 2.38% for the years 2009-2018 (*N*=2,269) Outside of these published studies, there is acknowledgement of the need for this–though very little empirical data indicates the extent of the problem with in our field.

2.2 Lack of Reproducibility

Though often lumped together, reproducibility (as defined by the NFS/IES report as well as others) differs from replication [8]. The report defines ways researchers can perform reproducibility studies, including the analysis of the data 1) using the same analysis procedures used in the original study and 2) using different statistical models [21]. The former can verify study results or identify errors within the analysis process or within the dataset. The latter can use the results from the different statistical models to again verify the results and conclusions drawn in the original study.

In the context of reproducibility and the sciences, Vuong and Ho (2019) describe the need for reproducibility based on studies conducted in economics, political science, psychology and more [12]. They cite a 2016 Nature's study that showed that 90% of the 1,576 scientists surveyed agree that there was a "reproducibility crisis" [12, p. 14]. Additionally, Laraway et al (2019) take a deep dive into reproducibility within the context of behavioral science and analysis, highlighting concerns within the community and offering suggestions to improve the practice [16]. Further, Yoccoz (2018) evaluates the reproducibility crisis in the field of ecology and evolution to find what is causing this crisis [35].

In the context of reproducibility and computing education research, with the conflation of these terms, it is difficult to know on the surface if the authors of articles understand their nuances and if reproducibility is also covered when the authors discuss replication–no formal definitions are provided. For example, another term, *reanalysis*, was found to describe reproducibility as defined by the NSF/IES report [25]. The nomenclature could also be adding to the lack of understanding of the terms.

Despite this, two studies call for reproducibility studies in the context of big data analysis of educational systems, including the 2015 ITiCSE Working Group paper [14, 24]. Though the terms described by the NSF and the working group differ (re-analysis versus reproducibility), a literature review was not conducted. However, case studies are provided to demonstrate the possibilities for reproducing and re-analyzing studies. Based on our inability to find meta-studies on reproducibility in the computing education research field, additional research (including systematic literature reviews) is needed to determine the state of reproducibility research in the computing education research literature.

2.3 Lack of Meta-Studies

Similar to the issues involving reproducibility and replication studies, single studies are more prone to errors or bias [11, 19]. Meta-studies synthesize the results of interventions or practices in ways that provide further validation of its impact on learners.

Without "good" data in independent studies and without the lifting of replication to a higher esteemed level, the authors of a 2016 study report that meta-analysis and synthesis of data needed to build strong theories cannot be conducted [2]. That is, even if more meta-analyses are conducted, if the findings from the individual studies have not been further validated or shown to be sound, what results then would a meta-analysis show? Further, as part of their conclusion, the authors have a call of action to the community to "...move toward meta-analysis of the literature for building of theories about CSEd." [2, p. 125] As presented in the next section, though, building more evidence within studies to ensure that the results of meta-analysis are sound requires significant change in how research is conducted across the community.

3 POTENTIAL SOLUTIONS

The dearth of these types of studies in computing education research is due to multiple cultural disincentives and research obstacles for conducting them. This is partially confirmed in a 2016 study of researchers [1]. Although computing education researchers seem to agree that there is value in replication studies, 73 such researchers also found that these researchers believe that replication studies are more difficult to publish, add less value with respect to incurring citations and applying for grant funding, and is not valued in the promotion process–which values innovative, original work [1]. They also found that these researchers in general are not interested in conducting replication studies.

In education science, the NSF/IES report lays part of the blame for these challenges on researcher disincentives, implementation challenges, and the difficulty in interpreting findings [21]. To help fix this, there are calls for replication studies be explicitly called for in Calls for Papers of conference and journals with the acknowledgement of the value of these types of studies called out [1, 10]. Ahadi et al. state that this would "...lead to more critical evaluation of new theories and better understanding of the generalizability of research findings." [1, p. 8] Other researchers calls for skipping the replicability issue entirely and instead encourages the focus on improving the quality of original studies [1, 15]. In this section, proposed solutions are briefly discussed, some of which have already been put into practice in other fields.

3.1 Improving individual studies

Several calls of action have been made across various fields to improve the research practices among individual studies [6, 15, 27, 31]. The practice of engaging in replication, reproducibility, and meta-analytic studies all hinge upon the quality of original, individual studies. Quality research practices are also equally important in the cases were no replication, reproducibility, or meta-analytic studies are conducted and decisions about policy or pedagogy are made based on one study.

3.2 Pre-Registration of studies

The pre-registration of studies is mentioned in several articles examining how to improve the practice [11, 17, 34]. This is the practice of requiring researchers to submit their research study methods and proposed analysis of the data prior to collecting data. This effort helps ensure that researchers are less likely to change hypothesis or data analysis measures to prevent questionable research practices (like "significance chasing") by supporting findings (novel results) that are "more publishable" [17, 27].

3.3 Transparency and better access to results through open science

One of the most supported changes to address the replication/reproducibility crisis is to move towards a model of transparency and open science [7, 11, 12, 17, 27]. Open science can be interpreted in several ways, including access since it gives the ability for those who cannot easily access research findings (teachers, policymakers, other researchers) full access. By so doing, teachers and policymakers can start to more easily base their actions on empirical findings [11, abstract] and can further enable public trust in the field [12].

Ho et al (2019) further call for open reviews and open dialogues about the research while Hillary et al (2019) and Yaffe (2019) call for data, code, and materials to be shared [11, 12, 34]. In addition to increasing credibility of the researchers and the work they conduct, these practices create further transparency and provide a stronger path to replicate and improve upon the original study.

3.4 Large-scale collaborative science

Large-scale collaborative science builds upon open science, specifically datasets, and can be used to "...understand which instructional practices work for whom and under what condition...." [11, p. abstract] This is noted as a key research practice needed to increase replicability and thereby the credibility of educational science [6]. Though this is happening within the learner space itself with tools designed to assist learning of computational thinking and computer science, extending this to the practice of computing education research will enable greater synthesis of data results [19].

3.5 Power analysis and effect size reporting

Washburn et al (2018) note that in the field of psychology formulating a power analysis and providing effect size are important steps to report sufficient data needed for study validation (via replication and reproducibility) and synthesizing results [31]. This is further supported by the American Psychological Association (APA) [3].

3.6 Tools to support reproducibility, replication, and meta-studies

Software systems that both store data for sharing and for analysis from studies and provide the tools for data analysis can be a standardized and powerful tool if created in a way that helps improve reproducibility of results within a study and cross-study synthesis and analysis [11, 33]. As a community, many computing education researchers have an understanding of what software development entails and what might be needed to design and create these types of tools [19].

3.7 Culture of Improved Practice

Washburn et al (2018), Ahadi et al (2016), and Hillary et al (2019) further bring to light the issues of cultural changes and ensuring that the researchers are properly incentivized to embrace these changes [1, 11, 31]. Without including the larger community in the conversation and carefully prioritizing, selecting, and promoting the changes that are a best fit for computing education, steering the education community in a direction where replication, reproducibility, and meta-analytic studies (and the necessary steps to make these studies happen) are valued will remain unaccomplished.

Tanweer (2018) proposed the concept of *exostructure* "...as a companion to infrastructure and a key mechanism enabling scaling in data science of the social. Exostructures are made up of the components of temporary, project-based collaborations intended to spawn replication or further investment in information and knowledge infrastructures." [29, p. iv], which is further supported by [22]. This is an interesting way of phrasing the "things around the methods and tools" that need to happen to ensure its propagation and success.

4 THE VALUE OF FURTHER DISCUSSION

The computing education research community propagates and encourages the belief that individual studies should be novel and innovative and that validating these individual studies by conducting replication, reproducibility, and meta-analytic studies is secondhand research not valued. As a community, we cannot expect different results without changing the culture and practice of computing education research and we cannot change the culture and practice without changing fundamental beliefs in the field. Hillary and Medgalia note that a "scientific community...motivated by this *crisis* may be at critical cross-roads for change engendering a culture of transparent, open science where the primary goal is to test and not support hypotheses about specific interventions." [11, p. abstract]

But changing these beliefs is not easy–with some being more easily adopted and accepted within the community than others. Washburn et al (2018) conducted a study that evaluated why psychology researchers don't adopt practices that will help address the crises (e.g., preregistering hypotheses/methods, making data publicly available online, conducting formal power analyses, reporting effect sizes) [31]. When asked whether or not they had ever engaged in these practices, they found that the least adopted practice (with an adoption rate of 27% of the 1,053 participants) was the preregistering of hypotheses/methods for a variety of reasons. The second least adopted practice (at a rate of 56%) was making data publicly available online. Almost all (99%) reported effect sizes in their studies and to a great extent (at 87%) conduct formal power analysis–which is to be expected in a more mature and rigorous field of behavioral research.

So, how do we as a community address this systemic issue? We can address overarching questions around the need for more empirical evidence with the goal of permanent change, such as:

- How "bad" is the lack of replication, reproducibility, and meta-analytic studies in the computing education research field? Can we empirically define this with evidence beyond the limited meta-studies that now exist?
- How do we define what "sufficiently corroborated" means for a particular intervention or pedagogy?
- Can we determine what our "false positive" rate is? How?
- What are acceptable standards for these type of studies? Will they be evaluated differently than original studies during the review process?

- What are the most impactful measures for the computing education research community? Which of these are most likely to be accepted by the community?
- What steps can leaders within the community adapt to start the process of promoting change?

Many suggestions have been made as researchers in other fields start to make changes to address these issues. Empirically defining what the state of our field is, examining the potential reasons for these issues, and considering the solutions in light of the current culture of our field will all be important in moving the field forward.

5 CONCLUSION

Computing education is the fastest growing field of study in K-12 education and it continues to evolve in post-secondary education as well. I invite the Koli community to join in an open discussion about the practices that the community currently and regularly engages in (those that have become standard and acceptable practices), how these practices can be improved, and together we can start to drive the field towards best teaching practices based on empirical evidence. And, in light of the fact that practices are being propagated into the K-12 computing education community without this empirical data, we can further consider what this means to those practices and the learners as well as the timing to improve these research practices–before our trust with teachers, administrators, and policymakers is eroded.

6 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1757402.

REFERENCES

[1] Alireza Ahadi, Arto Hellas, Petri Ihantola, Ari Korhonen, and Andrew Petersen. 2016. Replication in Computing Education Research: Researcher Attitudes and Experiences. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research (Koli Calling '16)*. ACM, New York, NY, USA, 2–11.

[2] Ahmed Al-Zubidy, Jeffrey C Carver, Sarah Heckman, and Mark Sherriff. 2016. A (updated) review of empiricism at the sigcse technical symposium. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. ACM, 120–125.

[3] American Psychological Association. 2010. *Publication manual of the American Psychological Association (6th ed.)*. American Psychological Association: Washington, DC, USA.

[4] Scott M Bartell. 2019. Understanding and Mitigating the Replication Crisis, for Environmental Epidemiologists. *Current environmental health reports* 6, 1 (2019), 8–15.

[5] Colin F Camerer, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A Nosek, Thomas Pfeiffer, et al. 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour* 2, 9 (2018), 637.

[6] Paulo Carvalho, Bryan G Cook, Joshua de Leeuw, Emily Fyfe, Hunter Gehlbach, Robert Goldstone, Matthew C Makel, David Mellor, Amanda Montoya, Benjamin Motz, et al. 2019. Actions for Increasing the Credibility of Educational Research. https://edarxiv.org/tv7rn/. (2019). Accessed: 2019-08-02.

[7] Open Science Collaboration et al. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015), aac4716.

[8] Chris Drummond. 2009. Replicability is not reproducibility: nor is it good science. (2009).

[9] John Fritz and Robert Carpenter. 2019. Triangulating Student Engagement with "Built & Bought" Learning Analytics. In *Emerging Trends in Learning Analytics*. Brill Sense, 158–174.

[10] Qiang Hao, David H. Smith IV, Naitra Iriumi, Michail Tsikerdekis, and Andrew J. Ko. 2019. A Systematic Investigation of Replications in Computing Education Research. *ACM Trans. Comput. Educ.* 19, 4, Article 42 (Aug. 2019), 18 pages. https://doi.org/10.1145/3345328

[11] Frank G Hillary and John D Medaglia. 2019. What the replication crisis means for intervention science.

[12] MT Ho and QH Vuong. 2019. The values and challenges of 'openness' in addressing the reproducibility crisis and regaining public trust in social sciences and humanities. *European Science Editing* 45, 1 (2019), 14–17.

[13] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough?: On the Extent and Content of Replications in Human-computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3523–3532.

[14] Petri Ihantola, Arto Vihavainen, Alireza Ahadi, Matthew Butler, Jürgen Börstler, Stephen H Edwards, Essi Isohanni, Ari Korhonen, Andrew Petersen, Kelly Rivers, et al. 2015. Educational data mining and learning analytics in programming: Literature review and case studies. In *Proceedings of the 2015 ITiCSE on Working Group Reports*. ACM, 41–63.

[15] Robert Kosara and Steve Haroz. 2018. Skipping the Replication Crisis in Visualization: Threats to Study Validity and How to Address Them: Position Paper. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*. IEEE, 102–107.

[16] Sean Laraway, Susan Snycerski, Sean Pradhan, and Bradley E Huitema. 2019. An overview of scientific reproducibility: Consideration of relevant issues for behavior science/analysis. *Perspectives on Behavior Science* 42, 1 (2019), 33–57.

[17] Ljiljana B. Lazarevic and Iris Zezelj. 2018. How open science norms improve scientific practices. https://bit.ly/2lUc0HO. Accessed: 2019-08-01.

[18] Matthew C Makel and Jonathan A Plucker. 2014. Facts are more important than novelty: Replication in the education sciences. *Educational Researcher* 43, 6 (2014), 304–316.

[19] Monica M. McGill, Tom McKlin, and Errol Kaylor. 2019. Defining What Empirically Works Best: Dynamic Generation of Meta-Analysis for Computer Science Education. In *Proceedings of the 2019 ACM Conference on International Computing Education Research (ICER '19)*. ACM, New York, NY, USA, 199–207. https://doi.org/10.1145/3291279.3339401

[20] Stephen R Merritt, Michael C Pante, Trevor L Keevil, Jackson K Njau, and Robert J Blumenschine. 2019. Don't cry over spilled ink: Missing context prevents replication and creates the Rorschach effect in bone surface modification studies. *Journal of Archaeological Science* 102 (2019), 71–79.

[21] National Science Foundation and Institute for Education Science. 2018. Companion Guidelines on Replication & Reproducibility in Education Research: A Supplement to the Common Guidelines for Education Research and Development. https://ies.ed.gov/pdf/CompanionGuidelinesReplicationReproducibility.pdf. Accessed: 2019-03-01.

[22] Gina Neff, Anissa Tanweer, Brittany Fiore-Gartland, and Laura Osburn. 2017. Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big data* 5, 2 (2017), 85–97.

[23] Institute of Education Sciences. [n. d.]. About the WWC. Retrieved August 26, 2019 from https://ies.ed.gove/ncee/aboutus/

[24] Misato Oi, Masanori Yamada, Fumiya Okubo, Atsushi Shimada, and Hiroaki Ogata. 2017. Reproducibility of Findings from Educational Big Data: A Preliminary Study. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. ACM, New York, NY, USA, 536–537.

[25] Owen Ozier. 2019. Replication Redux: The Reproducibility Crisis and the Case of Deworming.

[26] Craig Sale and Duane Mellor. 2018. A call for replication studies in Nutrition and Health.

[27] Alexandra Sarafoglou, Suzanne Hoogeveen, Dora Matzke, and Eric-Jan Wagenmakers. 2019. Teaching Good Research Practices: Protocol of a Research Master Course. (2019).

[28] P Schatz, K Bogar, and R J Elbin. 2019. Addressing the "Replication Crisis" in Sports Neuropsychology: Validation of the Two-Factor Theory of ImPACT. *Archives of Clinical Neuropsychology* 34, 5 (07 2019), 758–758.

[29] Anissa Tanweer. 2018. *Data science of the social: How the practice is responding to ethical crisis and spreading across sectors*. Ph.D. Dissertation.

[30] Olmo van den Akker. 2019. How do researchers interpret the results of multiple experiments? *Open Science 2019, Trier, Germany* (2019).

[31] Anthony N Washburn, Brittany E Hanson, Matt Motyl, Linda J Skitka, Caitlyn Yantis, Kendal M Wong, Jiaqing Sun, JP Prims, Allison B Mueller, Zachary J Melton, et al. 2018. Why do some psychology researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science* 1, 2 (2018), 166–173.

[32] Tobias Wingen, Jana Berkessel, and Birte Englich. 2019. No Replication, no Trust? How Low Replicability Influences Trust in Psychology. (2019).

[33] Samantha MW Wood and Justin N Wood. 2019. Using automation to combat the replication crisis: A case study from controlled-rearing studies of newborn chicks. *Infant Behavior and Development* 57 (2019), 101329.

[34] Joanne Yaffe. 2019. From the Editor: Do We Have a Replication Crisis in Social Work Research?

[35] Nigel Gilles Yoccoz. 2018. The replication and reproducibility crises: origins and consequences for studies of ecology and evolution. In *Septentrio Conference Series*. https://septentrio.uit.no/index.php/SCS/article/view/4525