D 6.3 Research paper on gendered practices in grant allocation processes:

Reforming peer review practices – lessons learned from the implementation of gender equality policies in Research Funding Organisations

D 6.3

Project no.:	824574
Project acronym:	GRANteD
Project title:	Grant Allocation Disparities from a Gender Perspective
Instrument:	H2020-SwafS-2018-2020/H2020-SwafS-2018-1
Start of the project:	01.01.2019
Duration:	58 months
Work Package:	WP 6
Due date of deliverable:	Month 58, October 2023
Organisation name of lead contractor for this deliverable:	JOANNEUM RESEARCH
Document version:	v03

Name	Organisation	Contribution
Helene Schiffbänker	JOANNEUM RESEARCH	Main author 1
Angelika Sauer	JOANNEUM RESEARCH	Main author 2
Helen Peterson	Örebro University	Main author 3
Peter van den Besselaar	TMC BV	Written comment v01
Michael Meuser	TU Dortmund	SAB review: Written feedback v01
Suzanne de Cheveigné	CNRS	SAB review: Written feedback v02

Dissemination level	
CO	Confidential, restricted under conditions set out in Model Grant Agreement

Consortium:

The GRANteD consortium consists of six partners:

JOANNEUM RESEARCH Forschungsgesellschaft mbH (JR) (project leader, Austria)
ÖREBRO UNIVERSITY (ORU) (Sweden)
AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC) (Spain)
DEUTSCHES ZENTRUM FÜR HOCHSCHUL- UND WISSENSCHAFTSFORSCHUNG GMBH (DZHW) (Germany)
TERESA MOM CONSULTANCY BV (TMC) (Netherlands)
FORSKNINGSPOLITIK SVERIGE AB (FPS) (Sweden)

For questions and comments please contact:

Helene.schiffbaenker@joanneum.at

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Reforming peer review practices – lessons learned from the implementation of gender equality policies in Research Funding Organisations

Helene Schiffbänker, Angelika Sauer (JOANNEUM RESEARCH)

Helen Peterson (Örebro University)

Abstract[1]

This GRANteD deliverable 6.3 is a draft research paper in WP6 to be published later as scientific paper. D6.3 builds on core findings from D6.1 on assessment practices in peer review panels and aims to link them to the discourse of reforming research assessment.

This paper aims to advance our knowledge of research funding organisations (RFOs) as promotors for change by analysing how innovative gender equality policies that RFOs have adopted to mitigate gender bias have been implemented in practice in peer review panels[2].

The paper focuses specifically on three innovative policies which each address one approach to gender equality, namely fixing the numbers, fixing the organisation and fixing the knowledge.

Based on qualitative research data from interviews and observations in the Horizon2020 project GRANteD, we have gained valuable insights into general implementation practices, gender-related risks, support factors and encountered problems when implementing the three policies.

These findings may also be relevant for ongoing activities to reform the research assessment process with the objective of enhancing inclusivity; notable examples include the Declaration on Research Assessment (DORA) and the Coalition for Advancing Research Assessment (CoARA).

We have identified some conclusions about adapting the assessment process, such as having a clear policy concept with appropriate assessment indicators available when implementing a new policy or strengthening the role of chairs, or the need for reformation of the way in which innovative policies are communicated and how capacities are built along the new policy requests.

[1] We thank all reviewers who have contributed to improve the text, colleagues in the GRANteD consortium and SAB members.

[2] We use 'panel' as general term for assessing bodies, while in some RFOs they are called 'board', 'expert team' or 'council'.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

1. Introduction

The role of research funding organisations (RFOs) to strive for change in the research ecosystem has been highlighted for the last two decades (Husu et al. 2004). RFOs play a key role in fostering gender equality in science when they implement policies to increase equality and inclusion and in mitigating gender bias in research funding. There is, however, not much information on how formal RFO policies are implemented in panel practices.

In this paper, we contribute to the understanding of how formal policies can bring about change by discussing the challenges that emerge on panel level when new, innovative policies are implemented. By conducting interviews with panel members as well as panel chairs and observing peer review panels in real time, we were able to study implementation challenges in more detail. In the following, we explore how three selected policies for transformation and reformation of peer review assessment have been implemented in various RFOs.

RFOs have been drivers for transformative initiatives for decades, in particular by implementing policies to mitigate gender bias in the assessment process. A gender-fair assessment process is key for a gender-fair allocation of research grants, which again establishes the basis for equal career chances of female and male researchers. Across countries and across RFOs, considerable differences can be witnessed, depending on national contexts and the specific strategic orientation (Husu and Peterson 2022). Some RFOs have been very active in implementing formal policies to mitigate gender bias, others are just at the beginning. The European Commission supports and encourages national activities for gender equality by providing incentives and by pushing new innovative standards, like introducing gender equality plans as a new eligibility criterion in Horizon Europe in 2021 (EC 2021a).

When implementing policies to mitigate gender bias in research assessment, RFOs address specific bias risks which are determental primarily for women, as pointed out by gender scholars. It is argued that it matters *who* is doing the assessment, so the gender of reviewers can have a gendered impact on the funding outcomes (Arceo-Gomez and Campos-Vazquez 2022). This is manifested in the assessment process as homosocial reproduction when men tend to support and select other men similar to themselves (Husu 2021; Van den Brink and Benschop 2012).

Furthermore, it has been discussed what is assessed, referring in particular to assessment criteria and the concept of excellence. Excellence has been discussed as biased due to the focus on quantitative metrics and it has been questioned whether the indicators applied are actually able to measure what is excellent research – even more so as it has been shown that researchers aim for boosting impact factors by intention (Martin 2016). Recently, efforts to broaden the concept of excellence (see more in section 3.2.1) and to go beyond publications and impact factors as relevant assessment criteria are strengthened, and it is claimed that science communication, societal impact or the quality of teaching and leading research teams should be included when assessing the merit of applicants. While those revisions might be beneficial for researchers in general, they might in particular be beneficial for women who – due to a gender-based work distribution – have more obligations beyond doing research. More focus on female researchers is needed as it was stated that the current research evaluation system fosters inequality in respect to women and minorities (LERU 2019).

Recently, the emphasis on reforming research assessment has been increased. Initiatives, such as the San Francisco Declaration on Research Assessment (DORA), argue against using predominantly journal-based metrics, such as the Journal Impact Factor, for assessing applicants and argue for a more responsible use of bibliometrics in research assessment as stated in the Leiden Manifesto (Hicks et al. 2015). They argue for taking into account broader contributions when discussing excellence, a more

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

qualitative research assessment (Wilsdon 2015) and a stronger focus on individual qualities of researchers, team performance and leadership, as implemented in Dutch research organisations (Global Research Council 2018). RFOs increasingly collaborate in driving for change and bottom-up activities, such as the Coalition for Advancing Research Assessment (CoARA), are gaining support also from RFOs. The latter highlights the importance of including peer reviews in the evaluation of scientific output and assigns a complementary function to quantitative indicators (CoARA 2022).

One tool associated with the demands of DORA and CoARA is the narrative CV. In contrast to conventional CVs, narrative CVs "supplement traditional types of biographical information with narrative elements through which researchers can tell more contextual stories about their background/career/career motivation" (Kaltenbrunner et al. 2023).

So far, we have argued that RFOs had made considerable efforts to design and put in place formal policies to mitigate gender bias and thus become change agents for more gender equality. Now, we go a step further and investigate how such policies are implemented in practice when peer review panels meet to assess applications.

We discuss three innovative policies to increase gender equality, which address three different levels on which gender in science is promoted (Schiebinger and Schraudner 2011): 1) the implementation of a quota for gender-balanced funding outcomes is a policy to *fix the number* of female grantees. 2) approaches to broaden the concept of excellence refer to the culture of RFOs and are policies for *fixing the organisation*. 3) the implementation of gender in research content and innovation (GiRI) is an innovative policy for making knowledge production and innovation more inclusive, so to *fix the knowledge.*

These GE policies, in addition to their primary objective of fostering gender fairness and inclusivity, also strive to enhance the overall quality of the research assessment process. For each of these policies, we discuss the following research questions:

(1) How do reviewers make sense[3] of and understand the policy?
(2) How do they apply the policy in practice in the context of peer review panels?
(3) Which factors support and hinder the successful implementation of each policy in practice?

The used data were collected within the Horizon 2020 project GRANteD. Based on interviews with panel members as well as panel chairs and the observation of peer review panels, we investigated how formal policies were implemented in practice.

Based on these findings, we try to identify some entry points for improving the policies in order to increase their impact and effectiveness. Further, we derive overall learnings for the design and implementation of additional (reforming) policies. They can be informative for any interested and engaged stakeholders of the science system, and also for other RFOs which plan similar activities. These endeavours have to be considered against the background that it was argued that "planned diversity interventions barely lead to sustainable change in the long run as it is difficult to change norms and daily practices of organisational members" (Van den Brink 2020, p. 379–380).

The paper is structured as follows: First we discuss the role of RFOs in the move to more gender equality along the 'three Fixes'. After describing our research approach, we provide the main findings for the selected policies, presenting first the context, then their implementation in practice and finally we discuss these findings. The role of chairs has been shown to be crucial in these implementation

[3] Steinbauer et al. (2015) have introduced sense-making as an important element in the decision making process which can be conscious and unconscious.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

processes, so it is analysed in a fourth part. In the conclusion, we discuss some general foods for thought before we list learning on two levels. We close with some implications for RFOs and for the broader aim of reforming research assessment.

2. RFOs as promotors for gender equality on three strategic levels

The role of research funding organisations (RFOs) in transforming science, research and knowledge is growing both within Europe and beyond. RFOs are key actors in the research and innovation sector, as they design research programmes, define assessment criteria and are responsible for distributing research budgets in the intended manner. When allocating research grants, a fair assessment process is key. To support a gender-fair assessment, RFOs have had to implement various policies to counteract long-existing male norms based on a male-dominated science system, which is still the case in leading positions, such as professorships and in decision-making bodies.

Building on a growing research corpus on gender bias in research assessment, RFOs have constantly adjusted their procedures and developed and implemented innovative measures, activities and policies. When designing new gender equality policies, RFOs make use of gender and feminist research to ensure that "feminist policy matters in practice" (Mazur 2016, p. 15). They are put in place to optimise the assessment process and to prevent gender bias risks.

When striving for gender equality, RFOS can intervene on different strategic levels. According to Schiebinger and Schraudner (2011), gender equality policies and measures can aim to fix the numbers, the institutions or the knowledge.

2.1.1. RFOs fix the numbers

Policies aiming to "fix the numbers" typically have a quantitative target, as they try to increase the share of women and other minority groups in the overall science system and in specific positions and for specific stakeholders. RFOs may strive for an equal representation of women and men in scientific boards and review panels, as having more women in decision-making bodies is crucial for raising gender awareness and for providing role models. This is why the European Commission suggests a 40 % quota (EC 2018). To increase the number of female reviewers, RFOs can define internal quota to ensure that gender is taken into consideration when recruiting remote reviewers. This policy is adopted in slightly different ways in the Austrian Science Fund (FWF), the Swedish Research Council (SRC), the Science Foundation Ireland (SFI) and the Polish National Science Centre (NCN).

Furthermore, policies for increasing the number of women *applicants* are relevant, as even when success rates for female and male candidates are similar, application numbers of female applicants might be low. To increase the number of female applicants, RFOs may present the funding calls in an inclusive and gender-neutral manner, as applied by the Estonian Research Council and the Academy of Finland. Certain RFOs have included additional information that specifically targets women researchers, such as extension regulations in case of care obligations and other leaves (ERC). RFOs also encourage RPOs to become active, as seen in Ireland, where research organisations were incentivised by the SFI to support and encourage excellent female researchers to apply for funding, since more applicants per research organisation could apply if more female researchers applied. Fritch et al. (2019) conclude that this policy has been very successful in increasing the number of female applicants.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Finally, RFOs have policies to increase the share or the number of female *grantees*. Here, different kinds of re-ranking policies and quotas to achieve a predefined number of female grantees can be found (see discussion in chapter 3.1).

An important measure to "fix the numbers" is to closely monitor gender-segregated application behaviour and gender-segregated funding outcomes. This monitoring system enables RFOs to identify future needs for action. Some RFOs enhance transparency by making monitored data publicly accessible on their websites (Husu and Peterson 2022).

2.1.2. RFOs fix the institutions

Gender equality policies to "fix the institutions" target the core processes of an organisation in order to foster structural change in research organisations. They can be categorised into two main types: Firstly, there are policies to adopt internal processes, like hiring, training and rewarding of RFO staff and management. Specific structures have been established to deal with gender issues, like staff units and/or working groups for gender issues. Secondly, the allocation of budgetary resources across various target groups and disciplines and the way funding programmes are structured can be seen as elements to fix the institutions.

However, the main emphasis in RFOs is on the assessment and allocation of research grants, so the most important process to "fix" concern primarily the peer review process and how it is organised. These policies address two distinct external stakeholders groups: researchers who potentially apply for grants, and remote reviewers and panel members who assess the grant applications. This strategic level includes a focus on the definition of assessment criteria. Accordingly, policies on this level attend to the operationalisation and broadening of the concept of excellence, like narrative CVs or a more qualitative assessment of applications. All policies and practices where reviewers negotiate, grade and assess the applications as well as decision making in panels are covered here, acknowledging the risks of bias appearing in practices during panel negotiations.

2.1.3. RFOs fix the knowledge

Policies for "fixing the knowledge" address the process of producing knowledge and innovation, which, albeit intending to generate objective and valid knowledge, are not impartial when it comes to women and other minority groups. The policy to address gender in research and innovation (GiRI) is a main intervention here. It aims to integrate a sex, gender or intersectional analysis into projects where it's relevant in order to produce scientific knowledge that is equally relevant for all genders. Applicants are required to implement a gender research dimension in their research approach, and reviewers are tasked with assessing the extent and quality of this integration. An existing body of literature indicates that it is imperative for RFOs to create an environment that fosters gender inclusivity in research and ensures that external stakeholders have the required tools to conduct gender analysis. For example, it is crucial for external stakeholders that RFOs provide clear definitions of terms that are relevant for gender analysis (Hunt et al. 2022). In the 'Benchmarking and assessment report on guidelines for sex/gender analysis' of the GENDERACTIONplus project, Korsvik et al. (2023) highlighted the current difficulties RFOs are facing in operationalising the GiRI concept.

From each of these strategic intervention levels, we have selected one innovative gender equality policy that is analysed in more detail (see 4.1, 4.2, 4.3), based on the GRANteD case studies in RFOs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

3. Data collection and analysis

This paper draws on five case studies conducted in RFOs across Europe. Case studies can be understood as an "in-depth examination of one or more subjects of study (cases) and associated contextual conditions" (Sovacool et al. 2018, p. 18). Studying practices in panels is one part[4] of these case studies, which provide a rich and detailed understanding of the review processes and how reviewers act there. While the scope of this qualitative analysis is limited and findings are more exploratory than explanatory (Yin 2013), our findings are nevertheless relevant in general, as they inform how (innovative) policies are implemented in panel practice.

The five RFOs cover a range of disciplines, including Humanities and Social Sciences, Medicine and Health as well as Natural Science and Technology. For each RFO, panel practices were analysed for one call of a selected funding programme:

- FWF – Austrian Science Fund, Austria: ESPRIT Career Advancement for Postdocs
- NCN – National Science Centre, Poland: SONATA
- SFI – Science Foundation Ireland, Ireland: Frontiers for Future Programme (FFP)
- SRC – Swedish Research Council, Sweden: International postdoc grant (IPD)
- SRDA – Slovak Research and Development Agency, Slovakia: General Call

The RFOs also provided different levels of gender policies. Three have a long tradition regarding gender equality and two show emerging engagement.

Overall, 104 semi-structured interviews were conducted with panel members and remote reviewers actively involved in the assessment process as well as with RFO staff and management members. All interviews were conducted in English, between December 2020 and November 2022. They were recorded and subsequently transcribed. Furthermore, five panel meetings in two RFOs were observed in a non-participant form, one meeting took place face-to-face, while four were conducted online, bringing specific challenges for data collection (Utoft et al. 2021, Howlett 2021). The integration of observational and interview data allows for comprehensive and detailed insights into panel work.

The data analysis was based on qualitative content analysis (Mayring 2015), where initially the transcripts from one RFO were split into small text units which were coded mainly along topics from the interview guideline (Silverman 2006). Then, a thematic analysis (Braun and Clarke 2006) was conducted to identify overall categories and themes and to develop more patterns. This analysis was used to create a coding tree in MAXDA for coding the other four RFOs, reflecting the themes and adding various meta- and sub-level codes associated with gender bias in assessment practises and the strategies employed by RFOs to address and minimise bias.

4. Findings – implementing innovative policies in practice

In the following, three innovative policies that RFOs have implemented recently to mitigate gender bias are discussed in more detail, first describing the policy context, then pointing out how panels have implemented this formal policy in practice and finally discussing some risks and potential learnings.

[4] Other parts are: policy analysis, applicant survey, RFO data analysis

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

4.1. Re-ranking

4.1.1. Context

When discussing the gender fairness of research assessment, success rates of female and male researchers are often checked for being equal. Most RFOs publish sex-disaggregated data that illustrate the proportion of researchers that apply for funding categorised by sex/gender. This data is monitored on an annual basis and is often categorised by funding programme or call. The purpose of publishing this data is to demonstrate that the chance to be funded is equal/very similar for female and male applicants. Following a broader European perspective, however, female researchers continue to experience comparatively lower levels of success, with their average success rates being 3.9 % lower than that of their male counterparts (EC 2021b).

A different approach is not to address the success rates – and by this, the ratio between applicants and grantees by sex/gender – but the absolute number or the share of female grantees compared to male grantees. For this, a re-ranking policy can be implemented, aiming to have more women granted by changing the ranking order of applicants approved for funding based on their sex/gender. Re-ranking policies intend to achieve gender-balanced funding outcomes at the end of the assessment process, following the argument that female researchers face career-related disadvantages along their career trajectory which need to be compensated for: "We hereby correct for bias prior to the assessment" (policy officer). A higher number of woman grantees can impact the representation of female researchers in two ways: First, a higher number of woman grantees can signal and serve as incentive for female researchers that success is possible, as a fixed number (or share) of female applicants is granted. This can increase the number of female applicants. Second, with a higher number of female grantees, more women are able to build a successful research career and to get into leading and decision-making positions. This can have a positive impact on the integration of women in the science system, in particular in fields where they have been under-represented so far.

In each of the three RFOs with a re-ranking approach, the formalisation[5] was different. Re-ranking in SRC means that applications by the underrepresented gender shall be prioritised when several applications are assessed as being of equal quality (SRC 2020). In FWF, a gender-balanced funding quota was implemented subsequent to the discontinuation of women-only programmes. SFI aims for more female applicants and has implemented a re-ranking approach, stating: "When ranking applications, in the event of applications receiving the same final score, SFI will give priority in the review process to applications from female candidates" (SFI 2019).

Recent academic discourse argues that equal success rates are not due to balanced assessment scores but specific policies aiming to re-order the applicants to be funded. Scholars claim that re-ranking strategies have resulted in the denial of financing a higher number of exceptional male researchers (Bol et al. 2022). Although we do not endorse these arguments, we were intrigued by the opportunity to examine the practical implementation of re-ranking and the deliberations and application of this policy by panel members.

4.1.2. Implementation in practice

In FWF, achieving a gender-balanced funding quota was tested in one funding instrument, intending to mainstream this gender equality policy into other funding programmes later on (Zimmermann 2023). In this funding programme, at the end of each panel meeting, taking place five times a year to assess applications received on a rolling basis, it was mandated that 50 % of the grant recipients are to

[5] In 4.1.2 we discuss in more detail how it is done in practice.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

be women, regardless of the proportion of female applicants. Regulations were specifically developed pertaining to the proportion of female applicants in areas traditionally dominated by males.

When observing three panel meetings, we found that panel members refrained from mentioning or questioning this re-ranking policy during negotiating applications. Only after all applications were discussed, the chair took care of achieving the quota by highlighting the initial gender distribution of the already approved applicants, consisting solely of applications assessed with grade A[6]. Due to fixed budget, the number of grantees was already fixed prior to the panel meeting. In one panel meeting, until the final number of applications, which could receive funding, had been reached, applicants assessed with grade A- were subject to further considerations. In this second round, it was spelled out how many female applicants are needed to fulfil the quote, and accordingly, this number of female applicants was selected from the pool of A-applicants. The practice was to add A- female applicants in the appropriate number, so no re-ranking took place. In a second panel meeting, the number of female and male applicants was equal after the first round of discussions. Consequently, no re-ranking had to be conducted. A negotiation was only needed in the third panel meeting. However, it was not possible to observe this discussion as the applicant involved had not provided informed consent to take part in the GRANteD research.

In SFI the re-ranking process was executed without any panel members involved. Here, re-ranking was performed by an automatic algorithm-based system. After all applications had been discussed, a first ranking list was generated, based on the fine-tuned average scores entered into the IT system. Then, women were prioritised within predefined subgroups called 'bands'. Bands were built on a 0.5 range (e.g. 5.0, 4.5, 4.0), average final scores of applicants were rounded to next 0.5 – a process called tie-breaking. That panel members remain unaware of the applicants who undergo re-ranking can be seen as an attempt to objectify the re-ranking process, isolating it from personal and potentially emotional discussions.

In SRC, re-ranking is described as a borderline condition to prioritise the underrepresented gender in cases of equal scientific quality (excellence) of applicants. In this case, the ideal outcome is that the number of grantees is balanced in terms of gender. However, interviews with panel members revealed that re-ranking was not practiced in each panel meeting and, furthermore, application of the re-ranking policy varied among different panels. In general, panel members encountered a lack of clear instructions regarding the implementation of re-ranking procedures, and explanations regarding the underlying policy rationale were not fully clear. One explanation provided for why the policy was not always and consistently applied was that a three-year average was used for monitoring success rates, which meant that slight gender variance in success rates were accepted as long as it did not result in a long-term trend.

4.1.3. Discussion

As discussed above re-ranking policies may take on different forms, but they all pursue the same overarching goal, namely a more gender-balanced outcome of the assessment process. Some policies leave more scope for action (e.g. SRC), others are designed in a way to strictly achieve the target (e.g. FWF).

Indeed, the implemented quota in FWF is a strong policy and a clear signal for increasing the representation of female researchers. As the panel has to achieve the fixed number of female grantees, there can be a tendency during the negotiation phase of the panel meeting to prefer female applicants.

[6] Grade A is the best grade.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Applying standards differently to female applicants, without clear guidelines and in an unsystematic manner, might lead to gender bias. Further, the focus on gender-balanced funding outcomes might reinforce panel members' focus on the applicants' gender. When taking care of gender in the assessment process was focussing on approving the requested number of female grantees, less attention was paid to a gender-neutral assessment process (gender sensitive language, GiRI etc), gender was rather limited to fixing the number of female grantees. It was further noticed that a quota might reinforce the horizontal segregation pattern. In the case of the FWF, for example, it got obvious that more females were granted in female-dominated disciplines (e.g. Humanities/Social Sciences and Biology/Medicine). Conversely, a high number of male grantees was observed in Natural and Technical Sciences. Accordingly, in these scientific fields with a minority of female researchers the quota did not manage to increase the share of female grantees. Consequently, it can be suggested to limit the implementation of quotas to specific scientific disciplines rather than averaging it across all fields.

In contrast to FWF, SRC leaves more scope of action when it comes to applying the re-ranking policy. The monitoring of success rates follow a three-year average. Slight gender variances in success rates are accepted as long as they do not lead to long-term trend. Furthermore, some panellists understand the re-ranking policy of the SRC as a further layer of control in the assessment process. This control layer checks not only for gender, but also and above all for the quality of the funded projects.

SFI's 'technical' approach towards re-ranking attempts to objectify the re-ranking process by isolating it from personal and possible emotional discussions. However, automatic algorithm based system may be prone to errors.

Our findings demonstrate that trying to fix the number of women grantees by utilising re-ranking policies shows that RFOs are concerned about gender unbalanced outcomes of assessment processes and that they want to change something about it. However, the findings also showed that these policies need to be monitored carefully in order to see how they are eventually understood by panel members. Monitoring data may give insights if re-ranking policies need adjustments of their design or other factors may need improvement, such as circulation of information about these policies.

4.2. Broader excellence

4.2.1. Context

Excellence is a merit-based concept, yet what is seen as merits and how they are measured is often vague and open for interpretation. What is defined as excellence depends on the context and on the specific scientific discipline (Cañibano et al. 2018), which is why different concepts of excellence are in place. As it is not yet clear what excellence means in detail, it is also not obvious how to measure it. However, excellence is usually narrowed down to quantitative terms and measured by bibliometric indicators (h-index, journal impact factors, citations). The h-index is however criticised for not taking epistemic differences between disciplines into account, and for limiting merit to publications, while other forms of knowledge production and dissemination are widely ignored. Moreover, the fact that excellence is discussed as a neutral concept, while being rather vague, can serve as 'rationalising myth' and thus provide space for gender-biased practices (O'Connor et al. 2020).

RFOs are key actors for questioning, (re-)defining and transforming standards and criteria for excellence (Schiffbänker et al. 2023). Quite recently, some RFOs have started broadening their assessment indicators, also taking into account the societal impact or the working culture in research performing organisation, e.g. when asking for mentoring activities in research teams, as practiced in FWF. Some RFOs have signed the DORA declaration (FWF, NCN, SFI), and in response to the increasing demand regarding the challenges in assessing capabilities of applicants in a correct, fair and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

transparent manner, SFI has introduced the narrative CV format. FWF has also adopted the CV format. Innovative CV formats invite applicants to report their merits in a more qualitative, narrative manner and reviewers to assess excellence without using the h-index.

Beyond that, the implementation of GiRI broadens the concept of excellence in respect to research content (see 4.3).

4.2.2. Implementation

When analysing how reviewers assessed excellence, we found that standards are still vague and differ within and across panels, with specific challenges in multidisciplinary and international panels due to geographic or disciplinary heterogeneity. To better align different standards, some panel chairs organised explicit calibration sessions before discussing applications, while others do this informally in the negotiation phase.

When being asked not to use the h-index, reviewers raised several concerns. On the one hand, they argued that such innovative approaches are "questioning the rules of the game" (reviewer, SFI), on the other hand, reviewers claimed that careers are built on publications and thus they need to be the core of each applicant's assessment. Some reviewers underlined the necessity to check back on Scopus or Google Scholar in case a proposal is not in their field of expertise. In this case, they need the h-index to have a quality benchmark for the applicant. Further, reviewers articulated a lack of trust and argued that applicants might not report their merits properly, which is why they need the h-index to verify the given information. Finally, a lack of alternative indicators for replacing the h-index was mentioned. Yet we found that when publications are supposed to count less, previous grants might become more important.

Moreover, reviewers raised concerns about the narrative CV as it could lead to new gender bias, based on gender disparities in self-presentation, specifically pertaining to female applicants who might exhibit lower levels of self-confidence, visible gender differences in writing and presentation style.

For these challenging issues, chairs were expected to provide answers and solutions.

One more aspect for ensuring the best assessment of merit or excellence was the formalised rebuttal introduced at SFI. Here, feedback from remote reviewers was sent to applicants, who were asked to comment on this feedback. In a second assessment round, remote reviewers had to adjust their scores according to the applicants' feedback. The panel and the panel chair were then responsible for checking whether the remote reviewers had integrated the rebuttal from applicants accordingly in their second assessment. This approach might be of high interest as it can be regarded as a more collaborative assessment approach, where applicants are granted the right to clarify specific parts of the application, which might be not that clear at first hand.

4.2.3. Discussion

Re-defining excellence and implementing innovative indicators to measure a broader understanding of excellence (beyond quantitative metrics) was shown to be a demanding endeavour, not least as long-established narratives and indicators are challenged.

We noticed resistances against such approaches and confusion regarding how to do the assessment differently. Chairs were expected to offer solutions for challenging questions – which might indicate that chairs need additional instructions and training. In this respect to language, it might be relevant

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

to analyse stereotypes in language and to update “institutional processes so that they are not negatively predisposed toward female scientific language." (Kolev et al 2020, p. 249).

Accountability is as key factor for (gender) fairness in the assessment of applications, which is why panel chairs or observers (from RFOs or external) should be made responsible for monitoring that assessment criteria are applied in the same way to all applications. Here, a key responsibility for chairs is to communicate with and provide guidance to panel members in order to develop a common understanding for how excellence is understood and assessed – with and without the h-index.

When implementing new and more experimental assessment criteria, sound information, guidelines as well as training and, above all, time and practice are needed. Therefore, practices of reviewers should be continuously monitored in order to identify potential for improvement.

4.3. Gender in Research and Innovation (GiRI)

4.3.1. Context

‘Gender in Research and Innovation’ (GiRI) is a rather new policy, aiming to integrate a gender perspective in the production of research content and innovation. For this more inclusive form of knowledge production, RFOs are key players as they “are responsible for promoting excellent research that benefits all of society (…) and make research more responsive to social needs” (Hunt et al. 2022, p. 1492). GiRI is an approach to increase and strengthen quality of research by addressing gender in the production of knowledge and innovation. Hence, it implies a certain need to change the ways of doing research (for applicants) and of assessing research (for reviewers). GiRI asks for includeing gender as a quality aspect and as key analytical and explanatory variable in research. Only when this broader approach is taken into account, research can be understood and evaluated as excellence. This way, GiRI in fact transforms how the concept of excellence is defined.

The policy builds on empirical evidence revealing that knowledge production and innovation are narrowed when researchers fail to incorporate diverse perspectives taking into account women and other minority groups along the research cycle. Schiebinger and Schraudner argued that “fixing knowledge by incorporating gender analysis into basic and applied research (...) assure(s) excellence and quality in research outcomes” (2011, p. 154). Sex and gender are relevant when formulating research questions, collecting, analysing and interpreting data as well as disseminating research outcomes and innovations (Håkansson and Sand 2021, Gibney and Schiebinger 2020). Adverse consequences can arise from the omission of sex, gender and diversity analysis, as the well-known example of Volvo illustrates, where the company improved its quality standards by developing new seat belts to fix this problem after research revealed that most automobile companies’ crash test dummies were not designed for pregnant women.

The European Commission has paved the way for this policy by incorporating it into the framework programme 6 (FP6), providing support material already in FP7 (EC 2011). Subsequently, the integration of the sex and/or gender dimension in knowledge production has become a policy target for the EC, also reflected in the ERA principles. Similar policies can now be found in various national RFOs.

When examining the implementation of GiRI policies in the three RFOs under study where they are in place, some variations in the organisation of the assessment process could be identified. In two RFOs (FWF and SFI), GiRI was scored as a sub-element of excellence by remote reviewers, while in SRC this is done by panel members. Hence, we were interested in how reviewers responded to this policy and what criteria they considered when assessing how GiRI was included by applicants in their applications.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

4.3.2. Implementation in practice

We found that most reviewers exhibit a lack of awareness of this policy. At the same time, the policy itself and the underlying aims were considered to lack clarity. Reviewers were often unfamiliar with the GiRI policy and had trouble understanding GiRI and its aims, as the following quote illustrates:

> *I have never been asked this question anywhere before. They [the RFO] wanted to know whether the research is going to impact gender .... I said this PI might be female. They said: no that's not what we want. We want you to comment on the projects' impact on gender equality." (remote reviewer)*

Instead, GiRI was associated with other gender concepts, like the numerical representation of women, the number of female researchers within teams, the low number of female applicants and the overall underrepresentation of female researchers in specific scientific fields. Reviewers active in countries with a tradition in affirmative action policies brought up the idea to extend the effort to increase the representation of women in research as previously done for other marginalised groups. Remote reviewers, particularly those with limited or inadequate awareness or expertise on gender, frequently misinterpreted GiRI as a metric for evaluating the composition of teams in terms of gender balance. Here, the reviewers' disciplinary background is relevant. Our findings indicate that the awareness and understanding of integrating more divers perspectives into research and innovations differed by scientific field, as argued by a reviewer in Life Sciences for who it is a new approach: *"Typically in my area of research we don't think about that at all, so that made me think for a while"* (remote reviewer). Accordingly, the assessment of the GiRI approach was challenging and it can be assumed that the assessment of reviewers was frequently not well-informed. This can be observed in reviewers hardly questioning the great number of applicants denying the relevance of GiRI in their research proposals.

Furthermore, the awareness of GiRI was also limited in peer review panels. In the observed panels, GiRI was hardly addressed. In one case, people just laughed when a panel member argued in a Life Science project that both male and female mice are used as probands. There was a clear lack of understanding as to why having both sexes for testing is beneficial and why it can increase quality and excellence.

4.3.3. Discussion

Studying how the GiRI policy was implemented in practice provides some learnings for finetuning the policy and how to introduce it. In this context, interventions are urgently needed as GiRI is still widely linked to the representation of women and not to gender in knowledge production, not so different to what was found in the FP7 interim evaluation (EC 2017).

One finding refers to the necessity to comprehend the policy objective. As GiRI is an innovative and maybe even transformative policy approach to incorporate more diverse perspectives throughout the research cycle, RFOs and other stakeholders of the science ecosystem need to increase their efforts in elucidating and effectively communicating the narrative of this policy (Schiffbänker 2023). Explaining the underlying arguments and presenting good practice examples from various scientific fields might help reviewers to develop awareness and creativity regarding how to assess GiRI as well as producing support material provided by RFO that contains guiding questions or illustrative examples (e.g. videos, good practice examples, guidelines). Reviewers need gender awareness and competences to assess whether a research proposal in a respective field addresses gender adequately, otherwise GiRI might be assessed inadequately or not at all. They need to know that most of the research ideas/proposals can be addressed from a male, a female or an intersectional perspective. As this might be challenging

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

in particular for reviewers in the 'hard' sciences capacity building needs to start there. Nevertheless, by taking this complexity into account, GiRI becomes an element to increase scientific excellence.

Capacity building is recommended for all stakeholders, reviewers, RFO staff and applicants. Collaborative capacity-building activities across disciplines can help to align the understanding of GiRI among reviewers from different scientific, social and cultural backgrounds.

Alignment in the formal implementation might further be needed across RFOs. Reviewers who are active in various RFOs argued that the assessment of GiRI was even more demanding as this formal policy was implemented in slightly different ways in the various RFOs, and that also the indicators to assess GiRI varied. Implementing uniform formal approaches across RFOs may thus potentially enhance the capabilities of reviewers for effective implementation.

Finally, it is important to mention that the GiRI policy, which aims to bring diverse perspectives to knowledge production, cannot be limited to gender. As gender intersects with other social categories, like race, age as well as cultural and ethical background, the GiRI approach can be expanded to those dimensions, as emphasised by Boytchev (2023) and Kraemer-Mbula (2020). This expansion entails the utilisation of more comprehensive and nuanced individual data sets, as well as the incorporation of additional indicators, in order to effectively spread this strategy (Woods et al. 2021).

4.4. Chairing

4.4.1. Context

When discussing how innovative policies to foster gender equality in research funding have been implemented in panels, it became evident that panel chairs play an important role in this regard. Hence, we take a closer look at their role, pointing out what panel members and RFO members expect from chairs in general and with respect to avoid (gender) bias.

The role of peer review panel chairs is "to facilitate discussion, moderate individual personalities and provide a fair and balanced presentation of each proposal" (Gallo et al. 2020, p. 2). Bethencourt et al. (2021) highlighted the crucial function of chairs in facilitating consensus in panels, emphasising that chairs not only organise and shape the debate but also influence the interactions among reviewers. Already back in 2001, Heilman argued that power dynamics inside panels are in place which can be attributed to a lack of structure in the decision-making processes of assessment bodies. This allows chairs to exercise a certain degree of discretion, particularly when it comes to adjusting the score subsequent to the panel discussion (Langfeldt 2022). Regarding gendered dynamics, chairs are perceived as possessing the potential to significantly influence structural aspects of the decision-making processes by determining which individuals to invite for input and at which stage of the negotiation process (Vinkenburg et al. 2021).

4.4.1. Implementation – Chairing in practice

Our empirical data illustrate the various tasks chairs are taking over in panels – very general tasks and tasks specifically relevant for implementing innovative policies. Before discussing the role of chairs for implementing new policies, it might be interesting to initially highlight some aspects of this role (see 4.1.1.1). Readers not interested in general expectations on chairs are invited to continue reading in 4.4.1.2 which focuses more specifically on the role of chairs when re-ranking, broadening excellence and implementing GiRI.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

4.4.1.1. Excursus: Chairing in practice – general [7]

To understanding the role of chairs in panel practice and the agency in this role, it seems relevant to know by which criteria they are selected. More precisely, by what criteria chairs themselves think they have been been selected. This might give some indication about what they see as relevant in their enactment as chair.

It is important to acknowledge that the formal process of selecting a chair is not standardised. The formal appointing regulations vary among the five RFOs investigated. For instance, in the SRC, the Director General appoints each chair, following the recommendations of the Scientific Councils' Secretaries General[8], while in the Austrian FWF, the president assumed the role of chair for all analysed panels.

With reference to a discourse on the representation of women in leading positions (see 2.1.1) it might be interesting to point out how women were represented in the function of panel chairs. When considering the inclusion of women in this significant role, RFOs exhibit varying degrees of success: Among the 39 panels covered by our research, 15 (38%) were chaired by women. However, there are notable disparities across different RFOs, as depicted in the following table:

Table 2: Female panel chairs

	Number of chairs of panels studied	Number of female chairs	Share of female chairs
FWF	3	0	0 %
NCN	25	11	44 %
SFI	2	1	50 %
SRC	3	2	66 %
SRDA	6	1	17 %
Total	39	15	38 %

In a first step it will now briefly discussed how chairs describe their responsibilities themselves and how panel members and RFO representatives see them. In our interviews we got interesting information about the personal understandings and experiences, as well as more factual information about the role in the panels.

When asked why they are in this position, panel chairs often could not clearly state the reasons for their appointment, speculating that it may have been due to their prior experience as vice-chairs or as panel members, or that they were probably put forward as candidates by other chairs. A female chair thought that it was simply "because I am a woman".

Further, some chairs also were rather unsure about their responsibilities and functions. This became evident by the very different descriptions of their roles provided by two chairs affiliated with the same RFO. This lack of clarity can create a situation where chairs interpret and manoeuvre their roles and responsibilities quite differently. At the same time, it points to a limited level of formalisation and accountability on the side of the RFO. Some chairs frequently participate in different RFOs that involve distinct assessment processes and different criteria. Accordingly, chairs participating in various panels have diverse responsibilities and may not necessarily adjust their understanding and assessment

[7] This part might become a separate paper later on.

[8] https://www.vr.se/english/about-us/organisation.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

practices to the requirements of each RFO in a consistent manner. Furthermore, we found that the tasks of chairs differ according to the role a panel plays in the assessment process: While some panels solely focus on discussing reviews rather than applications, others are responsible for final decision-making regarding funding, which brings divergent responsibilities for chairs. A further distinct element is that in some RFOs, clear support structures for chairs are defined, e.g. vice-chairs or administrative chairs. Additionally, other panel members take over defined formalised roles supporting the chair, e.g. internal or external observers to monitor gender issues, or other RFO staff members who have specific tasks, often closely linked to the funding programme.

Guaranteeing a fair assessment process free from gender bias is one core requirement imposed on chairs. In addition to organising and facilitating the meetings as well as steering and summarising negotiations in the panel, the chair is expected to align diverging positions among the panel members and reach a consensus to ensure unanimous support for the final funding decision or recommendations. In order to achieve this in practice, some chairs set up explicit calibration sessions prior to engaging in discussions regarding applications. Guided by the chair, panel members develop a (more) common understanding on how to effectively apply assessment criteria and to align different interpretations and indicators, avoiding that the same indicators are applied to measure different issues (e.g. for independence see Schiffbänker et al. 2022b). Moreover, chairs are expected to guarantee equal standards, which refers to adjusting the relative weight of the various formal assessment criteria in order to avoid informal weighting, such as merely considering publications, but also to aligning how scores are attributed by different panel members, as it was reported that some seldomly use the maximum score („just for a new Bourdieu...“), while others do it quite often. If different standards were to be employed, it is envisaged that chairs point this out and ensure that applications are rediscussed or reassessed. Insisting on reassessment might also be necessary when panels highlight specific strengths or weaknesses of an application, while these aspects are ignored in other negotiations. Also in cases where criteria were assessed ambivalently (e.g. when ‘over-ambitious’ was once seen as a positive aspect but identified as a flaw for another applicant), chairs are required to guarantee equal standards by pointing out these discrepancies.

In some panels, chairs have to handle national heterogeneity of panel members. When panel members are from different countries, the chair needs to harmonise different standards and connect them to the corresponding national norm. In the five RFOs investigated, this particular challenge took various shapes. While FWF panels are comprised *only* and SRDA and SRC panels *predominantly* of national reviewers, international reviewers were dominant in two other RFOs. Recently, there has been a transition in Poland, where NCN has changed its panel composition to include a greater number of international members. SFI panels exclusively consist of reviewers who are not Irish, but from various international backgrounds.

Beyond national heterogeneity, chairs also have to handle disciplinary heterogeneity in interdisciplinary panels or when panels cover a broad range of disciplines. The diverse scientific background of panel members gives rise to distinct epistemological cultures, which e.g. may lead to varying criteria for evaluating scientific excellence.

Furthermore, chairs are required to accommodate diverse personalities, individual preferences and experiences. For newcomers, chairs are expected to provide guidance and answer any inquiries beginners might have.

Managing power dynamics and gender dynamics is a further function attributed to chairs, ensuring that the assessment process and outcomes are unaffected by any associated bias. In respect to (gendered) power structures, time use in panel meetings is one important element. The time individual panel members allocate to push the applications they present and to comment on other applications

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

reflects the underlying dynamics within the panel. To balance differences and to guarantee even allocation of time, chairs need to structure the length of discussion, balance the time spent for each application as well as the time allocated to each panel member. We observed that some panel members make longer statements, defending and revisiting their positions several times, hence prolonging the deliberation process. This phenomenon was observed both among female and male panel members and might be based on their individual personality or their respective statuses or positions within the panel. In this context, chairs are expected to effectively mitigate the recurrence of arguments and lengthy monologues. For example, a panel chair interrupted a negotiation by arguing: "We stop here, we should allocate the same time to each application". While in most panels, keeping an eye on time is the responsibility of chairs, in some RFOs it is done by staff members or vice-chairs. Overall, it has to be ensured that female panel members are given equal opportunities in terms of time and space.

Another aspect that requires the attention of panel chairs are the different communication patterns of panel members. On the one hand, chairs need to encourage panel members to engage in the negotiation process and test their commitment: "Are you ready to fight for this proposal?". It is crucial to avoid that "the person with the less loud personality gets muted" (panel member). On the other hand, chairs are well aware of highly-engaged panel members: "Every expert wants to show that he or she did a very good job and went through the whole proposal and pointed out all the drawbacks of the project. […] They all want to show that they prepared very well for the discussion" (panel member).

Based on respective RFO policies, a further important task of panel chairs can be to ensure that no gendered language is used and to emphasise and promote a gender-neutral language. One RFO has implemented a policy that prohibits the inclusion of an applicant's name, sex or pronoun in the panel negotiations in order to avoid activating gender stereotypes. Panel members are not allowed to talk about an applicant e.g. by saying "*she* has published...". In this case or when personal information was used, the chairs – supported by RFO staff members – intervened several times to remind panel members to only use the applicants' identification numbers (ID) or the neutral term 'the applicant'.

Moreover, panel chairs have obligations when it comes to organising the final decision-making process. When deciding about an applicant's score and determining whether to fund an application or not, chairs have set up informal practices for how to achieve the final decision, based on the specific formal assessment process employed by the RFO. However, it was sometimes not obvious in the panel observations when exactly the panel deliberation has achieved a sufficient consensus so that the chair can close the negotiation. To increase transparency, RFOs and chairs are encouraged to explicitly communicate the process how to do the final decision making (raising hands, announcing the decision and waiting for possible objections etc). Typically, a more systematic approach involves the implementation of formal guidelines to govern the process of reaching final decisions.

Finally, panel chairs had to handle remote chairing during the Corona pandemic. As four of the five panels we observed met remotely, we are not able to compare online and onsite panels; yet we were able to witness that the utilisation of online platforms like Zoom or Teams facilitated the implementation of a quite strict moderation style, as chairs deliberately were giving the floor to one specific panel member. From a gender perspective, this could have both negative and positive effects: On the one hand, it was easier for chairs to ensure that all panel members were allocated equal speaking time and that they were not interrupting each other. On the other hand, it was harder to engage all panel members due to the more formal and "business-like" atmosphere (Peterson und Husu 2023, p. 374).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

4.4.1.1. Chairing in practice – innovative policies

After having discussed some general expectations towards panel chairs, we now analyse in a second step how chairs facilitate the implementation of innovative policies which have changed the assessment process (sometimes radically).

When we observed panel meetings, several policies were new to panel members: One was the implementation of SFI sitting panels which limited the role of panel members on assessing solely the quality of remote review(er)s, but not the quality of proposals anymore. In this case, chairs intervened when panel members brought in their own quality standards.

A second example was the implementation of a narrative CV and the request not to use the h-index for assessing applicants. For implementing this policy, the chairs were only limited in charge. This approach was most advanced implemented in SFI, but as here the role of the panel was adopted (see paragraph above), the panel chair was more concerned with the new panel function and less with the narrative CV. In FWF, SRC, SRDA and NCN the narrative CV was no topic when discussing the panel work. In a few cases chairs provide support und explanations to panel members and practical examples of how to navigate under new circumstances.

The chair also had a crucial role when a quota for gender-balanced funding outcomes was introduced in an FWF funding programme. It was the chair who highlighted the gender distribution after the first round of discussing applications. In a next step, the chair also steer the process of selecting the number of female applicants requested to meet the quota. In SRC the chair has the role to highlight when a re-ranking is needed. For the SFI re-ranking, chairs do not play a role as it is done IT-based.

Finally, the implementation of the GiRI policy in panels (see 4.3) was and is still challenging. While the scoring for GiRI as an element of excellence was predominantly done by remote reviewers, it was striking that GiRI received minimal attention during panel negotiations, chairs showing limited awareness for highlighting and promoting the relevance of this concept. Yet in one case a chair stopped an unfruitful discussion on GiRI, but did not stress why GiRI is relevant in the assessment process.

These examples illustrate how the implementation of innovative gender equality policies impacts the work of chairs and how thereby they steer the discussions and choose their priorities, following the objective to reduce gender bias.

4.4.2. Discussion: Re-thinking chairing

When analysing the role of chairs in peer review panels, we see that they operate in two ways. Panel chairs are assuming the leadership role in steering peer review panels, guided by various formal RFO gender equality policies and other regulations in place. This way, chairs are supposed to ‘serve’ RFOs by implementing the formal rules and policies set in place. Chairs are expected to transfer them to panel practices. As these policies are constantly adjusted, chairs need to adapt accordingly to be able to effectively put new GE policies into practice.

In parallel to that, chairs are also expected to facilitate and organise the panel work in order to enable panel members to do their assessment in an optimal manner. In this respect, they ‘serve’ the other panel members. So overall, chairs are intermediaries between the needs of RFOs and those of (RFO-external) panel members.

When RFOs have strong GE policies in place, the chairs – sometimes supported by observers or officers – are the ones who facilitate the practical application of these policies. Our research demonstrates that effective leadership is in particular crucial – and would even need to be more crucial - for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

implementing new gender equality policies which aim to transform narratives (GiRI, new formats to report merit) or to restructure assessment procedures, as our empirical findings demonstrate (re-ranking, no h-index). In this case, chairs provide support but also decide whether to intervene in case a new policy is violated or alternatively choose to grant time, allowing panel members to familiarise themselves with the new requirements and to adjust their practices accordingly.

Specifically for gender-fair funding outcomes (Vinkenburg et al. 2021), panel chairs might have a relevant role when monitoring the gender distribution of applicants throughout the various steps of the panel work. For this, some chairs use informal gender-specific ranking lists. Even so, at this point we should have in mind that as most panels met virtually, decision making faced specific challenges (Ortiz de Guinea et al. 2012).

Overall, chairing has been demonstrated to be a multifaceted endeavour, which is anticipated to grow in complexity in the near future, when considering the implementation of qualitative evaluation methods and/or the transition from a gender-focused perspective to an intersectional framework.

However, in a survey among reviewers, one out of three panel members expressed dissatisfaction with the performance of the chairs on the last panel they attended (Gallo et al. 2020). So it seems imperative to enhance the support provided for chairs, like specific chair trainings that are still rare. Also more instructions and support material might be needed. In this respect, it was highlighted by RFO staff that a range of instructional materials and guidelines was already provided, yet interviews revealed that support material was hardly known or applied by the potential users. It could be insufficient or unsuitable, so creating new attractive formats might be worth considering.

Finally, to increase impact, the role of chairs should be strengthened by formalising the tasks and responsibilities and by defining clear accountability. This can help improve panel work as up until now, limited awareness for this role and for the expectations that panel members and RFOs have with respect to chairs was identified. Tasks could also be shared among multiple actors (vice-chairs, administrative chairs, RFO coordinator, scientific officer), in line with Coveney et al. (2017), who found that appointing a 'technical' and administrative chair focussing mostly on the facilitation of the discussion is well received by panel members.

Correspondingly, reasons why innovative gender policies are not (adequately) applied in practice might not only be due to a lack of capacity and a limited understanding of gender – which could both be mitigated by trainings and instructions. Rather, one reason for this might be that panel chairs (and panel members) do not support gender equality policies, although they do not overtly oppose against these RFO policies. In this case, we cannot expect that when more tools (precise explanations of the meaning of the respective measures, trainings, best practice examples, etc.) are provided, the implementation will be more successful.

5. Conclusions

In recent decades, national authorities and research funding organisations (RFOs) have been increasingly active in fostering gender equality in science and in mitigating gender bias in research funding. Various policies have been implemented to – according to Schiebinger/Schraudner (2011) – 'fix the numbers' (of applicants, grantees, reviewers), to 'fix the institution' (the assessment process) or to 'fix the knowledge' (assessment criteria for excellence/merit). While considerable differences still exist between countries and RFOs, based on national and cultural context factors (see D5.1), some RFOs have proven 'pioneers' in putting innovative policies in place. Hereby, RFOs set new standards for applicants on how to do research and for reviewers on how to assess research. It was shown that

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

RFOs and their advanced gender policies are in a way 'spearheads of change' as they try to transform fundamental narratives of the research ecosystem (GiRI, narrative CV, re-ranking and others).

These efforts for a gender-fair assessment match with ongoing activities to reform the research assessment process, like DORA, the Leiden Manifesto (Hicks et al. 2015) or CoARA, pursuing similar goals. Also in this context, fostering inclusion is key by encouraging applicants to report broader contributions to the science system and by fostering a qualitative research assessment accompanied by quantitative indicators.

In our research on how written, agreed-upon policies are implemented in concrete everyday practices in peer review panels, we have identified some learnings that might also be useful for other policies to be implemented and in particular when implementing new policies for reforming the research assessment process. We have analysed three innovative policies:

(1) How re-ranking policies aim to balance funding outcomes and increase the number of grantees from underrepresented groups (women and others).
(2) For broadening the concept of excellence, the narrative CV was implemented. When studying its assessment, we learned that reviewers struggle with giving up the h-index as benchmark for assessing merits and as proof in case they do not trust what applicants reported qualitatively.
(3) Another challenge is the assessment of the request to take gender into account in all research proposals or to argue why it is not relevant. Reviewers were lacking clarity on this policy and its aims, and accordingly, assessing it in an informed manner was hardly possible. It needs to be underlined that most reviewers lack awareness about gender as a dimension of structural inequality and as an analytic dimension in their respective research fields. So this chance to broaden the understanding of excellence by integrating the sex/gender dimension into the research content and innovation processes, and to contribute to more research outcomes that are beneficial for all subgroups of society is not taken up yet.

So far, such policies can be found in RFOs with advanced gender equality policies, while in the Central-Eastern European countries we studied (Poland, Slovak Republic) there is still only an emerging approach to gender equality.

Studying the implementation of these innovative policies in practice revealed that putting innovative formal policies in place by RFOs does not mean they are necessarily implemented in practice in peer review panels. We found that the different stakeholders – panel chairs, panel members and remote reviewers – were not always sufficiently 'prepared' to implement new policies effectively. They lack understanding of the policy aims and of the way how to apply innovative policies in assessment practices. This suggests general learnings with respect to two areas of intervention: (i) learnings for policy design and effective implementation and (ii) learnings for awareness raising and for building reviewers' capacities.

5.1. Learnings for reforming the assessment process and chairing

Firstly, we identified some aspects for reforming the assessment process including chairing of panels. These learnings provide valuable information for adapting the design and the introduction of policies.

(1) Clear concept: When implementing a new concept or policy, it should be formulated precisely, with a clear political target, arguing why it is implemented and what bias factors it aims to address or which improvements are intended. Otherwise reviewers are not able to implement the policy effectively, as a lack of clarity promotes non-uniform standards. For well-informed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

comprehensive reviewing, it is necessary that the concept be clear, well communicated, easily comprehensible, readily adaptable and supported by accessible resources.

(2) Comprehensive communication of policies, policy aims and addressed gender risks: Our data indicates that when innovative policies are introduced, more focus should be placed on explaining and communicating the aims of this policy to reviewers, applicants and the wider research ecosystem. This would facilitate a deeper understanding among reviewers, hence enhancing their ability to effectively implement and endorse these policies. Reviewers might benefit from sound and easy-to-understand information on the policy to be implemented, which could help reviewers to better understand what to look at in the assessment process.

(3) Aligning policies across RFOs: When implementing new policies (GiRI, narrative CVs, etc.), RFOs are encouraged to develop common standards and to align assessment indicators. Uniform standards make it easier for reviewers to adapt to new policies in practice, and later enables a better comparability of impacts in evaluations. Otherwise, reviewers active in various national and international RFOs have to adapt to the specific demands of each RFO, which requires additional time and awareness.

(4) Alternative assessment, alternative indicators: When excellence is aimed to be broadened and merits are reported qualitatively in the narrative CV – which is in form of a story instead of a list – reviewers need information and instructions on how to adapt the assessment in practice (Fritch et al. 2021). When an established indicator is supposed to be abandoned, like the h-index, new approaches like 'responsible metrics' (Wilsdon et al. 2015) and alternative metrics, like citations in the news media or policy documents, should be provided to support establishing adjusted application behaviours.

(5) Reforming chairing and accountability mechanisms: Our research revealed that panel chairs play a crucial role in how panels work in general and also when it comes to the implementation of innovative policies. As chairs steer the negotiation and decision-making process, they are supposed to make sure that RFO policies are transferred to panel practices. As our data shows, there is limited awareness of this role and of the expectations from panel members as well as from RFOs. Thus, it is suggested to strengthen the role of chairs by formalising it through stating a clear role description, offering clear communication and specifying accountability. Further, specific briefings when taking over a new task could help as well as tailored capacity-building activities for chairs. Chairs are also supported when clear formal regulations are in place, e.g. explicit guidelines for the structure of panel meetings (SRC 2015).

(6) Increasing complexity: Gender is not the only inequality dimension that is relevant in the assessment process. Other dimensions intersecting with gender, like ethical and social-cultural background, disability or age, also need to be addressed in the near future, enabling a broader approach to diversity and inclusion.

(7) Beyond this, it might be worth thinking about establishing a more responsible approach of assessment. In short, the concept of response-ability refers to the ability to cultivate collective doing (Bozalek/Zembylas 2017, p. 63 f.). Thereby, the chair and the applicant enter into a kind of dialogue to advance research, overcoming the classical assessment from the distance, "opting instead for more modest, generous, imaginative and inventive responses involving curiosity and attentiveness or atonement, crafting new ways of doing, thinking and living in academia" (Shefer et al. 2023, p. 154). Response-able ways of reviewing "maximise the author's ability to respond to the comments made by the peer reviewer" (Shefer et al. 2023, p. 154). This idea can already be found in SFI's rebuttal where applicants respond systematically to the feedback from remote reviewers and the latter have to include the rebuttal information in their second assessment.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

5.2. Learnings for reforming awareness raising and capacity building

Secondly, we found that the implementation of innovative policies, the changing requirements and the increasing complexity that reviewers are faced with when assessing research grants ask for more activities to build the required capacities and raising awareness. Overall, reviewers need to be able to perform their reviewing in an informed manner to fight "distrust in peer review panels" (Langfeldt et al. 2023).

(8) Offering capacity-building activities along new requirements: By strengthening the capacities of reviewers and enabling them to ‚catch up' with new policy requirements (GiRI, new CV formats, innovative assessment indicators, intersectional perspective), their awareness can be raised and competence increased.

(9) Informing remote reviewers: Remote reviewers do not have the opportunity to engage in discussions with other panel members during meetings and thus have limited opportunities to familiarise themselves with new assessment requirements. This is why a specific focus should be put on their capacity building, even more so as they are not as easy to address and get committed for capacity-building activities, in particular when reviewing only occasionally.

(10) Attractive formats: To foster participation, RFOs are invited to create training formats that are effective and attractive for reviewers. For instance, preparatory sessions prior to panel meetings or self-checks for remote reviewers are best practices. For GiRI, online training with evidence-based capacity-building methods have been developed in medical research. There, pre- and post-training evaluations show that most participants increased their knowledge on how to integrate sex and gender in grant proposals (Tannenbaum, van Hoof 2018).

(11) Providing space to reflect on how to apply information in practice and how to concretely implement one's personal role, sharing experiences and fostering mutual learning might enhance reviewers' reflexivity, understood as "an ability and willingness to examine one's own presuppositions and to take on board novel perspectives" (Argyris and Schön 1978, p. 263).

(12) Nevertheless, capacity building might have some limitations: We need to be aware of the fact that chairs who take on this work-intensive and time-consuming task do this rather rarely to promote gender equality and are often eager to control their own 'zones of influence'. So a reason why innovative gender policies are not (adequately) applied in practice might not only be that reviewers lack capacity and have a limited understanding of gender, but also that they are active not supporters of GE policies, even while not showing overt opposition to such policies. If this is the case and resistances exist, only limited impact from providing more tools (precise explanations of the meaning of the respective measures, trainings, best practice examples, etc.) can be expected. While resistances might be specific to implementing (innovative) gender equality policies, other resistances might emerge for other policies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

6. References

Ahlqvist, V., Andersson, J., Söderqvist, L., Tumpane, J. (2014): A Gender-Neutral Process? A qualitative study of the evaluation of research grant applications 2014. In: *Swedish Research Council, Vetenskapsrådets rapporter 2015, Stockholm.* https://www.vr.se/english/analysis/reports/our-reports/2015-04-03-a-gender-neutral-process.html.

Argyris, C., Schon, D. (1978): Organizational learning: A theory of action perspective. Reading, MA, Addison-Wesley.

Begeny, C. T., Ryan, M. K., Moss-Racusin, C. A. und Ravetz, G. (2020). „In some professions, women have become well represented, yet gender bias persists—Perpetuated by those who think it is not happening". Science Advances 6, Nr. 26, p 6: eaba7814. https://doi.org/10.1126/sciadv.aba7814.

Bethencourt, A. M., Luo, J., Feliciani, T. (2021): Bias and truth in science evaluation: A simulation model of grant review panel discussions. Online verfügbar unter https://ceur-ws.org/Vol-2838/paper2.pdf.

Benshop, Y. (2021): Grand Challenges, Feminist Answers. In: *Organization Theory*, *12, 1-19.* https://doi.org/10.1177/26317877211020323.

Bol, T., Mathijs de Vaan, M., Arnout van de Rijt, A. (2022): Gender-equal funding rates conceal unequal evaluations, Research Policy.

Bozalek, V., Zembylas, M. (2017): Towards a response-able pedagogy across higher education institutions in post-apartheid South Africa: An ethico-political analysis. In: *Education as Change, 21(2), 62-85.* https://doi.org/10.17159/1947-9417/2017/2017.

Braun, V., Clarke, V. (2006): Using thematic analysis in psychology. In: *Qualitative Research in Psychology, 3(2), 77-101*.

Cañibano, C., Vilardell, I., Corona-Sobrino, C., Benito-Amat, C. (2018): The evaluation of research excellence and the dynamics of knowledge production in the humanities: The case of history in Spain. In: *Science and Public Policy, 45(6), 775–789.* https://doi: 10.1093/scipol/scy025.

European Commission (2021a). Horizon Europe Guidance on Gender Equality Plans. Directorate-General for Research and Innovation. Publications office of the European Union, https://data.europa.eu/doi/10.2777/876509

EC European Commission (2021b): Tracking progress on the path towards gender equality in research and innovation. In: *She Figures, Publications Office*. https://data.europa.eu/doi/10.2777/602295.

EC European Commission (2018) Guidance to facilitate the implementation of targets to promote gender equality in research and innovation, Luxembourg, Publications Office of the European Union.

EC European Commission (2017), Interim evaluation – Gender equality as a crosscutting issue in Horizon 2020, Publications Office, https://data.europa.eu/doi/10.2777/054612.

EC European Commission (2011) Toolkit gender in EU-funded research, Publications Office, https://data.europa.eu/doi/10.2777/62947.

Fritch, R., Hatch, A., Hazlett, H., Vinkenburg, C. (2021): Using Narrative CVs: Process Optimization and bias mitigation, Zenodo. https://doi.org/10.5281/zenodo.5799414.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Fritch, R., McIntosh, A., Stokes, N. und Boland, M. (2019) „Practitioners' perspectives: a funder's experience of addressing gender balance in its portfolio of awards". Interdisciplinary Science Reviews 44, Nr. 2 p. 192–203. https://doi.org/10.1080/03080188.2019.1603882.

Gallo, S. A., Schmaling, K. B., Thompson, L. A., Glisson, S. R. (2020): Grant reviewer perceptions of the quality, effectiveness, and influence of panel discussion. Research Integrity and Peer Review. Online verfügbar unter https://doi.org/10.1186/s41073-020-00093-0.

Gherardi, S. (2009): Knowing and Learning in Practice-based Studies: An Introduction. In: *The Learning Organization 16(5), 352–359.* https://doi: 10.1108/09696470910974144.

Gherardi, S. (2012): Why Do Practices Change and Why Do They Persist? Models of Explanations, 217–231. In: *Hager P., Lee A., Reich A.: Practice, Learning and Change: Practice-Theory Perspectives on Professional Learning, Professional and Practice-based Learning, Dordrecht*. Springer Netherlands.

Guinea Ortiz de A., Webster, J., Staples, D. S. (2012): A meta-analysis of the consequences of virtualness on team functioning. In: *Information & Management, 49(6), 301-308*. https://doi.org/10.1016/j.im.2012.08.003.

Harrison, H., Birks, M., Franklin, R., Mills, J. (2017): Case Study Research: Foundations and Methodological Orientations. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 18(1)*. https://doi: 10.17169/fqs-18.1.2655.

Howlett, M. (2021): Looking at the 'field' through a Zoom lens: Methodological reflections on conducting online research during a global pandemic. In: *Qualitative Research, 1-16*. https:// doi: 10. 1177/1468794120985691.

Hunt, L., Nielsen, M. W., Schiebinger, L. (2022): A framework for sex, gender, and diversity analysis in research. In: *Science, New York, 377(6614), 1492–1495*. https://doi.org/10.1126/science.abp9775.

Husu, L. & 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).

Jong, L., Franssen, T., Pinfield, S. (2022): 'Excellence' in the Research Ecosystem: A Literature Review. Working Paper, RoRI Working Papers 5(5). Research on Research Institute. https://doi: 10.6084/m9.figshare.16669834.v2.

Kolev, J., Fuentes-Medel, Y., und Murray F. (2020. „Gender Differences in Scientific Communication and Their Impact on Grant Funding Decisions". AEA Papers and Proceedings 110, p 245–49. https://doi.org/10.1257/pandp.20201043

Korsvik, T.R., González, L. and Dvorackova, J. (2023) Benchmarking and assessment report on guidelines for sex/gender analysis, GENDERACTIONplus Deliverable 4.1

Langfeldt, L. (2022): The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome. In: *Forsberg E., Geschwind L., Levander S., Wermke W.: Peer review in an Era of Evaluation. Understanding the Practice of Gatekeeping in Academia: Palgrave Macmillan, 297–326.*

Langfeldt, L., Reymert, I., Svartefoss, S. M. (2023): Distrust in grant peer review—reasons and remedies. In: *Science and Public Policy*, scad051. https://doi.org/10.1093/scipol/scad051.

LERU (2019) Equality, diversity and inclusion at universities: the power of a systemic approach

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Lockwood Harris, K. (2017): Re-Situating Organizational Knowledge: Violence, Intersectionality and the Privilege of Partial Perspective. In: *Human Relations 70(3), 263–285*. https://doi: 10.1177/0018726716654745.

Kraemer-Mbula, E. (2020): Gender diversity and the transformation of research excellence. In: *Kraemer-Mbula E., Tijssen R., Wallace M. L., & McClean R.: Transforming Research Excellence: New Ideas from the Global South, 79–91.* African Minds, Capetown, South Africa.

Mayring, P. (2015): Qualitative Inhaltsanalyse: Grundlagen und Techniken. 12. Überarbeitete Auflage. Beltz, Weinheim.

Mazur, A. G. (2016): Policy Analysis: Feminist Comparative Policy. In: *Encyclopedia of Public Administration and Public Policy, Third Edition, Taylor and Francis, New York, 1-7*. https://dx.doi.org/10.1081/E-EPAP3-120053606.

Peterson, H., Husu, L. (2023): Online panel work through a gender lens. Implications of digital peer review meetings. In: *Science and Public Policy 50(3), 371–381*. https://:doi: 10.1093/scipol/scac075.

Schiebinger, L., Schraudner, M. (2011): Interdisciplinary Approaches to Achieving Gendered Innovations in Science, Medicine, and Engineering. In: *Interdisciplinary Science Reviews, 36(2).* https://ssrn.com/abstract=1895434.

Schiffbänker, H. (2023): Implementing 'Gender in Research' as Inclusive Excellence Indicator – Practices in peer review panels [preprint]. 27th International Conference on Science, Technology and Innovation Indicators (STI 2023).

Schiffbänker, H., Sauer, A., Husu, L., Peterson, H. (2023): Synthesis Report on gender bias in each core-RFO, GRANteD Deliverable D6.1.

Schiffbänker, H., Van den Besselaar, P., Holzinger, F., Mom, C., Vinkenburg, C. (2022): Gender Bias in Peer-Review-panels: "The Elephant in the Room". In: *Jenkins F., Hoenig B., Weber S. M., Wolffram, A.: Inequalities and the Paradigm of Excellence in Academia, Routledge*.

Schiffbänker, H., Haas, M., Holzinger, F. (2022): The gendered nature of independence in the context of research funding and excellence. In: *SN Social Sciences, 2/275*. https://doi.org/10.1007/s43545-022-00563-w.

Science Europe (2017): Practical Guide to Improving Gender Equality in Research Organisations. https://:doi: 10.5281/ZENODO.4980161.

Shefer, T., Zembylas, M., Bozalek, V. (2023): Re-viewing Peer Reviewing: Towards an Affirmative Scholarship. In: *Scholarship of Teaching and Learning in the South, 7(1), 147–167*. https://doi.org/10.36615/sotls.v7i1.299.

Silverman, D. (2006), Interpreting Qualitative Data: Methods for Analyzing Talk, Text, and Interaction, Sage Publications, London

Sovacool, B. K., Axsen, J., Sorrell, S (2018): Promoting novelty, rigor, and style in energy social science: Towards codes of practice for appropriate methods and research design. In: *Energy Research & Social Science, 45, 12-42*. https://doi.org/10.1016/j.erss.2018.07.007.

Söderqvist, L., Ahlqvist, V., Andersson, J., Bergström, M., Johansson, H. S., Svantesson, M., Österberg, R. (2019): A gender-equal process. A qualitative investigation of theassessment of research grant applications. In: *Swedish Research Council, Vetenskapsrådets rapporter 2020, Stockholm*. https://www.vr.se/english/analysis/reports/our-reports/2020-05-07-a-gender-equal-process.html.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Steinbauer, R., Rhew, N. and Chen S., (2015) From stories to schemas: A dual systems model of leaders' organisational sensemaking, in: Journal of Leadership and organisations Studies, 22, 404–412.

Tannenbaum, C., van Hoof, K. (2018): Effectiveness of online learning on health researcher capacity to appropriately integrate sex, gender, or both in grant proposals. In: *Biology of sex differences, 9(1), 39.* https://doi.org/10.1186/s13293-018-0197-3.

Utoft, E. H., Søndergaard, M. K., Bendtsen, A.-K. (2021): A lack of mess? Advice on undertaking video-mediated participant observations. In: *Journal of Organizational Ethnography, Emerald Group Publishing Limited, 11(3), 243-258.* https://doi 10.1108/JOE-07-2021-0037.

Vinkenburg, C. J., Ossenkop, C., Schiffbänker, H. (2021): Selling science: optimizing the research funding evaluation and decision process. In: *Equality, Diversity and Inclusion: An International Journal*. https://www.emerald.com/insight/content/doi/10.1108/EDI-01-2021-0028/full/htmlAIT.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J. (2015): The metric tide. Report of the independent review of the role of metrics in research assessment and management.

Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, *19*(3), 321-332.

Zimmermann, B. (2023): Neue Wege in der Karriereförderung. Reflexion des Umstiegs von Frauenprogrammen auf ESPRIT. In: *BWBWF 2023: Von der Geschlechterpolitik zur diversitätsorientierten Gelichstellungspolitik im österreichischen Hochschul- und Forschungsraum.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.