

**Serial number** 0356

**Title** Examining the Legitimacy of Panel Peer Review to Assess Impact

**Session** The impact of 'impact' on UK higher education (Filippakou)

**Submitter** Dr. Ourania Filippakou  
Author Dr. Gemma Derrick

## **Examining the Legitimacy of Panel Peer Review to Assess Impact**

### **Abstract**

As an ambiguous object in evaluation, and one that is of increasing political and social prominence in research evaluation, Impact places new pressures on the traditional peer review mode of evaluation. Traditionally, peer review has concentrated on assessing common, more habituated norms of research excellence, but the inclusion of Impact as a formalised criterion, questions its legitimacy as the “gold-standard” tool of evaluation. Likewise, the consideration of peer-review as a group endeavour, rather than solely concentrating on assessing the reliability of its outcomes (Bornmann and Daniel 2005), has shifted how the science system values its outcomes, and the validity of these peer (or expert) decisions. This paper examines the legitimacy of peer review as an evaluative tool for Impact criteria, in the face of common drawbacks associated with group dynamics such as Groupthink.

### **Outline**

Realising the societal gains from publicly funded research requires a workable evaluation model that includes how research influences society outside academia (Impact). This paper considers the extent to which a demand for ambiguous aspects of research excellence as criteria, as is the case of Impact, are sufficiently accommodated within old tools of research evaluation, such as peer review.

Opposition to formally including an Impact criterion point to the questionable use of qualitative case studies and subjective judgements through peer review, being time consuming, unreliable and unnecessarily expensive (Martin 2011, Bornmann 2012, Bornmann 2013). The majority of what we currently know about peer review panel processes is based on the interpretation of traditional notions of research excellence (Chubin and Hackett 1990, Chubin 1994, Langfeldt 2001, Langfeldt 2004, Lamont 2009, Mow 2010, Olbrecht and Bornmann 2010). For these criteria, a certain level of subjectivity and inefficiency is acceptable as the decisions are perceived to be governed by a shared set of academic peers, and/or experts. Described as a “*system(s) of institutionalized vigilance*” (Merton 1973) in the self-regulation of the research community, panel peer review reaches a common judgment through, what Olbrecht and Bornmann (2010) described as mutual social exchange, where the final judgment is based on the common judgment of all evaluators, that are considered academic peers. In addition, how people work together towards a common goal, or “group behaviour”, lies behind much of what we currently understand about peer review panels.

In recent research regarding peer review, the change of focus from the attributes of submissions, to one that focuses on the way that academic groups arrive at a consensus, is revolutionary but there is a lack of empirical evidence of how criteria are operationalised in evaluative practice (Langfeldt 2001, Langfeldt 2004, Lamont 2009, Arensbergen 2014, Pier, Raclaw et al. 2017). As an evaluative object, Impact as a formalised criterion is ambiguous in its operationalisation within peer review panels. Conceptually, researchers are less habituated to the concept of research excellence beyond academia and this influences their approach to its evaluation. Insights from the practice of other, similarly

ambiguous criteria such as with “interdisciplinary research” have shown that panels tend to adopt a conservative approach during assessment (Langfeldt and Brofoss 2005, Luukkonen 2012). This begs the question of how panelists, when faced with an untested, ambiguous concept of Impact as an evaluation criteria, approached its assessment in a way that guaranteed the robustness of peer review, with the necessary objectivity of assessing an object that exists beyond academic experience and value. By examining this process more closely, the suitability of peer review as a tool for the evaluation of Impact is questioned.

Using a combined qualitative interview structure (Derrick and Samuel forthcoming), this paper examined the group-behaviour of evaluators from the 2014 UK Research Excellence Framework Main Panel A (which included biomedicine, health, and clinical disciplines) and associated sub-panels tasked with evaluating the “Impact” criterion. By examining incidents of group behaviours that were isolated through direct accounts from the participants, and accounts of evaluative practice of participants individually and at the group-level, this paper examines the extent that common drawbacks of group processes risk the legitimacy of peer review as an evaluative tool for Impact.

For evaluation panels, mediating this criterion against the backdrop of inexperience (Derrick and Samuel 2014, Samuel and Derrick 2015, Derrick in press, Derrick and Samuel in press) as well as the highly political nature of Impact and its crucial role in determining the future of research funding, was difficult. This led to a higher reliance of evaluators on their personal and professional experiences as well as their individual beliefs about the relationship between notions of research excellence and Impact, to guide the assessment of this new criterion (Derrick in press). Further, to achieve the type of group consensus around Impact vital for group, participatory peer review processes, panel members navigated the interplay of differing values, academic roles, reputations and as well as the perceptions of other evaluators during the evaluation.

When we consider peer review as a group activity, rather than guided by a single notion of research excellence, its role as “gold standard” evaluation tool appears questionable for Impact. Groupthink in particular, as a common drawback of group dynamics, is when groups become preoccupied with the goal of reaching a consensus rather than reaching the “correct” decision (Comer 1995, Esser 1998, Gallo, Sullivan et al. 2016). Therefore groups that are at risk of groupthink are not motivated to detect weaknesses in the decisions, nor critically appraise alternative decisions, instead reaching a state of collective delusion (Chong 2013) or mutual construction (Pier, Raclaw et al. 2017). Its existence can reduce the legitimacy of evaluation outcomes, process, as well as jeopardise the future of Impact as a formal research criterion, and the gold standard associated with peer review as an evaluation tool (Bailar 2011, Arensbergen 2014, Academy. September 2007). Groupthink is difficult to pinpoint in practice as it requires a separate, subjective judgement regarding the “correct” decision. The nature of peer review, and its basis on the democratic deliberation between experts, means that the evaluation outcomes are communally accepted as correct, as long as the process is considered fair (Chubin and Hackett 1990, Chubin 1994, Chong 2013). Instead, this paper questions the way in which the group reached a decision by exploring the nature of any shortcuts taken, as well as the opportunity lost from not exploring dissenting ideas around Impact as an evaluation object. To this end, this paper explores instances where the probability of groupthink leading the evaluation were higher through the exploration of the group-based errors in decision making of social loafing, shelving and satisficing. It also examines how these shortcuts influenced the generation of evaluation outcomes for the Impact criterion and by doing so explores the suitability of using traditional peer review for the assessment of the societal Impact of research.

## References

- Arensbergen, P. v. (2014). *Talent Proof. Selection Processes in Research Funding and Careers*, Den Haag: Rathenau Instituut.
- Bailar, J. (2011). "Reliability, fairness, objectivity and other inappropriate goals in peer review " *Behavioral and Brain Sciences* **14**(01): 137-138.
- Bornmann, L. (2012). "Measuring the societal impact of research." *EMBO Reports* **13**(8): 673-676.
- Bornmann, L. (2013). "What is the societal impact of research and how can it be assessed? A literature survey." *Journal of the American Society of Information Science and Technology* **64**(2): 217-233.
- Bornmann, L. and H.-D. Daniel (2005). "Committee peer review at an international research foundation: predictive validity and fairness of selection decisions on post-graduate fellowship applications." *Research Evaluation* **14**(1): 15-20.
- British Academy. (September 2007). *Peer review: the challenges for the humanities and social sciences*, A British Academy Report, The British Academy.
- Chong, P. (2013). "Legitimate judgement in art, the scientific world reversed? Maintaining critical distance in evaluation." *Social Studies of Science* **43**(2): 265-281.
- Chubin, D. E. (1994). "Grants peer review in theory and practice." *Evaluation Review* **18**(1): 20-30.
- Chubin, D. E. and E. J. Hackett (1990). *Peerless science: Peer review and US science policy*. Albany, State University of New York Press.
- Comer, D. R. (1995). "A model of social loafing in real work groups." *Human Relations* **48**(6): 647-667.
- Derrick, G. E. (in press). *The Evaluators' Eye: Impact Assessment and Academic Peer Review*. London, UK, Palgrave Macmillan.
- Derrick, G. E. and G. Samuel (in press). "The future of societal impact assessment using peer review: Pre-evaluation training, consensus building and inter-reviewer reliability." Palgrave communications.
- Derrick, G. E. and G. N. Samuel (2014). "The impact evaluation scale: Group panel processes and outcomes in societal impact evaluation."
- Derrick, G. E. and G. N. Samuel (forthcoming). "The in-vitro approach: A qualitative methodology for investigating peer review panel debates."
- Esser, J. (1998). "Alive and Well after 25 Years: A Review of Groupthink Research." *Organizational Behavior and Human Decision Processes* **73**(2/3): 116-141.
- Gallo, S. A., J. H. Sullivan and S. R. Glisson (2016). "The influence of peer reviewer expertise on the evaluation of research funding applications." *PLoS One* **11**(10): e0165147.
- Lamont, M. (2009). *How Professors Think: Inside the curious world of Academic judgement*. Cambridge, Massachusetts, Harvard University Press.

Langfeldt, L. (2001). "The decision-making constraints and processes of grant peer review, and their effects on the review outcome." *Social Studies of Science* **31**(6): 820-841.

Langfeldt, L. (2004). "Expert panels evaluating research: decision-making and sources of bias." *Research Evaluation* **13**(1): 51-62.

Langfeldt, L. and K. E. Brofoss (2005). "Evaluation of the European young investigator awards scheme."

Luukkonen, T. (2012). "Conservatism and risk-taking in peer review: Emerging ERC practices." *Research Evaluation*: rvs001.

Martin, B. R. (2011). "The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster?" *Research Evaluation* **20**(3): 247-254.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*, University of Chicago press.

Mow, K. E. (2010). *Inside the black box: research grant funding and peer review in Australian research councils*, Lambert Academic Publishing.

Olbrecht, M. and L. Bornmann (2010). "Panel peer review of grant applications: what do we know from research in social psychology on judgement and decision making in groups?" *Research Evaluation* **19**(4): 293-304.

Pier, E. L., J. Raclaw, A. Kaatz, M. Brauer, M. Carnes, M. J. Nathan and C. E. Ford (2017). "'Your comments are meaner than your score': score calibration talk influence intra- and inter-panel variability during scientific grant peer review." *Research Evaluation* **26**(1): 1-14.

Samuel, G. N. and G. E. Derrick (2015). "Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014." *Research Evaluation* **24**(3): 229-241.

