Research Evaluation, 2025, **34**, rvaf030
https://doi.org/10.1093/reseval/rvaf030
Article

Peer review across borders: benefits and challenges of international review panels in research funding organizations

Helen Peterson[1,*], **and Liisa Husu**[1]

[1]School of Humanities, Education and Social Sciences, Örebro University, SE-701 82, Sweden

*Corresponding author. Helen Peterson, School of Humanities, Education and Social Sciences, Division of Gender Studies and Sociology, Örebro University, Fakultetsgatan 1, Örebro, SE-701 82, Sweden. E-mail: helen.peterson@oru.se.

Abstract

Peer review by external experts is widely recognized as a legitimate and trustworthy academic practice, essential for ensuring the quality and rigor of research, providing more objective and less impartial assessments, and promoting transparent decision-making in science and academia. Research Funding Organisations (RFOs) usually rely on some form of peer review to evaluate the scientific quality of research proposals to allocate their limited resources. The peer review system is, however, also associated with several weaknesses, such as risks for bias and conflict of interest. This article explores the implications of replacing National Review Panels (NRPs) with International Review Panels (IRPs) in a national RFO, examining how this shift may impact the peer review process. Drawing on semi-structured interviews with staff from a national RFO in a European country and members of its IRPs, the article provides a nuanced analysis of both the potential benefits and challenges with substituting NRPs with IRPs. The results highlight how IRPs increase the distance between applicants and reviewers, which benefits the impartiality of the process. Nevertheless, this distance needs to be balanced by domestic panel members, chairs or research officers possessing appropriate knowledge of the local academic context, culture and structure. IRPs also introduce a greater diversity of perspectives into the assessments of applicants, which may promote objective and balanced assessments. The diversity may however also lower inter-reviewer reliability, and, in turn, complicate calibration practices and hinder the development of informal deliberative norms during the process of reaching decisions and consensus.

Keywords: Peer Review; Research Funding Organisations; International Review Panels; Research Evaluation; Calibration; Inter-reviewer Reliability; Conflict of Interest.

1. Introduction

Peer review is a key academic practice in which researchers with expertise in relevant fields critically assess the work of their peers before it is published or funded, to ensure that it meets high academic standards of quality and rigor (Hirschauer 2010). Many Research Funding Organisations (RFOs) use peer review as a key method for assessing the scientific quality of grant applications and the past performance of the applicants (Gallo, Sullivan and Glisson 2016; Recio-Saucedo et al. 2022). Key predefined criteria for this assessment often include originality and innovativeness of the research, the qualifications of the research team, feasibility, and potential impact of the project to advance the field (Feller 2013; Oxley and Gulbrandsen 2025). After reviewing the proposals, peer reviewers usually provide numeric scores and written assessments which contribute to a ranking of proposals (Pier et al. 2018; Meadmore et al. 2020; Hren et al. 2022). Ultimately, decisions on which projects to fund are made by the RFOs, based on these scores and rankings (Husu and Peterson 2022).

Although the peer review system is widely adopted to ensure an objective and fair assessment, it has also been recognized as influenced by conflicts of interests and reviewers' bias (Langfeldt, Reymert and Svartefoss 2024; Gould et al. 2025). Roumbanis (2021: 358) describes the evaluation of proposals as 'a rather arbitrary and biased practice', whereas Shaw (2024: 7, emphasis in original) more strongly argues that 'peer review is *intrinsically* biased', and Meadmore et al. (2020: 2) similarly declare that 'peer review is inherently subjective'. In addition to individual peer reviews, RFOs therefore often convene review panels or committees composed by expert reviewers to perform a *collective* peer review, thereby potentially mitigating any individual biases (Bendiscioli 2019; Meadmore et al. 2020; Peterson and Husu 2023). During these panel meetings the reviewers discuss their individual assessments and scoring in a transparent manner, in order to merge them into a common group decision that all panel members can support and stand by (cf Lamont 2009; Huutoniemi 2012; Kaltenbrunner and de Rijcke 2019).

While RFOs often rely on National Review Panels (NRPs) constituted by primarily experts from the country where the RFO is located, they can also engage International Review Panels (IRPs), constituted by experts from outside the funding organization's home country. Langfeldt and Kyvik (2011) already observed that the role of research evaluator had begun to change, becoming increasingly international over the preceding two decades. According to them, cross-border peer review 'serves to enhance review quality as well as to avoid conflicts of interest' (Langfeldt and Kyvik 2011: 208). In a more recent contribution, Shaw (2024) argues that generally, without focusing specifically on cross-border review, diversifying review panels can help mitigate bias, based on the assumption that more heterogeneous committees are less prone to shared blind spots or systematic bias. Győrffy, Herman

and Szabó (2020) explore strategies to enhance objectivity in peer review, suggesting that international reviewers may offer more 'objective and unbiased evaluation' due to their lack of national ties and 'independent overview of the field' (12). However, asides from these brief mentions of cross-border peer review in studies primarily focused on other issues, the topic of IRPs remains largely unexplored.

Gould et al. (2025) observe that, although peer review of journal manuscripts has been extensively studied, comparatively less attention has been given to the peer review of grant applications (cf Hren et al. 2022). Nevertheless, there is a growing body of research critically investigating the integrity of the peer review system used by RFOs (cf e.g. Feller 2013; Bendiscioli 2019; Győrffy, Herman and Szabó 2020; Roumbanis 2022; Acker, Ylijoki and McGinn 2024; Langfeldt, Reymert and Svartefoss 2024). Notwithstanding, this article identifies, and addresses, a significant gap in the literature on peer review in grant assessment: the use of IRPs in RFOs.

This article draws on qualitative interviews with staff members in a national, European RFO and the members of their IRPs. The aim is to provide an in-depth and nuanced analysis of how internationalization of review processes can impact peer review as an academic practice and the dynamics of review panels. The main research question addressed in the article is: What *benefits* and *challenges* do members of review panels and RFO staff identify with international review panels? The findings contribute to a deeper understanding of the complexities involved in cross-border peer review. By shedding light on the operational realities of IRPs, the article offers valuable insights for RFOs seeking to refine their processes and procedures in what has been described as an increasingly international research landscape (cf Stensaker et al. 2010; Langfeldt and Kyvik 2011; De Wit and Hunter 2015).

The article continues with an introduction to the framework for analysis which is followed by a section which describes the method and empirical data more in detail. After that the results and analysis are presented. The article ends with a final discussion.

2. Previous research

This article builds on ethnographic research that investigates peer review panels as a complex 'micro-political process of collective decision-making' (Lamont 2009: 246). Such studies have highlighted how this process unfolds through social and interactional practices such as persuasion, negotiation, strategic voting, disciplinary gatekeeping, horse-trading, and the formation of allegiances and alliances through which panellists exert influence to advance personal or collective interests (Langfeldt 2001, 2004, 2006; Lamont 2009; Huutoniemi 2012; Roumbanis 2019, 2022). According to Lamont these practices are 'unavoidable parts' (Lamont 2009: 156) of the 'complicated dynamics of group evaluation' (Lamont 2009: 110; cf Philipps 2024). They are essential because reviewers 'often arrive at very different judgements' (Hug and Ochsner 2022: 104) and because panel meetings 'are not restricted to rational deliberations' (Philipps 2024: 6).

Peer review panels are thus characterized by challenges with incommensurability, disagreement effect and low or weak inter-reviewer reliability (Huutoniemi 2012; Sattler et al. 2015; Roumbanis 2022). Seeber et al. (2021: 350) define low inter-reviewer reliability as 'a high level of disagreement between reviewers in the score assigned to the same research proposal'. This type of disagreement can be explained by different opinions about what constitute excellent science or different use of the rating scale (Pier et al. 2017, 2018). A high level of inter-reviewer reliability is desirable but difficult to achieve, especially with regards to how to interpret and use uncertain evaluation criteria such as assessing impact and societal outcomes from research (Derrick and Samuel 2017; Oxley and Gulbrandsen 2025).

Peer review meetings therefore provide a space for reviewers to calibrate, equalize, adjust, and merge individual assessments into a collective decision that all panel members can support and stand by (cf Lamont 2009; Huutoniemi 2012; Roumbanis 2017). This collective process enables the diverse perspectives, pre-conditioned expectations, standards and criteria that reviewers bring to be made transparent and open to (re)negotiation (Kaltenbrunner and de Rijcke 2019). Gould et al. (2025: 5) define calibration as a consensus discussion during which the members of the review committee reach 'common ground' and tune 'their individual interpretations of the application rating system to promote consistency and fairness' in how they rated each grant application. Huutoniemi (2012) describes the panellists' calibrating activities as involving adjusting 'their individual senses of quality to a group standard in order to form a concerted evaluation' (910).

To navigate the inherent subjectivity of peer review and to manage calibration processes, panellists develop customary rules of deliberation, that is, informal, often unspoken norms that guide both how proposals are evaluated and how panellists interact (Lamont 2009). In the face of ambiguous criteria and disciplinary differences, customary rules help panellists manage subjectivity, minimize conflict, and uphold the legitimacy of the process. Examples of these customary rules include striving for consensual decisions, persuading others through reasoned argumentation, rejecting overt bias and self-interest, ensuring equal voice among panellists, and deferring to disciplinary expertise when appropriate (Lamont 2009; Gallo, Sullivan and Glisson 2016). One especially significant customary rule is cognitive contextualization, which requires panellists to 'use the criteria of evaluation most appropriate to the field of discipline of the proposal under review' (Lamont 2009: 106; cf Laudel 2024). This customary rule is particularly crucial in interdisciplinary panels, where reviewers must navigate differing standards, histories, and methodologies to ensure fair evaluations across disciplines. When followed, these customary rules can contribute to regulating the internal dynamics of peer review meetings (Lamont 2009; Philipps 2024).

Lamont (2009) distinguishes these rules and practices, internal to the peer review process, from factors external to the peer review process itself, but which still can influence the outcomes of panel meetings and the fairness and legitimacy of panels' decisions. These external or extraneous factors originate or exist outside of the review process and involve panellists' preexisting networks and reputation, and different types of cognitive bias, such as homophily, that is, the tendency to favour work resembling one's own, which can skew evaluations towards familiar paradigms and disadvantage novel or unconventional research (Lamont 2009; cf Schmaling et al. 2024). Gender biases, such as attribution bias, where women's contributions are undervalued, and

same-group bias, has also been documented to subtly influence perceptions of quality and merit (Tricco et al. 2017). Many of these external factors are typically addressed in RFOs formal guidelines aimed at promoting fairness during the meetings and reducing partiality, bias, conflict of interest, structural inequalities and uneven personal influence and power dynamics in panel meetings (Husu and Peterson 2022).

Lamont's (2009) analysis of the U.S. context highlights how institutional size and geographic dispersion shape peer review dynamics and facilitates practices like blind review, helping to anonymize proposals and reduce bias. In the smaller, more homogeneous European higher education systems, personal networks and institutional affiliations may play a more visible role in peer review. Incorporating international panellists can help mitigate the influence of preexisting networks and reputation, but also broaden evaluative perspectives and reduce local biases, promoting diversity in scholarly assessments and 'facilitate complementary assessments' (Hug and Ochsner 2022: 106). At the same time, the inclusion of reviewers from varied academic cultures can complicate the development of shared customary rules. There is, however, a notable absence of qualitative research specifically addressing the dynamics of IRPs in the context of grant allocation. One of the few studies to touch on this topic quantitatively is Győrffy, Herman and Szabó (2020), whose statistical analysis on reviewer performance in grant allocation indicates that international reviewers performed less efficiently than national ones (measured as the correlation between reviewer scores and subsequent publication performance). They attribute these findings to the possibility that international reviewers apply standards from their own countries, leading to inconsistent evaluations (cf Huutoniemi 2012). The authors however call for further research to clarify the causes of this discrepancy.

With this previous research as an analytical frame, the article addresses this research gap and examines how members of IRPs navigate the formal requirements of peer review while grappling with the challenges of building shared informal norms across diverse academic and cultural contexts.

3. Method and empirical material

3.1 Empirical case

This article draws on a qualitative case study focusing on a European, national RFO (henceforth referred to as 'the RFO') and the peer review system in place at this RFO (cf Yin 2009). The case study was part of a large research project and included an initial policy analysis, finalized during early 2022, focusing on the written policies framing the system and reported on elsewhere (Husu and Peterson 2022). This was followed by qualitative interviews with RFO staff and reviewers which this article draws on.

The case study design provided a rich and detailed understanding of policies, processes and practices related to the implementation of this peer review system which centred around ~25 peer review panels consisting of a mix of domestic and international experts taking on the role as peer reviewers of grant applications. The panels were divided into groups based on three broad fields: Humanities, Arts and Social Sciences (henceforth abbreviated HS), Life Science (abbreviated LS) and Physical Sciences and Engineering (abbreviated PE). Each field consisted of between 6–10 review panels constituted by experts from adjacent disciplines such as history, archaeology, ethnology and anthropology (forming one of the HS panels); evolutionary and environmental biology, ecology, population biology, biodiversity (forming one of the LS panels) and mathematics, computer science, physics and statistics (forming one of the PE panels). The panels were thus to a certain extent multidisciplinary.

In general, the panels consisted of 14 to 15 members, although some panels were smaller and others larger. In 2022, only 19% of all the panel members were domestic but the composition of the panels differed slightly, with some panels consisting of only international experts. On average each panel had two to three domestic panel members. Each panel had a chair, and the proportion of domestic panel chairs was higher; 52%. Only 28% of all panellists and chairs were women. The panels were used to evaluate applications submitted to calls in around a dozen different funding schemes, most of them national calls. Although the panels had previously convened in person in the city of the RFO's offices, their meetings switched to digital during the pandemic (cf Peterson and Husu 2023).

The RFO follows a two-stage decision-making process across two panel meetings held months apart. Before the first meeting, panel members individually review and score a shorter version of the applications, including the applicant's CV, and submit their evaluations online. Once submitted, they can view the other reviewers' scores and comments. In the first meeting, all applications are discussed, and the panel collectively decides which move forward to the next stage. In the second stage, full applications are reviewed by all panel members and in addition by ad-hoc external experts, selected for their particular expertise on the topic of the application. These external experts submit only written reviews and do not participate in panel meetings. During the final meeting, the panel considers external reviews, discusses rankings, and agrees on which projects to recommend for funding.

3.2 Sampling strategy

A total of 32 interviews were conducted with RFO staff and reviewers. Five staff members were interviewed during January and February 2022. The staff members were selected due to their key positions in the organization which provided them with good insights into the policy development in the RFO and the peer review system. Three women and two men were interviewed. They were contacted directly by the members of the research project with an invitation to participate in an interview and all accepted.

In addition, interviews were also conducted with 27 members of the RFO's IRPs during summer and fall of 2022. They had all been involved in assessing applications for the same call, and they participated in panel meetings during the spring of 2022. After the RFO asked all their panel members for permission to share their contact information with the members of the research project, 20% agreed to this. The subsequent selection of panellists and panel chairs was made by members of the research project without involvement of the RFO and was guided by an intention to include panel members from all three different fields (HS, LS and PE) and from different panels in all three different fields, but also a desire to include equal number of women and men as far as possible. A total of 27 panellists and 12 panel chairs were contacted via mail and telephone by project members with an invitation to participate in the project. While the majority

agreed to be interviewed, some declined due to lack of time or did not respond to the request.

3.3 Participants

The efforts to recruit participants from the IRPs resulted in that interviews were performed with seven panel chairs (4 women and 3 men) and 20 panellists (9 women and 11 men). These 27 panel members belonged to 11 different review panels: 5 HS panels, 3 LS panels and 3 PE panels. Nine panel members from each of the three fields (HS, LS, PE) were interviewed.

The age of the interviewees ranged from 37 to 73. While all 5 of the interviewed staff members were from the country of the RFO, only 8 of the 27 interviewed panel members were domestic. Five of the eight domestic panel members were panel chairs. The remaining 19 panel members interviewed represented 15 different nationalities, from Europe and beyond. Out of the 27 panel members, a majority, 18, were professors and the remaining 9 were associate professors or equivalent. Their background was in 24 different disciplines, including: archaeology, astronomy, biochemistry, economics, environmental sciences, food science, neurobiology, immunology, law, literature, psychology, sports science, and veterinary pharmacology. Most of them had been expert reviewers or panel members for this particular RFO previously and for several different calls, and all of them had experiences from being a reviewer for other RFOs, both international and national RFOs in other countries.

No more details about the informants will be revealed, in order to preserve anonymity in accordance with research ethical considerations.

In the article the informants are referred to with a case number. For interviews with staff members the case numbers S1, S2 etcetera are used. Panel members (panellists and panel chairs) are referred to with the field of the panel, HS, LS or PE, followed by the number of the panel, followed by the case number for that specific panel member, according to the following: HS1-P1, which reads Panellist number 1, in panel 1 in the field of Humanities, Arts and Social Science (HS). Panel members with the role of panel chair are instead referred to as PC, for example: HS2-PC. The numbers referring to the panels are different from the actual numbers assigned by the RFO.

3.4 Semi-structured interviews

All 32 interviews were of qualitative, semi-structured character. They were all performed using digital communication platforms and they were all recorded using this technology, after permission was granted by the interviewees. All participants had a high degree of technological literacy and were thus familiar with how to use the digital platform and comfortable with it (cf Maldonado-Castellanos and Barrios 2023). The interviews lasted between 1 and 2 h and were performed in English, which was not the native language for the majority of the informants or the interviewers. Most of them were however fluent in English. One of the staff members requested a colleague to act as a translator which was agreed. The majority of the interviews were conducted by the authors of this article.[1]

Three different interview protocols were used: one for the interviews with the RFO staff, one for panel members, and finally one for panel chairs. These interview protocols guided the interviewee into certain topics, but the structure of the interview was fluid, allowing for the interviewee to narrate more freely on the topics. The interviews with the RFO staff included questions on policy development, policies in place and their implementation, and processes and procedures related to the peer review system. The interview protocols for panel members (panellists and chairs) included the following topics: background of the informant; becoming a panel member (reviewer/chair) for the RFO; reviewing and evaluating applications in practice; interpreting criteria and identifying excellence; structure and procedure of the panel meeting; discussions, dynamics and reaching agreements during the panel meeting; particularities of digital panel meetings; unconscious bias; gender and diversity aspects in research and peer reviewing.

The interviews provided rich and detailed accounts of being a RFO staff member, panellist and chair. They captured both more general information about the work in the panels, how the evaluation and decision-making processes were organized and about the interviewees' own understanding and experiences.

3.5 Thematic analysis

All interviews were transcribed and analysed by the authors using a thematic analysis which followed a six-phase analytic process, starting with a familiarization of the data and a thorough reading through of the interview transcripts and a detailed (albeit not line-by-line) coding (cf Terry et al. 2017). This procedure involved interpreting the data in a meaningful way to generate, identify and construct key themes which captured important patterns and trends (but also deviations, variations, and nuances) in the data. The major themes reappeared in most of the interviews. Identifying these themes facilitated a deeper understanding of the peer review process in the RFO.

Guiding the analysis and the construction of these themes were the main research questions of the overall project and the themes of the interview protocols, focusing on individual evaluation practices, collective decision-making, the role of the chair and how to identify excellence (cf Fereday and Muir-Cochrane 2006). In addition, some themes were constructed in a more inductive manner, using the empirical data as the starting point. The topic of this article is an example of such a theme, focusing on the different aspects related to the inclusion of international members in the review panels. Although the research project did not initially include a focus on IRPs it emerged as a highly relevant analytical theme during the fieldwork. The codes which together constituted this as a major theme were divided into two sub-themes: advantages and challenges with IRPs. These advantages and challenges were further explored and examined using the analytical framework presented in this article.

Because the 27 interviewed panel members belonged to 24 different disciplines and 11 different panels the thematic analysis did not identify clear patterns related to disciplines or panels. The analysis therefore does not include a focus on distinct internal disciplinary evaluative and epistemic cultures (Lamont 2009). Instead, the interviews were primarily analysed in order to identify patterns of similarities and differences across the material, without taking into account the different disciplines of the panel members or which panels they belonged to.

3.6 Ethics

Ethical considerations guided the project throughout the fieldwork and beyond, in order to ensure the protection of the participants' rights and integrity. Informed consent was obtained from all participants before the interviews. All participants agreed to participate voluntarily and were informed, both in written form and verbally, about the aim of the research, the procedures involved and any potential risks. Furthermore, the participants were informed that they could opt out of any questions during the interview (cf Nii Laryeafio and Ogbewe 2023).

The RFO had signed a memorandum of understanding to participate in the study on peer review systems and grant allocation. The RFO is based in a different country from that of the authors, and there were no preexisting contacts between them and the participants prior to the study. The RFO was not informed about which panel members took part in the study, and participants were made aware that the project was conducted independently of the RFO (cf Hammersley and Traianou 2012).

To protect the identity of the interviewees and to ensure secure data handling some information about the participating RFO has been excluded and details about the individual participants which could identify them has been omitted. In the presentation of the findings below, quotes from the interviews are included to illustrate prominent codes and themes, some of which have been edited to remove information which could reveal the identity of the informants or the RFO.

4. Findings and analysis

In exploring the RFO's current review practices, the transition from National Review Panels (NRPs), composed primarily of domestic researchers, to International Review Panels (IRPs) with predominantly international reviewers emerged as a notable development. The staff described this as a strategic shift and policy change initiated by the RFO a couple of years previous to the study: 'In the last few years, the policy towards the panels has changed. Five years ago, they were all domestic' (S2). Another staff member continued to explain what the policy change had implied: 'Now the policy has been to replace domestic researchers with foreign experts' (S3). This staff member also emphasized the effort that had been made to pursue this change: 'For the several last years there was a consistent and successful policy of making the experts almost entirely international' (S3). A third staff member outlined the specifics of their IRPs: 'We have more than 90% and sometimes even 100% [international panellists]. Most of the time there's just one [domestic] panellist' (S5).

The following analysis expands on the findings related to this transition to IRPs. It is structured around two key subthemes identified in the thematic analysis: the benefits and challenges of IRPs. The benefits include reducing conflict of interest, mitigating bias and increasing cognitive diversity. The challenges involve reduced inter-reviewer reliability and diverse evaluative standards, national contextualization and domestic sovereignty, negotiating authority and maintaining collegiality.

4.1 Benefits with international review panels

The members of the RFO staff and the panel members associated the shift to IRPs with several benefits, primarily addressing external factors seen as corrupting the fairness and robustness of the peer review process (cf Lamont 2009; Langfeldt and Kyvik 2011; Shaw 2024)

4.1.1 Reducing conflict of interest

When asked about how to deal with conflict of interest in the peer review process, one of the RFO staff members referred to the benefits with IRPs: 'We also have foreign experts in the panel, so we have no problem with that [conflict of interest]' (S5). Another staff member explained how replacing NRPs with IRPs was driven by a desire to engage panellists who 'would not be involved [in the local academic setting]' (S3), ie not have any conflict of interest. The staff member continued to explain how this had been a problem in their previous NRPs: 'Sometimes, disciplines or fields are small so there are [risks] … I [as a panellist] could be negotiating or could do things in favour or against [an application]' (S3). Staff members thus associated IRPs with the implementation of consistent and universalistic evaluation standards, which they saw as less susceptible to being compromised by conflict of interest (cf Langfeldt and Kyvik 2011).

Both domestic and international panel members also recognized that the move from NRPs to IRPs reduced conflict of interest. One of the international panellists suggested that the limited size of the domestic higher education sector was problematic with regards to conflict of interest and motivated the shift to IRPs: 'If it would be only domestic [panel members], you could have huge conflict of interest because the universities and the institutions there are really few' (LS2-P1). Another international panellist outlined a similar line of argument: 'I think it will reduce bias if people will be from abroad. They will not know the candidates' (PE1-P3). Another international panellist also expressed support of IRPs with reference to conflict of interest, emphasizing how the increased distance from the domestic field of science reduced the risk for clientelism: 'I think it will be less biased if everybody will be from abroad, or some domestic researchers who are abroad' (PE1-P1). A domestic panel chair had previous experience of conflict of interest in the NRPs and had noted the difference in the IRPs: 'Conflict of interest is much less common than it used to be' (HS4-PC).

The move from NRPs to IRPs was also consistent with recent changes to the RFO's formal policy on conflict of interest, which had broadened the definition to include entire HEIs rather than just departments as explained by a staff member: 'Conflict of interest has been redefined. Originally you couldn't be involved in evaluating proposals from your department or faculty, but now it has been extended to the whole institution' (S3). This change made it more challenging to rely on domestic reviewers, who would have to declare conflict of interest for an increasing number of applications. For IRPs conflict of interest was significantly reduced, as explained by a panel chair: 'Also, if you are from the same institute, you need to leave the room when the proposal will be evaluated. But it doesn't happen [often] because all the experts are from abroad' (HS2-PC). This meant that the panels could stay intact, with all panel members participating in the discussion and evaluation for a larger number of applications, which would strengthen the peer review process and make it more equal and robust.

A staff member also suggested that the problem with conflict of interest was not easily solved with the strategy to let the panellists with conflict of interests leave the room during the deliberations:

> Even if you were excluded from evaluating the proposal from your institution, it doesn't mean that outside the door you cannot make a deal with another panel member, who you know, so we really wanted to avoid that. Because this was happening. (S3)

Merely excluding 'involved' (S3) panel members from participating in formal discussions of certain applications would not be sufficient to uphold universalistic standards if informal alliances and horse-trading continued to occur due to preexisting academic networks and relationships (cf Gould et al. 2025; Philipps 2024). These informal deals were further described by one of the international panellists: 'I know you; you know me, I know your friends and you'll give me a favour, I return you the favour later' (PE1-P2). The influence of these alliances on the peer review process was however mitigated by the increased inclusion of international panel members.

4.1.2 Mitigating bias

In addition, IRPs were seen as effective in mitigating several different types of bias understood to interfere with the peer review process (cf Langfeldt and Kyvik 2011). According to one international panellist, 'the foreigner's opinion' (PE1-P4) in the IRPs was particularly welcomed as it was understood to be fair and objective: 'and preventing bias' (PE1-P4). The benefits with IRPs were especially outlined in relation to institutional bias. Reports and statistics from the RFO illustrate that funds consistently over time had been allocated to one of the country's most elite HEI. Institutional bias means that 'being affiliated with a prestigious university can keep a proposal above the bar' (Lamont 2009: 227). Some of the informants also shared examples of what they interpreted as institutional bias manifested in the Matthew effect meaning that 'being affiliated with a prestigious university can keep a proposal above the bar' (Lamont 2009: 227). One international panellists described being involved in deliberations in a previous panel with lower international representation: 'In the previous panel that I was on, you did get people saying: 'Well, obviously, this is really good because it comes out of this particular school, and we know that this works really well'' (HS3-P2). In contrast, in the current IRP the Matthew effect was noticed and questioned: 'There have been lots of comments like: 'The University of [name removed] seems to just get all the funding'' (HS3-P2).

By limiting the influence of preexisting local academic networks and favouritism tied to institutional prestige, IRPs were seen as mitigating the influence of institutional bias. International panellists also emphasized their impartiality and their objective approach to the domestic, prestigious, HEI: 'The institute isn't important to me. I don't think: it's from [domestic elite HEI], for example, so automatically it should be good. [...] The place doesn't matter' (LS1-P3). Another international panellist demonstrated both concerns about institutional bias, and a personal distancing from such bias, declaring objectivity in this aspect but also ignorance with regards to the elite status of any local HEIs:

> I've heard a suggestion that some of the more traditional universities have a bias. [...] I don't know if it's true or not. That they get more funding than the others because there's favouritism amongst the reviewers, like an unconscious bias. I certainly wouldn't have a bias in that sense. [...] Certainly, I don't know anything in the [domestic] system that would give any bias in that sense. (PE2-P2)

IRPs were also associated with increasing the objectivity of the peer review process by addressing attribution bias, particularly the problem with undervaluing women's achievements (cf Győrffy, Herman and Szabó 2020). Similar to the institutional bias the increased distance between reviewers and applicants was emphasized as a positive characteristic of IRPs. One of the international panel chairs explained how the IRPs could mitigate bias: 'I think that most of the international people don't even know whether the researcher is female or male, just according to the name' (PE2-PC). This was also supported by an international panellist: 'For myself, a lot of European names, I couldn't tell [if the applicant was a woman or man] even when knowing the name' (LS1-P2). One of the domestic panellist also agreed with this: 'People who are not locals, they can't recognise who's male or who's female' (LS1-P1). Another international panellist, from a different part of Europe, expressed similar ignorance: 'Most of the time I can't understand if they are male or female by their name because the local names of these countries are so difficult to understand if they are male or female' (LS2-P1). This suggests that international reviewers are less likely to recognize domestic first names and therefore may not automatically infer an applicant's gender. This alters the typical pathway through which attribution bias operates, potentially reducing its impact on the peer review process (cf Shaw 2024). While framed as a path to objectivity, the IRPs were also seen as contributing important perspectives.

4.1.3 Increasing cognitive diversity

The analysis of the benefits of IRPs highlights not only their preventive qualities in that they can mitigate conflict of interest and bias, but also that they can promote objectivity in the peer review process by introducing diverse perspectives that challenge shared assumptions, foster intellectual pluralism, and broaden interpretations of research excellence (cf Langfeldt and Kyvik 2011; Győrffy, Herman and Szabó 2020; Shaw 2024). Panel members highlighted these promotive factors of IRPs and how they encouraged evaluative breadth, particularly when assessing the global relevance, impact and competitiveness of proposals. A domestic panel chair emphasized the value of global benchmarking in this sense: 'To compete with international researchers, we need international experts who will provide international feedback on our projects' (LS1-PC). The IRPs were thereby expected to stimulate the international competitiveness of science and help national research meet global standards of excellence. The chair continued to explain this competitiveness:

> The project could be very well assessed by the people living in [the country of the RFO], but it could be assessed very poorly by the ERC [the European Research Council], and we need that feedback: "Hey, this is not a very competitive grant. Let's just change some ideas and some concepts to make it more important not for domestic research, but internationally." It's absolutely a good idea to have international panellists. (LS1-PC)

An international panellist similarly reflected on the benefits of heterogenous panels: 'I think more diversity is better' (PE2-P2). Another international panellist suggested: 'I would

say the more diversity, the better' (HS5-P1) in order to avoid review panels becoming 'like a small in-group' (HS5-P1) and too local and narrowminded in its assessments. Another international panellist suggested the cognitive precedence of IRPs: 'International people know what's important from the international perspective' (PE2-P1). This line of argument was grounded in an increased internationalization in higher education: 'Science now has to be sort of a global issue. You just have to view the problem in science with a global eye, not just with a national eye' (LS2-P1). One of the domestic panel chairs elaborated on these benefits with IRPs even further and suggested that IRPs need to include panellists from outside of Europe in order to achieve a 'balance' with regards to a 'wider spectrum of opinion':

> Because then you really have a very diversified number of participants from different cultures and universities and so on, then probably the final outcome would be more objective, I would say, because then it does not really reflect the attitude of, for example, Europe, the EU, but it's much more inclusive. (HS4-PC)

One of the international panellists seemed to confirm this difference in interpreting and understanding the evaluation criteria concerning impact and relevance. The panellist explained being 'baffled' when reviewing some of the project proposals with regards to international impact, and described a proposal with a narrow, localized focus: 'It focused very much on a small regional way in [the country of the RFO], and I couldn't actually see the international impact' (HS1-P2). A domestic panellist had noted similar differences between international and domestic panel members: 'They [international panellists] do look more at the impact as certainly they don't like projects which are regional for [homeland of RFO]' (PE2-P3). This highlights how cognitive diversity can expose blind spots in proposals that may resonate locally but lack broader significance (cf Roumbanis 2017).

The perceived benefits of IRPs thus included their increased cognitive diversity which could challenge entrenched norms, foster more inclusive definitions of excellence and broaden impact leading to richer discussions and more nuanced judgments (cf Langfeldt, Reymert and Svartefoss 2024). A domestic chair also described how international panel members would: 'give some good examples of how they evaluate projects in their country' (LS2-PC). This may lead to: 'a discussion, or maybe even we talk about how it is in their country and maybe if we want to change something or they have some suggestions, maybe we should do this or that' (LS2-PC). In this way, IRPs could also introduce alternative perspectives on the structure of peer review processes, offering comparative insights that could inspire reflection and potential refinement of existing practices.

This cognitive diversity, however, also increased complexity in the peer review process, contributing to challenges. The findings regarding these challenges are elaborated further below.

4.2 Challenges with international review panels

The interviews highlighted several challenges associated with IRPs, primarily related to how the inclusion of panellists with diverse national backgrounds complicated the deliberations, reduced inter-reviewer reliability, and required careful calibrations (cf Lamont 2009; Derrick and Samuel 2017; Hug and Ochsner 2022; Oxley and Gulbrandsen 2025).

4.2.1 Reduced inter-reviewer reliability

Albeit the cognitive diversity of IRPs was appreciated in relation to the evaluative criterion of impact, it was also recognized that the panellists' embeddedness in diverse national academic traditions influenced the evaluative standards they applied which could lead to challenges with reaching consensus and low inter-reviewer reliability (cf Derrick and Samuel 2017; Seeber et al. 2021; Oxley and Gulbrandsen 2025). A domestic chair reflected: 'I think there could be bigger differences between the evaluation of different projects when the experts are from all over the world. Mostly from Europe, but from different countries' (LS3-PC). Another domestic panel chair described how panellists' different expectations on the specific domestic context could influence their evaluative standards: 'Some of them judge us as people in [a special part of] Europe and some of them judge us as researchers that should apply for the ERC' (LS1-PC). An international panellist suggested that: 'there is a kind of academic culture that you contribute to the panel' (HS3-P1), further reflecting on national differences: 'Even though we are in Europe, and we all share very similar standards, I would say there are still national differences: in practice, in scope, universities might be organized in different ways … ' (HS3-P1). Speaking more in detail about how these national differences influenced peer reviewing, the panellist noticed how different types of publications were given precedence: 'Certain countries give a lot of importance to journal articles, others to books for example' (HS3-P1). Another international panellist similarly mentioned how publication traditions differed between national contexts: 'No one here [country of the RFO] writes books' (HS5-P1). The panellist continued to highlight a misalignment between the evaluative criteria for publications applied during the peer review process and the norms specifically of the domestic research system:

> The international output is lower than what would be expected in [other national] research system. For the applications I reviewed […] their publication output was very low. Even an assistant professor applying for these grants, had three or four papers out there and not really in high-prestige journals. (HS5-P1)

A domestic panel chair highlighted these discrepancies in evaluative standards for publications, noting that while some panellists would assess an applicant's output as 'really high,' others expected: 'there has to be like 30 publications' (LS3-PC). One of the international panellists also noted disagreements within the IRPs around the value placed on internationalization and mobility: 'Some other countries give a lot of importance to whether the candidate has spent time abroad. In some countries that's an issue. We don't have that' (HS3-P1).

The diverse evaluative standards extended to broader definitions of excellence. An international panellist challenged the notion that excellence required being 'a Nobel Prize laureate' (LS1-P4), criticizing some domestic colleagues for setting unrealistic standards: 'People, particularly domestic colleagues, were really setting the bar high. I just thought it's not realistic because nobody is at that level. You don't need to be at that level to be considered excellent' (LS1-P4). A domestic

panellist similarly appreciated that IRPs had 'improved' evaluative practices by shifting away from a rigid 'black-or-white approach' (PE2-P3): 'For people from abroad, it doesn't matter if the paper is published in an international journal or in a domestic local journal' (PE2-P3). This shift meant that deliberations to a lesser extent focused on: 'talking about the quality of the journal' and instead: 'more about the quality of the individual papers' (PE2-P3).

Different traditions in how excellence and track records were evaluated across national borders could thus lead to low inter-reviewer reliability and increased need for calibration and panel negotiations and even shifts in how evaluation criteria were interpreted and applied (cf Pier et al. 2017). Challenges with negotiations in IRPs were however also identified, as explained by an international panel chair: 'In each country the system is different in how they evaluate research. This is where the problems arise. […] This could be the problem of multinational panels because the system is always different' (PE2-PC). According to a domestic chair, the panels' multinational character meant that: 'It's hard to find the matching criteria to evaluate scientific achievements' (LS2-PC). Similarly, an international panellist described: 'So, it's a different culture that doesn't really translate and it's really hard to judge' (HS2-P2). This meant that diverse evaluative standards sometimes created challenges in establishing common ground through calibration exercises. Previous research has highlighted how these types of challenges result in low inter-reviewer reliability and high intra-panel variability (Pier et al. 2017; Recio-Saucedo et al. 2022). A domestic panel chair emphasized the importance of aligning these diverse evaluative standards: 'We need to calibrate this [the evaluation] on the same level' (LS1-PC). While calibration in peer review panels usually involves the development of certain customary rules of deliberation (Huutoniemi 2012; Gould et al. 2025), this panel chair proposed that it rather entailed informing the international panellist: 'how to evaluate', implying a domestic sovereignty with regards to the criteria used (LS1-PC).

4.2.2 National contextualization and domestic sovereignty

Diverse research systems and evaluative standards, which can lead to low inter-reviewer reliability, particularly required calibration, prompting the development of informal customary rules during deliberations. Lamont (2009: 106) defines cognitive contextualization as an important customary rule requiring panellists to use: 'the criteria of evaluation most appropriate to the field or discipline of the proposal under review' (cf Mallard, Lamont and Guetzkow 2009). The related rule about disciplinary sovereignty (Lamont 2009) refers to the deference given to panellists whose expertise closely aligns with the topics of a proposal, granting them greater authority in its evaluation (cf Huutoniemi 2012). However, in the IRPs, two other key guiding principles emerged: national contextualization and deferring to domestic sovereignty. An international panellist referred to the domestic sovereignty when describing their own 'cultural bias' (HS5-P1) in the review process: 'For me, it's a cultural bias that I don't understand their research culture as an outsider' (HS5-P1). Another international panellist emphasized the importance of national contextualization when evaluating across national borders and explained that it was necessary to take the specific 'academic culture' (HS3-P1) into account, when: 'you are judging someone from another country with a different system, and you have to understand that in the context' (HS3-P1).

These customary rules also underscored the need: 'to keep domestic panellists [in IRPs] because, as an outsider, you have no idea how the domestic research system works. […] it is important to have someone who's an insider of the culture and can give you some context' (HS1-P2).

The principle of domestic sovereignty could even render it superfluous for domestic panellists to read the RFO guidelines, according to a domestic panellist: 'I don't really need to read such guidelines. It's quite obvious for me what should be mentioned in the review, what is important, what is not important' (HS2-P1). In contrast: 'providing those guidelines on how to review, what to mention, what is and what is not important was much more essential [for the international panellists]' (HS2-P1). Conversely, a domestic panel chair emphasized the sovereignty of domestic panel chairs in relation to international panellists:

> It's not obligatory, but it's better to have him [ie the panel chair] from [country of RFO] because he understands how science works in this country. Then we can easily calibrate the [international] panellists and their judgments because we know how the system works here. It's good to explain to them how it works. (LS1-PC)

This quote suggests the importance of calibration exercises which particularly takes into account the diversity in evaluative practices and national and cultural differences and situates the peer review in the particular national setting (cf Mallard, Lamont and Guetzkow 2009). Huutoniemi (2012) similarly reports on how the British reviewers in her study suggested a 'British bias' and hesitated in 'imposing criteria that we would use in our own national context on this situation' (Huutoniemi 2012: 906).

To have research officers, who are firmly embedded in the domestic context, was highlighted as crucial in bridging the gap between 'outsiders' and 'insiders'. Their familiarity with local academic norms allowed them to guide panellists in understanding context-specific expectations during the panel meetings. An international panellist emphasized: 'That's why it's good to have the research officer in the panel that actually knows the academic [domestic] context well and has the ability to explain what is expected from a CV at each career stage' (LS2-P2). A domestic chair similarly described the need to answer questions from international panellists who: 'say they don't know how it works in [country of RFO]' (LS2-PC).

The panel chair was generally seen as pivotal for this national contextualization (cf Gould et al. 2025). Accordingly, panellists noted a diminished role when domestic chairs were replaced by international scholars who struggled to assert authority without domestic sovereignty in the IRPs: 'In the domestic panels, usually, the chair was important' (PE2-P3). Domestic chairs were seen as more engaged, actively leading discussions, whereas international chairs: 'really just left everything for the research officer' (PE2-P3), relying on the research officer's domestic expertise. The domestic panellist's rhetorical question: 'What's the point of the chair?' (PE2-P3), suggested the weakened, more passive, position of the international chairs.

4.2.3 Negotiating authority and maintaining collegiality

Collegiality and what Lamont (2009: 240) describes as 'the reciprocal recognition of authority' have been identified as central to the development of customary rules of deliberations in peer review process (cf Huutoniemi 2012; Hug and Ochsner 2022). Within the IRPs, however, these dynamics were described as particularly challenging, making the negotiation of authority and collegiality more complex.

A domestic chair reflected on the diversity of IRPs: 'From a practical point of view, as a chair, when you have a more homogenous panel, it's much easier to manage' (HS4-PC). This could be interpreted as more effort was needed to develop, establish and maintain customary rules of deliberations in more heterogenous panels, such as IRPs (cf Pier et al. 2017, 2018; Seeber et al. 2021). The findings also indicate that the informal rules of national contextualization and domestic sovereignty established in IRPs sometimes clashed with the rules of disciplinary sovereignty and deferring to expertise, which in turn put strain on the customary rule about collegiality and respectful interaction (cf Lamont 2009).

An international panellist described domestic panellists displaying 'a real sense of ownership' during deliberations, and 'a real sense that they are doing something to protect quality of science in their country' (HS3-P2). The panellist interpreted this to be a manifestation of: 'this kind of national pride which we haven't really experienced in [my country] as much' (HS3-P2). Another international panellist similarly shared experiences of lack of trust: 'I did feel that there was a degree of defensiveness from some domestic colleagues on the panels' (HS1-P1). Nevertheless, this panellist provided some nuances by giving voice to the complexity of the panel dynamics: 'I've heard people say there's a feeling that the international panellists don't take the domestic panellists seriously or they feel they're being patronised' (HS1-P1). A third international panellist expressed appreciation of the culture of deliberation in IRPs, emphasizing the authority of the international panellists:

> What I like very much there is that, when you have a domestic professor that has a different opinion than foreign professors, generally, I would say in 100% of cases, the foreign professor is winning and the domestic professor says, "Okay. You are my guest. You're invited. I'm trusting you. […] Generally, if you are not reaching a consensus, the domestic professors generally decline their opinion. (PE1-P4)

This could be interpreted as the opposite to 'a sense of ownership', but also as overreliance in the authority of international panellists. The negotiation of authority also extended to interactions with external ad-hoc experts. A domestic panellist described the 'important difference' observed between NRPs and IRPs, while outlining that the domestic panels: 'accepted the external experts are smarter than we are, because they are so into the subject' (HS2-P1). The panellist continued to explain: 'People from international panels just don't agree with them [the external experts] and can talk a lot about why they are wrong. […] They feel like they can defend against those who cannot respond because they just gave their written opinion' (HS2-P1). This could imply that there is more room for rhetorical persuasion and greater contestation in IRPs, which means panellists who are more skilled at argumentation may exert disproportionate influence. These dynamics illustrate the complexity of building trust and authority within IRPs, where national identities and academic traditions intersect with informal rules of deliberation (cf Langfeldt, Reymert and Svartefoss 2024).

5. Discussion

5.1 Summary

This article has highlighted some benefits of utilizing IRPs and some challenges of the collective review process taking place in IRPs. The findings suggest that the international composition of IRPs supported the formal integrity of the peer review process by reducing bias and conflicts of interest. In this RFO, reducing the number of domestic reviewers had significantly decreased the risk for conflict of interest and bias in the evaluation process, as described by the RFO's staff members and panel members alike. The diverse national and institutional backgrounds of reviewers in IRPs reduced the influence of several problematic factors understood as disturbing the peer review process: preexisting academic networks and bias towards domestic elite HEIs. The IRPs also entailed a type of language barrier which reduced the likelihood of gender-based, attribution bias. In addition, IRPs introduced external perspectives, academic traditions and diverse cultural perspectives, not rooted in the local research context. These perspectives broaden the evaluative lens and the definitions of scholarly excellence, challenging entrenched assumptions, fostering cognitive diversity, identifying blind spots, and enriching the quality of deliberation. IRPs, therefore, serve not only to reduce bias by diluting local networks and mitigate potential conflicts of interest but also as a strategic tool to promote richer, more globally attuned evaluations, achieving greater fairness, and inclusivity.

The IRPs were however also associated with challenges due to these diverse perspectives and the varying levels of familiarity with the national research system. These challenges were primarily associated with weakened inter-reviewer reliability, increased demands for calibration and negotiations of informal customary rules in IRPs (cf Huutoniemi 2012). The inclusion of international peer reviewers challenged the informal dynamics of the panel meetings and the context-sensitive deliberations. It proved more demanding to develop shared customary rules to guide deliberations and decision-making in the IRPs due to several interrelated internal factors: diverging understandings of excellence and track records, and the inherent fragility of trust and authority in diverse panels. The findings thus highlight how the increased heterogeneity brought about by IRPs can add complexity to the already intricate panel deliberations and increase strain on the panel chairs.

5.2 Contributions to previous research

This study contributes to the growing body of literature on peer review by exploring how IRPs function in practice, thereby filling an existing gap and responding to calls for more research (Győrffy, Herman and Szabó 2020; Gould et al. 2025). The findings underscore the micro-political dynamics inherent in IRPs, revealing how evaluative standards, authority, and expertise, are negotiated within these diverse review panels. Adding to previous studies on review processes in RFOs, which have primarily focused on NRPs, this article introduces the concepts national contextualization and domestic sovereignty while adding new dimensions to the

existing understanding of inter-reviewer reliability (cf e.g. Feliciani et al. 2022; Sattler et al. 2015). This contribution expands our understanding of the specific challenges in IRPs related to developing informal customary rules for effective deliberation. The findings thus nuance the expectations on increased review quality in cross-border peer review (cf Langfeldt and Kyvik 2011). The study further highlights how the diverse perspectives and evaluative standards in IRPs add new dimensions to the calibration exercises and the development of the informal customary rules of deliberation (cf Lamont 2009; Mallard, Lamont and Guetzkow 2009). National contextualization and domestic sovereignty emerged as key rules in these deliberations, but also potentially conflicting with principles like deference to expertise and collegiality. Overall, the analysis of IRPs deepens our understanding of peer review deliberations, adding new dimensions and nuances to an already complex interactional context characterized by low inter-reviewer reliability, disagreements, uncertainty and negotiations (cf Seeber et al. 2021; Roumbanis 2022).

5.3 Practical recommendations for RFOs

As the trend towards internationalization in research evaluation continues, it is likely that more RFOs will make the shift from NRPs to IRPs. This shift aligns with the still growing recognition of the benefits of cross-border collaboration and a more global perspective in assessing the quality and impact of research. For smaller countries especially, IRPs also offer a practical solution to issues like reviewer fatigue and reviewer burden (cf Lamont 2009; Bendiscioli 2019; Recio-Saucedo et al. 2022). This can be particularly important in niche fields with limited domestic expertise. The shift to IRPs also aligns with the growing trend of conducting meetings online. The increasing reliance on digital meetings supports the convening of IRPs with academics from diverse geographic locations, which could otherwise require considerable resources from the RFO (cf Peterson and Husu 2023).

For RFOs aiming to continue to improve their review systems, a nuanced understanding of both the advantages with enhanced objectivity and the operational complexities and challenges of IRPs is crucial. RFOs could consider several different strategies to optimize their use.

This study highlights how the international character of the review panels reduced inter-reviewer reliability and increased the need for calibration with regards to the criteria used in the evaluation of the applications and for assessing the merits of the applicants (cf Pier et al. 2017; Feliciani et al. 2022). RFOs should consider different strategies to institutionalize and formalize calibration practices, to ensure that panel chairs and research officers facilitate these discussions to promote consistency throughout the review process (cf Lamont 2009; Gould et al. 2025).

Echoing Lamont's (2009) call for continual self-evaluation and education, RFOs may want to implement training programs for members of their panels that address both formal review criteria and the informal dynamics of panel deliberation (cf Sattler et al. 2015; Derrick and Samuel 2017; Feliciani et al. 2022; Recio-Saucedo et al. 2022). Providing detailed review guidelines is particularly crucial for international reviewers unfamiliar with local research systems. The guidelines could then be expanded to include a clarification of expectations regarding applicant CVs, publication norms, and career trajectories within the specific national context.

It should also be noted that managing diverse panels places additional strain on the important role of panel chair (cf Gould et al. 2025). RFOs may therefore offer specific training and increased resources to support chairs navigate the complexities of IRPs, including conflict resolution and fostering inclusive deliberation. Similarly, the presence of research officers in panel meetings was found indispensable in contextualizing local norms for international reviewers. RFOs are therefore encouraged to ensure that research officers are empowered to actively support panels by clarifying local standards and practices.

While the benefits of IRPs are substantial, their challenges are equally significant but manageable. With thoughtful implementation, including structured calibration, comprehensive training, and robust support systems, RFOs can harness the strengths of IRPs while mitigating their complexities.

5.4 Strengths and limitations

This article offers several strengths that contribute to the growing body of research on peer review in higher education contexts. First, it is among the first articles to focus exclusively on IRPs, offering a concentrated exploration of their unique dynamics. While previous research has examined certain limited international elements within peer review, this study provides a focused lens on IRPs as a distinct evaluative setting, helping to fill a gap in the literature. Second, the study is grounded in rich qualitative data derived from in-depth interviews with both domestic and international panel members and RFO staff members. This approach allows for a nuanced understanding of the specific challenges and benefits associated with IRPs, insights often overlooked in more quantitative assessments. Third, the study sheds light on several key factors influencing the dynamics of the peer review process, which play a critical role in shaping panel outcomes, an area that is increasingly receiving attention in peer review research. Fourth, the study makes a theoretical contribution by expanding on our understanding of how diversity within panels can simultaneously enrichen and complicate evaluative practices. Finally, the findings hold practical relevance for RFOs, offering actionable insights into how IRPs function and the considerations necessary for improving fairness, cohesion, and efficiency in such panels.

Despite these strengths, the study has certain limitations that warrant consideration. First, there are limitations in generalizability. The study focuses on a single RFO at a particular point in time, and while the findings offer valuable insights, they may not fully capture the dynamics present in other funding agencies with different institutional or cultural contexts. Second, the theme of IRPs emerged during the interviews and become a focus during the analysis process. It was thus not the primary aim of the research study. If it had been, more specific questions on this topic would have inquired more deeply into this, rather than only follow up questions. On the other hand, this also meant that the topic could be explored in a more open manner with the results less affected and limited by interviewer pre-understandings, or by how the questions were formulated or asked. However, with a more explicit focus on IRPs the results may have been more nuanced, broad and rich. Third, participant self-selection could have led to overrepresentation of certain perspectives while silencing others. Fourth, it was not possible to take disciplinary differences into account in the analysis, although they most likely would have proved relevant. Fifth, the study relies

on retrospective accounts, which introduces the potential for memory bias or selective reporting. Participants' reflections on past panel meetings may not fully capture the complexities or subtleties of real-time deliberations, particularly in contentious situations. Finally, the absence of direct observation of panel deliberations limits the study's ability to analyse peer review panel dynamics, and spontaneous interactions that are often crucial in shaping consensus-building and decision-making processes (cf Philipps 2024). Future research could address this by incorporating direct observations of IRP meetings, similar to Lamont's (2009) and Roumbanis (2017) ethnographic approach.

Overall, while these limitations suggest areas for further research, they do not detract from the study's contributions. Instead, they highlight the complexities inherent in studying peer review processes and underscore the value of continued inquiry into the evolving dynamics of IRPs.

Acknowledgements

The article builds on a rich qualitative data set, produced within the Horizon 2020 project GRANteD (Grant Allocation Disparities from a Gender Perspective) (2019–23). The project analysed the occurrence and causes of gender bias in research funding in Europe. The GRANteD project included five empirical, multi-level and multi-method, case studies in which gender bias and gender equality policies were studied in-depth, drawing on both qualitative and quantitative approaches to data collection. The case study research design allowed the inclusion of five public RFOs in the European Research Area. This article draws on the findings from one of these case studies.

Funding

This work was supported by the European Commission and received funding from the European Union's Horizon 2020 research and innovation programme [grant agreement No 824574].

Note

1. Four of the 27 interviews with panel members were performed by a third member of the project.

References

Acker, S., Ylijoki, O. H., and McGinn, M. K., eds. (2024). *The Social Production of Research: Perspectives on Funding and Gender*. Abingdon, Oxon, UK: Taylor & Francis.

Bendiscioli, S. (2019) 'The Trouble with Peer Review for Allocating Research Funding. Funders Need to Experiment With Versions of Peer Review and Decision-Making', *Science Policy*, 20: 1–5.

Derrick, G., and Samuel, G. (2017) 'The Future of Societal Impact Assessment Using Peer Review: Pre-Evaluation Training, Consensus Building and Inter-Reviewer Reliability', *Palgrave Communications*, 3: 1–10.

De Wit, H., and Hunter, F. (2015) 'The Future of Internationalization of Higher Education in Europe', *International Higher Education*, 83: 2–3.

Feller, I. (2013). 'Peer Review and Expert Panels as Techniques for Evaluating the Quality of Academic Research', in A. N. Link and N. S. Vonortas (eds) *Handbook on the Theory and Practice of Program Evaluation*, pp. 115–142. Cheltenham, UK: Edward Elgar Publishing.

Feliciani, T. et al. (2022) 'Designing Grant-Review Panels for Better Funding Decisions: Lessons from an Empirically Calibrated Simulation Model', *Research Policy*, 51: 104467.

Fereday, J., and Muir-Cochrane, E. (2006) 'Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development', *International Journal of Qualitative Methods*, 5: 80–92.

Gallo, S. A., Sullivan, J. H., and Glisson, S. R. (2016) 'The Influence of Peer Reviewer Expertise on the Evaluation of Research Funding Applications', *PloS One*, 11: e0165147.

Gould, J. S. et al. (2025) 'Threats to Grant Peer Review: A Qualitative Study', *BMJ Open*, 15: e091666.

Győrffy, B., Herman, P., and Szabó, I. (2020) 'Research Funding: Past Performance is a Stronger Predictor of Future Scientific Output Than Reviewer Scores', *Journal of Informetrics*, 14: 1–13.

Hammersley, M., and Traianou, A. (2012). *Ethics in Qualitative Research: Controversies and Contexts*. London, UK: Sage Publications.

Hirschauer, S. (2010) 'Editorial Judgments: A Praxeology of "Voting" in Peer Review', *Social Studies of Science*, 40: 71–103.

Hren, D. et al. (2022) 'What Makes or Breaks Competitive Research Proposals? A Mixed-Methods Analysis of Research Grant Evaluation Reports', *Journal of Informetrics*, 16: 101289.

Hug, S. E., and Ochsner, M. (2022) 'Do Peers Share the Same Criteria for Assessing Grant Applications?', *Research Evaluation*, 31: 104–17.

Huutoniemi, K. (2012) 'Communicating and Compromising on Disciplinary Expertise in the Peer Review of Research Proposals', *Social Studies of Science*, 42: 897–921.

Husu, L., and Peterson, H. (2022). *Synthesis report on contextual factors, gender equality policy analysis and gender bias risk analysis* (Deliverable 5.1., GRANteD—Grant Allocation Disparities from a Gender Perspective).

Kaltenbrunner, W., and de Rijcke, S. (2019) 'Filling in the Gaps: The Interpretation of Curricula Vitae in Peer Review', *Social Studies of Science*, 49: 863–83.

Lamont, M. (2009). *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Langfeldt, L., Reymert, I., and Svartefoss, S. M. (2024) 'Distrust in Grant Peer Review—Reasons and Remedies', *Science and Public Policy*, 51: 28–41.

Langfeldt, L., and Kyvik, S. (2011) 'Researchers as Evaluators: Tasks, Tensions and Politics', *Higher Education*, 62: 199–212.

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: 820–41.

Langfeldt, L. (2004) 'Expert Panels Evaluating Research: Decision-Making and Sources of Bias', *Research Evaluation*, 13: 51–62.

Langfeldt, L. (2006) 'The Policy Challenges of Peer Review: Managing Bias, Conflict of Interests and Interdisciplinary Assessments', *Research Evaluation*, 15: 31–41.

Laudel, G. (2024) 'Where Do Field-Specific Notions of Research Quality Come from? ', *Research Evaluation*, 33: rvae027.

Maldonado-Castellanos, I., and Barrios, L. M. (2023) 'Ethical Issues When Using Digital Platforms to Perform Interviews in Qualitative Health Research', *International Journal of Qualitative Methods*, 22: 1–10. Article: 16094069231165949.

Mallard, G., Lamont, M., and Guetzkow, J. (2009) 'Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review', *Science, Technology, & Human Values*, 34: 573–606.

Meadmore, K. et al. (2020) 'Decision-Making Approaches Used by UK and International Health Funding Organisations for Allocating Research Funds: A Survey of Current Practice', *PLoS One*, 15: e0239757.

Nii Laryeafio, M., and Ogbewe, O. C. (2023) 'Ethical Consideration Dilemma: systematic Review of Ethics in Qualitative Data

Collection through Interviews', *Journal of Ethics in Entrepreneurship and Technology*, 3: 94–110.
Oxley, K., and Gulbrandsen, M. (2025) 'Variability and Negligence: Grant Peer Review Panels Evaluating Impact Ex Ante', *Science and Public Policy*, 52: 254–68.
Peterson, H., and Husu, L. (2023) 'Online Panel Work Through a Gender Lens: Implications of Digital Peer Review Meetings', *Science and Public Policy*, 50: 371–81.
Philipps, A. (2024). 'Dealing With Potentials and Drawbacks of Peer Review Panels: About the Intertwined Layers of Determinacy and Indeterminacy', *Research Evaluation*, 1–8.
Pier, E. L. et al. (2018) 'Low Agreement Among Reviewers Evaluating the Same NIH Grant Applications', *Proceedings of the National Academy of Sciences*, 115: 2952–7.
Pier, E. L. et al. (2017) 'Your Comments Are Meaner Than Your Score': Score Calibration Talk Influences Intra-and Inter-Panel Variability During Scientific Grant Peer Review', *Research Evaluation*, 26: 1–14.
Recio-Saucedo, A. et al. (2022) 'What Works for Peer Review and Decision-Making in Research Funding: A Realist Synthesis', *Research Integrity and Peer Review*, 7: 2.
Roumbanis, L. (2017) 'Academic Judgments Under Uncertainty: A Study of Collective Anchoring Effects in Swedish Research Council Panel Groups', *Social Studies of Science*, 47: 95–116.
Roumbanis, L. (2019) 'Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-Making Mechanisms for Allocation of Research Grants', *Science, Technology & Human Values*, 44: 994–1019.
Roumbanis, L. (2021) 'The Oracles of Science: On Grant Peer Review and Competitive Funding', *Social Science Information*, 60: 356–62.
Roumbanis, L. (2022) 'Disagreement and Agonistic Chance in Peer Review', *Science, Technology, & Human Values*, 47: 1302–33.
Sattler, D. N. et al. (2015) 'Grant Peer Review: Improving Inter-Rater Reliability With Training', *PloS One*, 10: e0130450.
Schmaling, K. B. et al. (2024) 'Perceptions of Grant Peer Reviewers: A Mixed Methods Study', *Research Evaluation*, 33: rvae050.
Seeber, M. et al. (2021) 'Does Reviewing Experience Reduce Disagreement in Proposals Evaluation? Insights From Marie Skłodowska-Curie and COST Actions', *Research Evaluation*, 30: 349–60.
Shaw, J. (2024) 'Fund People, Not Projects': From Narrative CVs to Lotteries in Science Funding Policy', *Research Evaluation*, 33: rvae035.
Stensaker, B. et al. (2010) 'The Impact of the European Standards and Guidelines in Agency Evaluations', *European Journal of Education*, 45: 577–87.
Terry, G. et al. (2017) 'Thematic Analysis', *The SAGE Handbook of Qualitative Research in Psychology*, 2: 17–37.
Tricco, A. C. et al. (2017) 'Strategies to Prevent or Reduce Gender Bias in Peer Review of Research Grants: A Rapid Scoping Review', *PLoS One*, 12: e0169718.
Yin, R. K. (2009). *Case Study Research: Design and Methods*, 4th edn. London, UK: Sage.

Research Evaluation, 2025, 34, 1–12
https://doi.org/10.1093/reseval/rvaf030
Article