



**QUEEN'S  
UNIVERSITY  
BELFAST**

## Replication of quantitative psychological and educational research: a philosophical analysis

Cantley, I. (2023). Replication of quantitative psychological and educational research: a philosophical analysis. In *AERA Annual Conference 2023: Proceedings* (AERA Annual Conference Proceedings). AERA. <https://doi.org/10.3102/2001119>

**Published in:**  
AERA Annual Conference 2023: Proceedings

**Document Version:**  
Publisher's PDF, also known as Version of record

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**  
Copyright 2023 The Authors.  
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

**General rights**  
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

**Open Access**  
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

# **Replication of Quantitative Psychological and Educational Research: A Philosophical Analysis**

## **Abstract**

This paper outlines the replication crisis in psychology, and highlights its significance for quantitative educational research. Greater methodological rigour, aimed at addressing methodological deficiencies in the conduct of scientific research, has been suggested as a response to the replication crisis. Following a review of various calls for greater methodological rigour, the current paper argues that there is potentially a more fundamental reason for the replication crisis. The case is made that the non-separability of a measured attribute from the measurement instrument, and the irreducible uncertainty in unmeasured attributes, may be the principal reasons for the replication crisis in psychology and education. Implications of such an explanation for the crisis are articulated in relation to the evidence-based policy agenda in education.

## **1. Introduction**

Since the beginning of the 21<sup>st</sup> century, scientific research has experienced a crisis of confidence in the form of what has become known as the ‘replication crisis’. This refers to the fact that, in many areas of scientific endeavour, including psychology and the allied discipline of education, difficulties have arisen when attempting to reproduce previous research findings (Hughes, 2018; Frias-Navarro et al., 2020). The replication crisis has potentially deleterious implications for educational research if it transpires that a large proportion of published quantitative research findings are unreliable (Makel & Plucker, 2014; Makel et al., 2021).

## **2. Theoretical basis and methodology**

The current paper briefly reviews various calls for greater methodological rigour that have been made as a response to the replication crisis. Whilst most responses advocate a range of strategies for addressing methodological deficiencies in research and publication practices, this paper fundamentally challenges, at a paradigmatic level, the premises upon which the notion of replication is predicated. I draw upon Wittgenstein’s later philosophy of mind to offer a novel philosophical critique of the quests for objectivity and universalism in psychology and education that are implicit in replication of empirical research findings.

## **3. Greater methodological rigour as a response to the replication crisis**

The replication crisis has been linked to various dubious research practices including, for example, the use of small samples, non-registration of research protocols before the research is conducted, absence of statistical power planning, and post-hoc formulation of hypotheses to ensure statistically significant results are obtained (Frias-Navarro et al., 2020; Wiggins & Christopherson, 2019). Consequently, it has been suggested that, when publishing their findings, scholars should be required to more fully explain their methodology and to refrain from concealing vital aspects of the approach taken (Simmons et al., 2011). It has also been argued that dubious research practices could be addressed by requiring researchers to pre-register their research protocols (including hypotheses, methodological plans and analytical plans) before data collection commences, so that methodological decisions are divorced from their potential impact on the study’s findings (Frias-Navarro et al., 2020). In addition, there have been calls to mandate the sharing of complete data sets, and the code used to analyse the

data, to permit other researchers to verify the findings of a study, and as a further antidote to dubious research practices (Frias-Navarro et al., 2020).

Prestigious academic journals usually prioritise the publication of original and statistically significant findings. Such outlets tend to be reticent to publish non-significant results and replication studies, which consequently leads to lacunae in the literature and a distinct bias towards statistically significant findings (Frias-Navarro et al., 2020; Wiggins & Christopherson, 2019). To address the reticence of journals to publish null results, Nosek and Lakens (2014) suggested that protocols should be established to ensure that both significant and non-significant findings are published for studies conducted in a sufficiently rigorous manner.

Therefore, those seeking to address the replication crisis in psychology and education have tended to focus their reform efforts on improving methodological practices, and to be less critical of the objectivist epistemological underpinnings of measurement practices in quantitative psychological and educational research.

#### **4. Philosophical explanation for the replication crisis**

The basic premise upon which the notion of replication is predicated presupposes that universal principles and laws underpin an objective social reality. In this section, however, I challenge the validity of such a worldview in the context of psychological and educational research by drawing upon fundamental aspects of Wittgenstein's later philosophy of mind. The arguments below are developed using the measurement of cognitive abilities for illustrative purposes, but similar arguments could be propounded for the measurement of other intentional psychological predicates.

Wittgenstein contends that a psychological trait, such as ability, cannot *just* consist of a mental process: "In the sense in which there are processes (including mental processes) which are characteristic of understanding, understanding is not a mental process" (Wittgenstein, 2009, §154). In particular, Wittgenstein's discussion of rule-following leads to philosophical dilemmas if a finite mental object, such as an image or formula, is viewed as the source of a student's ability. This then precipitates further conundrums if, for example, a standardised test of mathematical achievement is deemed to be an appropriate measuring instrument for *checking up* on students' mathematical abilities, as summarised below.

Bruner (1996, p. 129) has defined learning as developing the capacity to follow rules so as to “go beyond the information given”. Wittgenstein (2009), however, demonstrates that a paradox arises if the origins of a student’s ability to act according to a given rule is construed as a mental object such as a formula or image. Wittgenstein posits that, if a finite mental object were the source of the student’s ability to correctly follow the rule, any behaviour on the part of the student could be viewed as either agreeing with, or violating, the requirements of the rule if a suitable interpretation of the object is adopted.

This was our paradox: no course of action could be determined by a rule, because every course of action can be brought into accord with the rule. The answer was: if every course of action can be brought into accord with the rule, then it can also be brought into conflict with it. And so there would be neither accord nor conflict here.

(Wittgenstein, 2009, §201)

For example, suppose that a question on a standardised test of mathematical achievement requires students to calculate the value of  $a^3$  when  $a = 4$ . Clearly, a student could attach the usual interpretation to the formula, i.e., replace  $a$  by 4, then calculate  $4 \times 4 \times 4$ , and respond with the answer “64”. Alternatively, the student could interpret the formula in various unconventional ways, and suggest answers such as 12 (obtained by multiplying 4 by 3), 7 (obtained by adding 4 and 3) or 81 (obtained by calculating the value of  $3 \times 3 \times 3 \times 3$ ), or a multitude of other possibilities. In other words, the student could either interpret and respond to the question by following the normal (correct) mathematical conventions, or else they could fail to adhere to the normal (correct) conventions and consequently give an answer that is inconsistent with normal conventions. The inability of a finite mental object (e.g., mental image or formula) to guide the student in the correct use of a rule cannot be addressed by positing that the student must be able to interpret the object correctly since any interpretation of the object could itself be interpreted in multiple different ways, hence leading to an infinite regress (Wright, 2001).

Wittgenstein’s analysis of rule-following implies that, before a student proffers an answer to a test item, such as the sample mathematics question referenced above, criteria for determining a correct application of the relevant rule do not exist. In other words, the student is in a superposition of two states before giving an answer to the question: the student is both correct *and* incorrect. Wittgenstein contends that a rule cannot be followed privately (i.e., in the mental realm) since criteria for the correct application of the rule do not exist in the

mental realm. Instead, Wittgenstein argues that established practices determine whether a rule has been applied correctly or incorrectly:

‘Following a rule’ is a practice. And to think one is following a rule is not to follow a rule. And that’s why it’s not possible to follow a rule ‘privately’; otherwise, thinking one was following a rule would be the same thing as following it. (Wittgenstein, 2009, §202)

For example, in the sample mathematics test item referenced above, the criteria associated with the mathematical practices of algebraic substitution and calculating the cube of a number are invoked when the student proffers an answer of “64”, or a different answer, to decide if the answer is correct or incorrect. Clearly, an answer of “64” would be judged to be correct, but any alternative answer would be deemed incorrect as it would not conform with the relevant disciplinary practices. When the student responds to the test item, he or she transitions from indefinite ability with respect to the item (i.e., they are both correct *and* incorrect) to a definite ability with respect to the item (i.e., they are either correct *or* incorrect).

Therefore, the correctness, or otherwise, of a student’s response to a particular test item is judged by comparing the response with the answer that accords with the relevant disciplinary customs or practices, and the student’s ability relative to the test item is indefinite before the student responds to the item. Consequently, it can be concluded that the measurement process influences the measured value of the student’s ability since the relevant ability measurement is actually a joint property of both the student and the device used to measure the ability, i.e., the items on the test. Thus, it appears that the relevant student ability does not exist as a thing-in-itself, and that it is only meaningful to refer to it relative to the instrument used to measure it, i.e., the relevant test paper. The prospect of validity in psychological or educational measurement thus appears to be a pipe dream rather than a possibility, since it is problematic to abstract a measurement of a psychological predicate away from the instrument used to measure it (Cantley, 2015; 2017; 2019).

It is plausible that the suggested indeterminism in unmeasured psychological or educational predicates might be challenged by incorporating arguments pertaining to brain neural processes. Clearly, such processes are measurable using sophisticated technological approaches. However, the measurements would be first-person/third-person symmetrical, rather than adhering to the first-person/third-person asymmetry which Wittgenstein (2009)

contended is a hallmark of intentional psychological predicates. In the context of psychological predicates, there are no criteria for first-person ascriptions, but third-person ascriptions are based on criteria. In the case of physical phenomena, such as brain neural processes, both first-person and third-person ascriptions are based on the invocation of criteria (Malcolm, 1986).

This implies that measurements of intentional psychological predicates made using technological approaches, such as brain scans, would be incongruous with current conceptions of psychological attributes. Entities that demonstrate first-person/third-person symmetry (i.e. measurements garnered using technological methods) are not logically equivalent to those that are asymmetrical with respect to first and third-person ascriptions (i.e. psychological phenomena). However, even if one assumes that there is a systematic relationship between subjective mental states and physical brain states, a philosophical conundrum remains. The reasoning presented above regarding Wittgenstein's analysis of rule-following means that, since subjective mental states cannot be identified using criteria, i.e., they are indeterminate, subjective mental states cannot be equated to brain states. Therefore, it would appear that, both prior to and after measurement, the uncertainty in intentional psychological predicates is irreducible since it cannot be reduced by measuring physical properties of the brain because of the indeterminism in the corresponding subjective mental states.

The non-separability of psychological predicates from the instruments used to measure them, coupled with the irreducible uncertainty in unmeasured psychological predicates, casts considerable doubt on the objectivist ideals associated with positivistic psychological and educational research. Indeed, these issues may lie at the heart of the replication crisis in relation to quantitative psychological and educational research. Clearly, any attempt to improve the robustness of the methodological approaches employed in empirical psychological and educational research is commendable, but solely relying on methodological improvements, without adopting a critical stance to objectivism, may be insufficient to adequately address the replication crisis.

## **5. Conclusion**

The 21<sup>st</sup> century has heralded a significant focus on evidence-based policymaking in education (Lingard, 2013; Pellegrini & Vivianet, 2021), with large amounts of funding being

invested in educational research to identify and support the development of effective practice. Given the numerous concerns that have been expressed about the replicability of quantitative educational research, it is imperative that all plausible explanations of the replication crisis, including both theoretical and methodological issues, are thoroughly assessed through a critical lens. Taking such an approach, and dealing appropriately with the outcomes, would reduce the probability of education policy decision-making being based on dubious evidence that may potentially lead to the inappropriate deployment of significant public resources.



## References

- Bruner, J. (1996). *The culture of education*. Harvard University Press.
- Cantley, I. (2015). How secure is a Newtonian paradigm for psychological and educational measurement? *Theory & Psychology*, 25(1), 117-138.  
<https://doi.org/10.1177/0959354314561141>
- Cantley, I. (2017). A quantum measurement paradigm for educational predicates: Implications for validity in educational measurement. *Educational Philosophy and Theory*, 49(4), 405-421. <https://doi.org/10.1080/00131857.2015.1048668>
- Cantley, I. (2019). PISA and policy-borrowing: A philosophical perspective on their interplay in mathematics education. *Educational Philosophy and Theory*, 51(12), 1200-1215.  
<https://doi.org/10.1080/00131857.2018.1523005>
- Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production? *European Journal of Education*, 55(4), 618-631.  
<https://doi.org/10.1111/ejed.12417>
- Hughes, B. M. (2018). *Psychology in crisis*. Palgrave.
- Lingard, B. (2013). The impact of research on education policy in an era of evidence-based policy. *Critical Studies in Education*, 54(2), 113-131.  
<https://doi.org/10.1080/17508487.2013.781515>
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304-316.  
<https://doi.org/10.3102/0013189x14545513>
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both questionable and open research practices are prevalent in education research. *Educational Researcher*, 50(8), 493-504. <https://doi.org/10.3102/0013189x211001356>
- Malcolm, N. (1986). *Wittgenstein: Nothing is hidden*. Blackwell.
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3), 137-141. <https://doi.org/10.1027/1864-9335/a000192>

- Pellegrini, M., & Vivanet, G. (2021). Evidence-based policies in education: Initiatives and challenges in Europe. *ECNU Review of Education*, 4(1), 25-45.  
<https://doi.org/10.1177/2096531120924670>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, 22(11), 1359-1366.  
<https://doi.org/10.1177/0956797611417632>
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39(4), 202-217. <https://doi.org/10.1037/teo0000137>
- Wittgenstein, L. (2009). *Philosophical investigations: The German text with an English translation* (Edited by P. M. S. Hacker & J. Schulte; translated by G. E. M. Anscombe, P. M. S. Hacker & J. Schulte; 4<sup>th</sup> edition). Wiley-Blackwell.
- Wright, C. (2001). *Rails to infinity*. Harvard University Press.