

心理學研究的開放科學：挑戰與發展

趙軒甫*

心理學研究的目的是希望能夠對於人類的行為與心智的運作有更多的瞭解。透過研究，逐步累積與建構心理學的知識。在過往，心理學的學術性期刊，作為學術社群中知識傳播與溝通的媒介，較為容易看到的是統計顯著的、理論或實徵新穎的發現。而在近年來，可以看到國際學術社群的省思與改變。舉例來說，多個心理學領域的國際研討會開始有特別針對研究可重製性 (research reproducibility) 的主題演講、工作坊或論壇。此外，學術性期刊可以看到再現研究 (replication study)、註冊研究 (registered report)，以及其他開放科學 (open science) 的實踐的蓬勃發展。這些重大的改變，反映了學術社群對於研究可再現性的反思與相應行動。

一、危機與轉機

雖然科學研究的目的是追求知識、論文發表的目標則是知識的溝通與交流，由於論文發表與研究者的取得學位、求職、評鑑、升等、獲得計畫經費補助等產生了連結，這使得論文發表的意義變得複雜許多。Nelson 等人 (2018) 便指出 2010 年到 2012 年之間發生的數起事件的重大影響。舉例而言，Bem (2011) 發表了一系列實驗顯示人們能夠預知未來、而受到未來的事件所影響。此一發表引發了後續的質疑與反思 (如：Schimmack, 2012)，學界也進一步思考這些支持證據是如何累積的 (Nelson et al., 2018)。此外，Diederik Stapel 的研究造假，再加上其有選擇性報告部分而忽略其他研究結果 (Nelson et al., 2018) 的「有問題的研究行為」(questionable research practice)，也引起重視與反思。反思的方向包含如何提升研究的透明性與開放性。

Open Science Collaboration (2015) 的心理學研究再現的報告是另外一個重要事件。在此，我們先針對可重製性 (reproducibility)、穩健性 (robustness) 與

* 中原大學心理學系教授

再現 (replication) 進行簡介。根據 Nosek 等人 (2022)，可重製性指的是用與原先的研究報告相同的資料、相同的分析方式，能否得到與相同的結果。穩健性指的是針對同一筆資料，使用不同的分析程序能否得到一致的結果。再現指的是再進行一次研究蒐集資料，檢視能否得到與原始研究一致的發現。

Open Science Collaboration 透過多個實驗室的合作，嘗試再現了 100 篇刊載在心理學頂尖期刊的研究發現，結果顯示再現性 (replicability) 偏低。舉例而言，如果以虛無假設的統計檢定 (null hypothesis statistical testing) 的結果作為指標，只有 36.1% 的研究是可再現的；另外，倘若以原始研究的效果量 (effect size) 是否落在再現研究的效果量的 95% 信賴區間 (confidence intervals) 來估計，則可獲得 47.4% 的再現性。雖然此一再現性偏低的報告有多方的解讀、批評與討論，包含部分的再現研究的方法與原始研究有所出入等考量 (Gilbert et al., 2016)、研究發現的情境敏感度 (contextual sensitivity) 對於再現性的影響 (Van Bavel et al., 2016)，也包含以預測區間 (prediction intervals) 重新估算顯示有 77% 的研究是可以再現的 (Patil et al., 2016)，無論如何，此份報告提醒學界對於研究結果的再現性的重視。

虛無假設的統計檢定本身，也是學界開始檢討的目標。過往的研究，通常是基於統計檢定所得到之 p 值是否小於 0.05 的判準，來做出是否達到顯著、能夠拒絕虛無假設的決策。然而，此一決策的基礎有可能被誤用。舉例而言， p -hacking 指的是由於研究者有過度的自由作為，藉由如何移除界外值、要納入哪些變項進行分析、是否繼續徵求參與者、是否把兩個或兩個以上的情況合併等考量進行多方嘗試，直到顯著。這是有問題的研究行為，可能導致錯誤地拒絕虛無假設 (Simmons et al., 2011)。Head 等人 (2015) 藉由分析文獻中 p 值的分布，指出多個學術領域可能有 p -hacking 的存在，顯示這是值得重視的問題。

二、解決方案的推動與嘗試

針對上述之研究再現性等挑戰，開放科學的推動與實踐在學術社群日益蓬勃，期許能夠讓研究更透明與開放、也讓研究成果更能夠被再現。以下簡述近年來學界的因應行動。

促進透明性與開放性 (Transparency and Openness Promotion) 的指引。開放科學中心 (Center for Open Science) 提出了此指引，希望能夠提升研究的透明性與開放性 (Nosek et al., 2015)。此指引涵蓋了透明性、預先註冊 (preregistration)、再現等取向。此指引已可見於多本期刊之投稿規範，而展現其影響力。

預先註冊。預先註冊指的是在研究開始之前，就預先陳述研究假設、研究設計與分析計畫。後續研究者便依循此一規劃執行研究。其實務考量可參考 Nosek 等人 (2018)。

預先註冊的好處是由於研究的假設、設計與分析均已事先規劃好，可以避免由於研究者的過高的自由作為而出現有問題的研究行為，如：p-hacking、HARKing (hypothesizing after the results are known; Kerr, 1998；得到研究結果之後才提出研究假設)。

註冊研究。與預先註冊相似，註冊研究需要事前提供研究假設、研究設計與資料分析計畫。此外，期刊端會預先針對預先註冊的研究計畫進行同儕審查，倘若被接受，便會是原則性接受 (in principle acceptance) 的狀態。後續執行研究，最終結果 (如：是否達到統計顯著) 不會影響是否被刊登。針對註冊研究，目前已經有明確的指引可以參考 (<https://osf.io/pukzy>)。透過此一設計，可以減少學術期刊的發表偏誤 (publication bias) 或者是抽屜問題 (file drawer problem)，也就是顯著的結果容易被發表而不顯著的結果只能被擱置收藏的情況。此外，藉此也可以避免對於統計顯著的過度追求。

統計方法的考量。針對過往的過於依賴虛無假設的統計檢定並採用 p 值是否小於 0.05 的標準，學界開始思考是否有更好的方法。首先，新統計學 (new statistics) 強調了效果量、信賴區間、後設分析 (meta-analysis) 與開放科學的使用。新統計學建議相較於過度依賴 p 值，針對所研究的現象估計其效果量，並藉由信賴區間來呈現此估計的精準程度，會是更合宜的方式。此外，針對同一個研究問題，可能有多個實驗室、多個研究的執行，而有多組研究結果。藉由後設分析，可以統整多組研究結果以提供更為精準的估計，並檢驗可能的調節變項 (Cumming, 2014)。

針對統計檢定的 p 值，有學者主張將 p 值的標準由原先的 0.05 調整為更嚴格的 0.005 (Benjamin et al., 2018)。另一方面，Lakens 等人 (2018) 主張清楚說明與佐證研究設計與統計方式的考量更為重要。

最後，相較於上述古典統計 (Frequentist) 取向的考量，貝氏統計 (Bayesian) 取向在學界逐漸獲得重視。值得注意的是，貝氏統計的使用仍有需要留意與反思之處 (如：Cumming, 2014; Simmons et al., 2011)。

三、開放與可重製科學在臺灣

國內的學術社群也不會自外於國際學術社群的變化。以下從國內研究者的

再現研究、多個團體的行動、學會與學術期刊的推動、與課程的規劃等面向進行簡介。

在臺灣，有時因為研究者希望從再現已經發表的重要發現啟始，進一步進行延伸，而有再現研究的執行。此再現的過程有可能遭遇到失敗。失敗的成因很多，可能意涵有重要的調節變項值得進一步研究 (Van Bavel et al., 2016)，也有可能反映該現象本身需要進一步檢視。舉例而言，顏志龍針對其再現恐懼管理理論 (terror management theory) 研究的一連串失敗，分析並提出該理論不一定成立的反思 (顏志龍, 2012)。

除了上述的反思，國內學者也開始積極且自發地推動開放科學。舉例而言，陳紹慶與來自於國內多所大學的研究者共同設立了台灣心理科學協作群 (Taiwan Collaboration for Psychological Scientific Research; 網址為 <https://tcpsr.netlify.app/>)，致力於在臺灣心理學研究的開放科學的推廣與交流。此外，國內學者與國際學術界多方合作。舉例而言，陳紹慶擔任了 The Society for the Improvement of Psychological Science 此一以促進心理科學的方法與實務為主要目標的國際學會的執行委員，並與龔俊嘉共同參與了該學會提升國際參與的工作小組報告的制訂 (Steltenpohl et al., 2021)。最後，國內目前已經有兩個團體 (臺北醫學大學與台灣心理科學協作群) 自發的籌組了呼應 ReproducibiliTea 這項旨在討論與推動開放科學的期刊讀書會。這些都展現了國內學者在臺灣、也在國際間的積極參與。

國內的學會與學術性期刊也開始相關之推廣。舉例而言，台灣心理學會透過與台灣心理科學協作群的成員合作，推出網路實驗交換平台工作坊、可重製性研究黑客松工作坊這些呼應了開放科學的重要推廣；也設立了 Effect size 效果量工作坊，呼應了新統計學的取向。而在學術性期刊，以《中華心理學刊》為例，已經於民國 107 年公告新增註冊研究之投稿文類 (姜定宇等, 2018)，並進一步在民國 112 年設立註冊研究專刊公開邀稿，希望推動開放科學的進展。雖然國內的推動沒有國際學術社群快速，期許透過這些行動，能夠讓研究者們有更多機會瞭解近年來在國內外學術社群的變化。

此外，近年亦已有開放科學相關的課程於大學開設或規劃 (如：張仁和與李怡青在臺灣大學心理學系所開設的「社會心理學研究法」課程；台灣心理科學協作群著手規劃了心理學研究可重製教程 (<https://tcpsr.netlify.app/course/>)。課程的開設是很重要的開始，因為傳統的心理學的方法學課程與教科書尚未充分涵蓋開放科學相關的新進展。我們期待在教學、研究、期刊多方的共同努力之下，有助於開放科學的相應思考與實踐。

致謝

感謝張仁和、陳紹慶與另外一位學者對於本文的寶貴建議，謹此致謝。

參考文獻

- 姜定宇、趙軒甫、姜忠信、李思賢 (2018)。〈註冊研究：中華心理學刊新增投稿文類〉，《中華心理學刊》60 卷 1 期，頁 i-iii。https://doi.org/10.6129/CJP.201803_60(1).0000。
- 顏志龍 (2012)。〈渴望心理學的科學革命：從「恐懼管理理論」的研究歷程反思心理學研究之現狀〉，《本土心理學研究》38 期，頁 103-129。
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407-425. https://doi.org/10.1037/a0021524
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Fehr, E., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10. https://doi.org/10.1038/s41562-017-0189-z
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. https://doi.org/10.1177/0956797613504966
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson. T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, 351(6277), 1037. https://doi.org/10.1126/science.aad7243
- 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), e1002106. https://doi.org/10.1371/journal.pbio.1002106
- 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
- Lakens, D., Adolphi, F., Albers, C., Anvari, F., Apps, M., Argamon, S., Baguley, T., Becker, R., Benning, S., Bradford, D., Buchanan, E., Caldwell, A., Van, C. B., Carlsson, R., Chen, S., Chung, B., Colling, L., Collins, G., Crook, Z., ... Zwaan, R. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168-171. https://doi.org/10.1038/s41562-018-0311-x
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69, 511-534. https://doi.org/10.1146/annurev-psych-122216-011836
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, 348(6242), 1422-1425. https://doi.org/10.1126/science.aab2374
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), Article aac4716. <https://doi.org/10.1126/science.aac4716>
- Patil, P., Peng, R. D., & Leek, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*, 11(4), 539-544. <https://doi.org/10.1177/1745691616646366>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Steltenpohl, C. N., Montilla Doble, L. J., Basnight-Brown, D. M., Dutra, N. B., Belsky, A., Kung, C.-C., Onie, S., Seernani, D., Chen, S.-C., & Burin, D. I. (2021). Society for the Improvement of Psychological Science global engagement task force report. *Collabra: Psychology*, 7(1), 22968. <https://doi.org/10.1525/collabra.22968>
- Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, 113(23), 6454-6459. <https://doi.org/10.1073/pnas.1521897113>