

Editorial

Open Science and Educational Technology Research

Jason M. Lodge

The University of Queensland, Australia

Linda Corrin

Swinburne University of Technology, Australia

Gwo-Jen Hwang

National Taiwan University of Science and Technology, Taiwan

Kate Thompson

Queensland University of Technology, Australia

Over the last decade a spate of issues has been emerging in empirical research spanning diverse fields such as biology, medicine, economics, and psychological science. The crisis has already led to fundamental shifts in how research is being conducted in several fields, particularly psychological science. Broadly labelled the ‘replication crisis’, these issues place substantial doubt on the robustness of peer-reviewed quantitative research across many disciplines. In this editorial, we will delve into the replication crisis and what it means for educational technology research. We will address two key areas, describing the extent to which the replication crisis applies to educational technology research and suggestions for responses by our community.

Background

It is commonly accepted that the replication crisis initially emerged from social psychology (Klein, 2014). Several key studies seem to have contributed to the emergence of the crisis, central among them was a paper purportedly showing evidence of psychic phenomena (*psi*; Bem, 2011). From the initial debate about the robustness of the research published in the *psi* paper, there rapidly emerged problems with other ostensibly robust findings in the psychology literature (Pashler & Wagenmakers, 2012). Attempts to replicate long established ‘effects’ revealed that many of these phenomena could not be observed under similar conditions despite many being reported essentially as fact in psychology textbooks. There have been systematic attempts to understand the scale of the problem (for example, Open Science Collaboration, 2015). Some estimates suggest that as much as half of the empirical research in psychology, with apparently generalisable findings, cannot be reliably replicated (Klein, 2018).

The scope and scale of the problem has since become apparent. While the initial focus was on the replication of findings in social and experimental psychology, other systemic issues emerged and the problem began to be discussed in other disciplines (for example, in Medicine; Mobley et al., 2013, in Economics; Camerer et al., 2016). The replication crisis is therefore not just about the robustness of findings in psychological science. The issue is much broader with implications for the way that research is conducted, communicated, and the implications for practice. It is timely to explore the major concerns that have emerged as part of the crisis and consider what these mean for research in educational technology. We will start by focussing on the issues underpinning the overarching problem with the replicability of quantitative studies.

Questionable Research Practices

Most of the issues underpinning the replication crisis appear to be associated with what have been labelled ‘questionable research practices’ or ‘QRPs’ (John, Loewenstein & Prelec, 2012). The major QRPs that have been identified related to the researchers are low power, p-hacking, hypothesising after the results are known (HARKing), poor statistical practices and the fabrication of data. These are all related to the way that experimental studies are designed and implemented. In essence, the kinds of studies that have been implicated tend to be controlled studies and involve testing an intervention against a control or alternate condition. These designs include the kinds of randomised control trials common in health and medicine

through to interventions tested on pre-existing groups, otherwise known as a quasi-experiment. An example of a quasi-experiment in education would be when a new approach or tool is tested in one class and compared to existing tools or approaches in another. In addition to QRPs, publication bias plays a key role in amplifying the impacts of the replication crisis.

One of the biggest issues that emerged from the replication crisis, and perhaps the core problem, is that many published studies are low powered. Power refers to the ability of a study to determine whether there is or is not an effect occurring at greater than chance levels. In other words, it is difficult to determine how generalisable or robust the findings are from low powered studies because often the number of participants in the study is too few to reliably detect an effect (or not). Without sufficient participants, and therefore statistical power, the likelihood of detecting an effect when there isn't one, or not detecting an effect when there is one increases markedly. Other factors, such as the design of the study, will have an impact the ability of the study to determine whether there has been some observable effect or not (Cohen, 1988). However, small samples are the most common issue leading to low powered studies (for review see, Button et. al., 2013).

Otherwise known as the 'data dredging' approach, p-hacking involves setting the conditions of a study in such a way as to maximise the likelihood of finding an effect, independent of whether the effect is real. As the name suggests, the aim of the approach is to manipulate the design and/or execution of a study or analysis protocol to find a statistically significant result. While there may be justification for carrying out an exploratory study, to do so with the explicit purpose of finding a statistically significant effect will markedly increase the chances of a false positive result. A tell-tale sign that p-hacking is occurring is that there is a disproportionate number of p-values accumulating just below $p = .5$ in a body of published research. This pattern is precisely what has been found in psychology research literature (for example, Masicampo & Lalande, 2012) and more broadly in the scientific literature (Head et al., 2015).

HARKing is somewhat more difficult to determine on the basis of a published article alone. However, it is evident that there are cases of people conducting studies and then formulating the hypotheses after the fact while representing the process the other way around (see Kerr, 1998). The motivation for doing so is that it is much easier to commit to what you are looking for after you already have evidence to support that finding. There are exploratory-type studies where it makes some sense to hypothesise after the data are collected and analysed. This approach would make sense in a pure data-mining study, for example, where there might not necessarily be a clear hypothesis at the beginning of the process. The problem arises when researchers falsely present their work as though the hypothesis were developed a priori.

One other issue that has emerged through the replication crisis is the long-accepted use of certain practices that are either questionable or incorrect (for a critical review see, Gigerenzer, 2018). Gigerenzer (2018) argues that much of the blame for poor statistical practices is because researchers are often taught statistical rituals rather than the theoretical foundation required for statistical thinking. Many examples of poor statistical practice have been observed, and other approaches have been the topic of extensive debate. While these have included a variety of practices, an example of one is the assumption that a Likert scale is a continuous measure (for example, Jamieson, 2004). In a technical sense, it is not correct to assume, for example, that a five-point scale from 'strongly disagree' to 'strongly agree' is continuous and therefore means can be calculated.

The replication crisis has also brought to light some rare examples of researchers completely falsifying data. Perhaps the most prominent example of this is work by Diederik Stapel, which was found to be fraudulent (Levelt, Noort & Drenth, 2012). An investigation was conducted into the work published by this researcher, who ultimately admitted to just making data up to fit his hypotheses. Although rare, these examples are perhaps the most troubling issue that has emerged as part of the crisis.

While not necessarily a QRP, research publication trends seem to be exacerbating the problems we have described here. Novel findings are perceived as more interesting and are more likely to be accepted for publication than either replications, or, particularly, failed replications (Franco, Malhotra & Simonovits, 2014). The greater likelihood of findings that seem to have found evidence for some sort of phenomena that has not been described previously to be published means that the overall body of literature is not an accurate reflection of the robustness of phenomena reported in it.

Implications for educational technology research

Given the replication crisis emerged from positivist domains that rely heavily on experimental methods, some authors have suggested that there are few, if any, implications for educational research (Morrison, 2019). Arguably, education and educational technology as research areas do not rely as much on experimental designs and generalisability has been an ongoing issue for debate (Henderson, Redmond & Heinrich, 2018). What works in one institution, or one digital learning environment cannot, a priori, be assumed to work in precisely the same way elsewhere. Despite the clear differences in approaches used in educational technology research and research in other disciplines. What is common are approaches that are focused on interventions of some kind. Many research studies reported on in AJET, and other educational technology journals, focus on the impact of technologies on students, teachers, and other parties. Some specifically require authors to produce an implications section. In many of the articles, the authors then make claims about the impact of the technology beyond the study reported in the paper. One example would be to state that the findings of the use of a tool or approach in one context suggest the potential for that tool or approach to have impact in another discipline context. Therefore, these articles are inherently addressing replicability and generalisability, at least at some level.

The key issues wrapped up in the replication crisis apply to the design and publication of research carried out by our community. Low powered studies are common. There is some suggestion in the analysis of p-values that there is likely to be widespread publication bias and could well be p-hacking taking place (Chow & Ekholm, 2018). Poor statistical practices have been an ongoing topic for debate in educational and educational technology research (for example, Oluwatayo, 2012). There does also seem to be evidence of publication bias in educational research more broadly (Ropovik, Adamkovic & Greger 2021). Educational technology research differs in many ways from psychological science and many of the other disciplines that have been caught up in the replication crisis. However, there is sufficient evidence to suggest that many of the issues that have emerged through the crisis are also evident in this context. We therefore argue that the replication crisis is something that we, as a research community, need to be paying attention to and address.

The response

Underlying all the issues wrapped up in the replication crisis are issues related to how scientific practices contribute to the discovery or creation of new knowledge, and the kinds of incentives that the publication system has created for researchers. Understandably then, the initiatives aimed at addressing the issues are complex, multifaceted, and will take some time to mature. Major reforms in research practices have been brought together under two overlapping banners. The first of these is what is being referred to as the ‘Open Science’ movement (Landrain et al., 2013). Open Science refers to a set of practices aimed at making as many elements of the research process as transparent and publicly available as practically possible. The Open Science Foundation (osf.io) is the primary organisation progressing this agenda. Aligned with this movement is the emergence of a new field, metascience. As the name suggests, this field is focussed on the study of science and scientific methods. This field takes a systematic approach to understanding and evaluating the robustness of empirical research findings across disciplines. These then are used to support increased transparency and new statistical practices.

Increased transparency

One key response to the replication crisis involves a call to increase the transparency of the research process. In order to enable this in a standard approach, a pre-registration process has been adopted in areas such as medical research. This process can involve publicly committing to a hypothesis and a set of methods prior to carrying out a study. Alternatively, several journals offer to pre-review methods before studies are carried out. In both cases, researchers provide extensive detail about the study before data are collected to reduce the likelihood of p-hacking or HARKing. If a journal accepts a pre-reviewed study protocol, in most cases, the journal then commits to publishing the results, no matter how they come out. This approach then also helps to address problems associated with publication bias.

There are many ways of increasing transparency. Figure 1 outlines the aspects of the research process that are available via traditional publication (the black shaded elements). The remainder of the approaches

(greyed out) are not traditionally shared. These components of the process could easily be made open (assuming that there are no ethical or commercial reasons for not doing so).

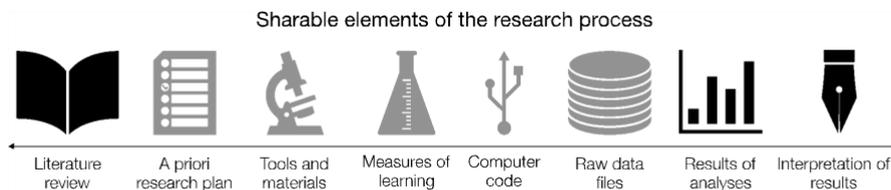


Figure 1. Elements of the research process that are (black) and could be (grey) open (from Alhadad, Searston & Lodge, 2018, with permission)

'New statistics'

The other major branch of the response to the replication crisis is a renewed emphasis on statistical practices. This includes a broad call for more systematic reviews and meta-analyses (Cumming, 2014). However, these approaches only deal with parts of the problem and, while providing a sense of the overall state of the field and whether p-hacking and publication bias are at play, they do not address the issues evident in individual studies. Alongside the increased focus on metascience, there are also calls for more rigorous statistical practices in individual studies. This includes the routine reporting of effect sizes and confidence intervals (for example, Cumming, 2013), through to arguments that 'frequentist' or traditional probability-based approaches be replaced by Bayesian analysis methods (for example, García-Pérez, 2016).

Implications for AJET

What does the replication crisis and its response mean for the future of AJET and educational technology research? As competition increases for the limited number of manuscripts we can publish in issues, smaller-scale quantitative studies are also less likely to be published relative to the total number of manuscripts submitted. However, small scale qualitative or case study research is critical for advancing the field. Our point here is not that small scale research per se is not high quality. Small scale research that is specifically aimed at detecting cause and effect relationships and/or finding generalisable effects is something that will attract scrutiny. As an editorial team, the lead editors and associate editors are deliberately mindful of the quality of the methods being used in all studies reported in manuscripts submitted to the journal. However, given the replication crisis, we are mindful of those making generalisable claims on the basis of quantitative data. We are also actively engaged in the Open Science and Metascience communities so that we can keep up with the latest developments in these emerging areas.

Further to the standard operating processes we already have in place, we would like to encourage authors to consider options for increased transparency. The Open Science Foundation provides options for pre-registering studies on their site (osf.io). There are also options on the site for uploading materials, measures and datasets. As the movement towards Open Science continues, these options are increasingly important considerations for the publication of quantitative articles across different disciplines and fields. They provide editors, reviewers, and ultimately the research community with greater confidence that any effect found in a quantitative study is robust with at least some sense that the findings are applicable beyond the context in which the study was conducted.

It would be counter to the philosophies that underpin our community to suggest that options for transparency be made mandatory. Many high-quality studies in educational technology research are not aimed at finding replicable, generalisable effects. For those that are, we would strongly encourage authors to increase the robustness of their results by seeking large and diverse samples, focusing on the highest quality research design and analysis, and considering options for openness and greater transparency.

As a high impact journal, we also need to be mindful of the role we play in publication bias in the field. Our primary concern as editors is on the quality of the research in submitted manuscripts. It is tempting to err towards publishing papers for their novelty or because they might attract substantial attention and citations. This approach is ultimately a disservice to the field, to the community of practitioners who apply

the research published in these pages and, ultimately, to students. The replication crisis might not be a direct result of the traditional practices in journal publication, but the evidence suggests that the situation has been made demonstrably worse through these practices. As editors, we want to ensure that we maintain the outstanding level of impact that AJET has had over time and is the result of the extraordinary efforts of our predecessors. One lesson of the replication crisis for journal editors that is seemingly clear is that the highest standards of quality must take precedence over what gets the most clicks. We only hope that we can do our part to ensure that quality triumphs over attention-seeking.

Acknowledgements

We would like to extend our gratitude to the wonderful team of AJET associate editors who manage reviews and revisions of manuscripts. We would also like to extend our thanks to the AJET copyeditors, Antonina Petrolito and Kayleen Wood, who do an amazing job, at times, under tricky circumstances. We are also thankful for the ongoing support for the journal provided by the ASCILITE Executive and ASCILITE members. Last but certainly not least, we thank all the volunteer reviewers who provide their expert advice on manuscripts submitted to AJET. As lead editors, we get a lot of the attention, but we could not do what we do without the contributions of all of you. So, a massive thank you to associate editors, copyeditors, to ASCILITE and to AJET reviewers. We very much appreciate all that you do for AJET.

References

- Alhadad, S., Searston, R. & Lodge, J. (2018). Interdisciplinary Open Science: What are the implications for educational technology research? In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri & J. Tai (Eds.), *Open Oceans: Learning without borders. Proceedings ASCILITE 2018 Geelong* (pp. 303-308).
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407–425.
<http://dx.doi.org/10.1037/a0021524>
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365-376. <https://doi.org/10.1038/nrn3475>
- Camerer, C. F., A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmeld, T. Chan, et al. 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351 (6280):1433–6. <https://www.science.org/doi/10.1126/science.aaf0918>.
- Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, 30(3), 727-744.
<https://doi.org/10.1007/s10648-018-9437-7>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29.
<https://doi.org/10.1177/0956797613504966>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502-1505. <https://doi.org/10.1126/science.1255484>
- García-Pérez, M.A. 2016. Thou shalt not bear false witness against null hypothesis significance testing. *Educational and Psychological Measurement* 77: 631–662.
<https://doi.org/10.1177/0013164416668232>.
- Gigerenzer G. (2018) Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1 (2), 198-218.
<https://doi.org/10.1177/2F2515245918771329>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, 13(3), 1-15.
<https://doi.org/10.1371/journal.pbio.1002106>
- Henderson, M., Redmond, P., & Heinrich, E. (2018). A caution about causation. *Australasian Journal of Educational Technology*, 34(5). <https://doi.org/10.14742/ajet.5030>
- Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical education*, 38(12), 1217-1218.
<https://doi.org/10.1111/j.1365-2929.2004.02012.x>

- John, L. K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23 (5):524-532. <https://doi.org/10.1177/0956797611430953>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3): 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Klein, S. B. (2014). What can recent replication failures tell us about the theoretical commitments of psychology? *Theory & Psychology*, 24(3):326-338. <https://doi.org/10.1177/0959354314529616>
- Klein, R. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1 (4): 443–490. <https://doi.org/10.1177/2F2515245918810225>
- Landrain, T., Meyer, M., Perez, A. M., & Sussan, R. (2013). Do-it-yourself biology: challenges and promises for an open science and technology movement. *Systems and synthetic biology*, 7(3), 115-126. <https://doi.org/10.1007/s11693-013-9116-4>
- Levelt, W. J. M., Drenth, P., & Noort, E. (Eds.). (2012). Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. Commissioned by the Tilburg University, University of Amsterdam and the University of Groningen. Retrieved from: <http://hdl.handle.net/11858/00-001M-0000-0010-258A-9>
- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly Journal of Experimental Psychology*, 65(11), 2271-2279. <https://doi.org/10.1080/17470218.2012.711335>
- Mobley, A., Linder, S. K., Braeuer, R., Ellis, L. M., & Zwelling, L. (2013). A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PloS one*, 8(5), e63221. <https://doi.org/10.1371/journal.pone.0063221>
- Morrison, K. (2019). Realizing the promises of replication studies in education. *Educational Research and Evaluation*, 25, (7-8), 412-441, <https://doi.org/10.1080/13803611.2020.1838300>
- Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, 2(2), 391-391. Retrieved from: <https://www.richtmann.org/journal/index.php/jesr/article/view/11851>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://www.science.org/doi/10.1126/science.aac4716>
- Pashler, H. & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528 –530. <http://dx.doi.org/10.1177/1745691612465253>
- Ropovik, I., Adamkovic, M., & Greger, D. (2021). Neglect of publication bias compromises meta-analyses of educational research. *Plos One*, 16(6), e0252415. <https://doi.org/10.1371/journal.pone.0252415>

Corresponding author: Jason M. Lodge, jason.lodge@uq.edu.au

Copyright: Articles published in the *Australasian Journal of Educational Technology* (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

Please cite as: Lodge, J. M., Corrin L., Hwang, G-J., & Thompson, K. J. (2021). Open science and educational technology research. *Australasian Journal of Educational Technology*, 37(4), 1-6. <https://doi.org/10.14742/ajet.7565>