technopolis
group

July 2024

Accompanying process evaluation of FWF's Emerging Fields

Final report

Peter Kolarz, Anete Vingre, Diogo Machado, Laura Sutinen, Tobias Dudenbostel, Erik Arnold

www.technopolis-group.com

July 2024

Accompanying process evaluation of FWF's Emerging Fields

Final report

Peter Kolarz, Anete Vingre, Diogo Machado, Laura Sutinen, Tobias Dudenbostel, Erik Arnold

Table of Contents

Tables

Figures

Headline summary

Launched in 2022 as part of the Excellent=Austria initiative, the first call of the Emerging Fields (EF) Programme attracted 45 applications from a range of multidisciplinary Austrian research teams. Five of these were granted a total of €31m via a three-stage selection process. This report presents the findings of the evaluation of that selection process.

Our headline conclusion is that the Emerging Fields (EF) programme has served a critical signalling and galvanising function. By launching the programme, the FWF provided an opportunity for Austria-based researchers to propose a wide range of highly novel research ideas. Many of these were fully formulated only in response to the EF call.

In the wider context of the Excellent=Austria initiative, the EF call has thereby fulfilled an important function by adding funding for basic research and by providing a vehicle for innovative ideas to take shape.

The three-stage EF assessment process itself was well designed and works well. In line with the original design-intentions, Stage 1 had a strong focus on innovative aspects of the research ideas. By focussing on short synopses, this stage also ensured that the burden of extensive peer reviewing of full applications was minimised (though applicants still had the burden of writing the full applications).

Stage 2 was a full formal external expert peer review focused primarily on scientific quality, while the Stage 3 Jury hearings allowed a broad range of questions to be asked in relation to many criteria but also provided insight into team composition, dynamics and hierarchy. The designated decision-making body, the FWF Scientific Board, played only minimal roles at Stages 1 and 3 but a major role at Stage 2.

We find a small number of challenges. First, applicants were dissatisfied with the feedback provided and what they saw as lack of process transparency, especially at Stage 1.

Second, while the initial pool of applications and applicants was diverse, outcomes in terms of gender and discipline are less so, with no humanities or arts applications being successful and no female coordinators among the five winning awards. While we find no overt evidence of bias or discrimination, the FWF needs to continue to monitor these figures in future calls.

Finally, there was much need throughout the process to define several key terms and how to treat them (novel, interdisciplinary, field, what it means to be 'emergent', etc). This extended to applicants and institutions being able to contextualise the scheme and what exactly was expected, extending to the difference or similarity to the FWF's SFBs.

In the final section of this report, we set out our full list of recommendations. These include creating synergy between the EF scheme and the SFBs, as well as better definition and clarification of key terms ('novel', 'interdisciplinary', etc) and what role they play in the assessment. We make various additional recommendations for minor process adjustments.

Executive summary

The Austrian Science Fund's Emerging Fields (EF) programme seeks to fund particularly original, innovative, paradigm-shifting research. Launched in 2022 as part of the Excellent=Austria initiative, the first call of the Emerging Fields (EF) Programme attracted 45 applications from a range of multidisciplinary Austrian research teams.

The selection process involves the following steps:

- First, review of 3-page 'synopses', i.e. short outlines of the project idea, by an international Jury of 16 experts
- Second, peer review by at least 3 external reviewers
- Third, a presentation by shortlisted applicants to the international Jury

This report presents the findings of the evaluation of the selection process of the first call of the EF programme. This evaluation was tasked critically to review the process end-to-end, to identify strengths and weaknesses in its design, and to provide evidence-based recommendations to the FWF and its supervisory bodies on how to improve these processes for the next EF call.

We used a mixed methods approach for this evaluation, consisting of the following main components:

- Desk review of the programme and other documentation
- Composition analysis of applications, applicants and reviewers
- Bibliometric and text analysis of applications to construct novelty indicators and assess the originality of applications
- Text analysis of application and synopsis reviews
- Observation of the hearings and Jury meetings
- An online survey of all lead applicants
- An online survey of all external reviewers
- Semi-structured interviews with Jury members, representatives of applicant institutions, policymakers and FWF staff

Overall findings

Our research finds ample evidence to indicate very high levels of scientific novelty across the pool of applications submitted to the Emerging Fields call. This conclusion is supported both by multiple sources:

- Our indicators from our analysis of synopses and applications show that the EF programme attracted highly novel proposals compared to the 'state of the art' more generally, as well as compared to FWF's core funding scheme (Principal Investigator Projects – Einzelprojekte) in particular
- Likewise, Jury members we interviewed frequently commented on the high level of novelty of the synopses they reviewed (and subsequently on the presentations at the Stage 3 Jury hearings). Of our Stage 2 peer reviewers surveyed, around 60% judge the EF application they reviewed to have a higher level of novelty than what they would consider the norm, splitting roughly equally between the 'slightly' and 'significantly' answer options

Further, the EF scheme did not simply act as a signalling-point for already existent novel and potentially transformative research ideas, but may have been a catalyst for such ideas to be formulated in the first place: of the applicants who participated in our survey, 46% claimed that

their project idea was entirely new and motivated by the EF scheme. A further 35% said that the idea had existed prior to the EF call but no funding had been applied for. Just 8% report that their idea existed before the call and that it had received prior funding.

We note the following headline figures:

- The total scheme success rate is 11%. Though not entirely untypical of competitions for PI-initiated project funding, we note that this rate is quite low even for such large awards. The success rate is staggered roughly equally across the three assessment stages. At each stage, roughly half the applications are sifted out

- On average, applications had 5.2 team members and requested €5.4m. Eligible applications can range from 3 to 7 team members and from €3-6m. Applications therefore tend towards the upper limits on both these dimensions, but not to an extreme extent

- The social sciences and humanities were most strongly represented among the initial 45 applications, while there are fewer applications from biomedical sciences with technical sciences occupying an intermediary position. These figures remain stable through the first assessment stage. However, applications from the natural and technical sciences are especially unsuccessful in Stage 2 while social sciences and humanities are especially unsuccessful in Stage 3

- Submitted applications had on average 2.4 female and 2.8 male team members and 47% of applications had a female lead coordinator. While this gender balance favours men in absolute terms, wider inequalities in the overall science system are more pronounced than the gender balance among EF applicants. The gender balance (both for coordinators and overall number of female team members) after Stage 1 is similar to that of the applications. However, the average number of female team members per proposal reduces after Stage 2 and there are no female coordinators among the five winning applications

These are small numbers and we have not found evidence of direct bias or discrimination in relation to gender or discipline. However, it will be important to continue monitoring these figures in future calls.

Applicants' satisfaction rates are highest for the FWF's support during the application process and with the user-friendliness of the application portal. Areas where applicants' satisfaction levels are comparatively low are the effort needed to prepare applications, the overall efficiency of the process, and the clarity and completeness of feedback received.

Satisfaction levels are lowest among applicants who failed at Stage 1. A key issue here is that applicants have to prepare their synopsis and full application for initial submission but if their application fails at the first assessment stage, only the synopsis will be read. The FWF could request only the 3-page synopses to be submitted at Stage 1, with only successful candidates then invited to submit their full application. However, this would lower barriers to entry, leading most likely to a much larger influx of sub-par synopses, requiring substantial additional reviewing work. Given the already low success rate, coupled with the overall pool of highly innovative ideas, we judge that the current approach is the right one.

The overall assessment process was lengthy. It took around 18 months from the initial call to the final funding decision (and over 12 months for the assessment process itself), which is long compared both to Austrian and international norms for PI-initiated research programmes. This was due to ongoing budget negotiations at the time, so the dates for approvals could only be finalised after submissions. The delays mostly affected Stage 2, which should have been about half as long, which would have resulted in a time-to-grant of around nine months. This shorter timeline is planned for the next EF call.

On clarity and completeness of feedback, we note that applicants generally indicate that they have low confidence in the thoroughness and expertise of those who reviewed their

applications. Satisfaction and confidence levels are lowest for applicants who were rejected at Stage 1 (see below). We do not find there to be an issue with the Jury members or reviewers themselves. Instead, we identify a need for greater transparency about reviewer selection and potentially for more structured and detailed feedback especially at the first assessment stage, where confidence and satisfaction levels are at their lowest.

Stage-by-stage

The intention behind the EF process design was as follows:

- Stage 1 (Jury review of 3-page synopses) was intended primarily to assess and reward the novelty of the proposed research ideas
- Stage 2 (external peer review of full applications) was intended primarily to assess scientific quality in the conventional sense (feasibility, robustness, etc)
- Stage 3 (Jury hearings) was intended primarily to assess team composition

Stage 1 broadly functioned well, especially in terms of its prioritisation of novelty as a focus of attention. However, there is a need to define key terms and explain several details of the process better. Several of the recommendations we set out at the end of this report therefore relate to Stage 1.

To further discuss the approach to assessing the innovative aspects of the applications, the Jury chair organised two informal Jury meetings. This was not an intended formal part of the process but happened at the Jury chair's request. Our interviews indicate that this was an important and helpful additional step to discuss how to conduct the reviewing work.

Our evaluation found no evidence of insufficient expertise in the Jury to review each synopsis. However, the uneven distribution of workload highlights that there is at least a slight mismatch between the Jury composition and the overall thematic/disciplinary mix of the EF synopses.

Following Jury review, all high-scoring applications were passed to Stage 2 and all low-scoring ones were rejected. The FWF Scientific Board was substantively involved only in those applications that received substantially split scores. Of the eight applications discussed, three were passed to Stage 2 and five were rejected. The novelty of research ideas appears to have been the main criterion determining progression from Stage 1.

The picture of actual Stage 1 outcomes is a little more nuanced in relation to applications' characteristics. Our novelty indicators do not show evidence of bias against novel ideas in the Stage 1 review process, though they also do not suggest any drastic novelty appetite within the process either. In other words: when we compare the novelty-levels of successful and unsuccessful synopses at Stage 1, there are no significant differences evident from our analysis.

However, our analysis of reviews strongly suggests that novelty played an important role in Stage 1 verdicts: written reviews of synopses focus much more on the criterion of novelty than on scientific quality, with risk/feasibility and team characteristics occupying intermediate positions in terms of the space dedicated to these criteria in the reviews. Moreover, successful synopses are associated with a slightly greater focus in reviews on novelty while unsuccessful synopses are associated with a greater share of reviews dedicated to quality issues.

From the applicants' perspective, we find no major issues with the operation of Stage 1, though one major issue from the point of view of applicants lies with feedback. In a similar vein, applicants rejected at Stage 1 are far more likely to have low confidence in the thoroughness of their application assessment. As mentioned, we find no evidence of bias or of 'poor fit' between application topics and Jury member expertise. But we judge there to be issues around transparency, which may have affected applicants' perceptions: to the applicants, it would have been unclear how the Jury was composed, how applications were allocated to Jury members, how the different criteria were weighed or translated into the eventual rankings,

or indeed how these rankings translated into final pass/fail verdicts. This may substantially increase sensitivity to any perceived bias or unclear process elements.

The 24 applications passed from Stage 1 to Stage 2 were each sent to three international academic experts for external peer review. The experts provided full written reviews, including scoring on a standard 5-point scale. External reviewers were identified and selected by the FWF Scientific Board with assistance from the heads of the FWF scientific departments. Following completion of all external reviews, the Board met to take decisions.

While novelty was a key focus of the assessment and decision-making in Stage 1, Stage 2 focussed strongly on the scientific quality of applications in the conventional sense. This applies both to the external reviews themselves as well as to the Scientific Board discussion. The role of the Scientific Board was much expanded compared with Stage 1, as there were no 'automatic' passes or rejections of applications based solely on external reviewers' scores. Every application was discussed, often in some detail, regardless of the overall level or divergence among external reviewers' scores.

We can report an almost entirely positive picture in terms of external reviewers' own views and perceptions on Stage 2. Aggregate process satisfaction scores in our survey of reviewers are consistently high on all aspects we surveyed.

Stage 3 was the shortest part of the assessment process, spanning less than two months from the communication of Stage 2 decisions and culminating in the Jury hearings in Vienna. Ten applicant teams were invited, of which five were funded.

Across the hearings and especially in subsequent discussion, there was much need for debate around the core terms: there was considerable discussion around what exactly novelty means, and it was evident that it means different things in different contexts, including in different disciplines. There was also discussion about what it meant to be 'emergent', or at what point something can be described as a 'field'. The importance of interdisciplinarity was also a point of discussion, as it was unclear whether interdisciplinarity itself should be rewarded. These were all important questions that informed decision-making. The hearings and subsequent discussion acted as a forum to discuss them. Elements of the discussion could likely have been expedited if some of these concepts had been more clearly defined from the beginning.

Overall, interviewees felt that the hearings were well organised and excellently supported by FWF staff. Jury interviewees noted that having this third assessment stage in this form was very important due to the ability to probe into any unclear methodological aspects. Interviewees also consistently identified meeting the project teams as the most important aspect. This included both the ability to get a clearer sense of who was doing what, but also to observe the team dynamics, collegiality and hierarchy.

Wider significance

The EF programme was designed as part of the Excellent=Austria initiative, which sought to provide additional funding for basic research in Austria. Parts of the initiative have an emphasis on attracting and retaining the best international talent, while other parts of the initiative are focused more on fostering ideas among researchers already based in Austria. EF firmly addresses the latter of these aims, while also adding the elements of fostering highly novel, high-risk/high-reward research ideas. The programme was additionally explicitly intended to be open to interdisciplinary, transdisciplinary and arts-based research proposals.

The true added value of the EF scheme lies in its provision of funding for exceptionally novel ideas. Our interviews and document review suggest that there are certainly other large research awards available to researchers in Austria. However, these are focused much more on 'ordinary' basic research. The programme unleashed creative potential in the Austrian research base that otherwise may have remained at least partially undiscovered.

Our interviews of all stakeholder groups consistently highlight that a programme with these characteristics was indeed needed, and the observed demand and success rate further confirm this. In light of the highly novel profile of applications and the substantial share of interdisciplinary, transdisciplinary and arts-based applications, we also conclude that the programme managed to attract the breadth of ideas that was intended.

Only 11% of surveyed applicant coordinators believe that their EF project idea would end completely without the EF funding. However, almost three quarters of respondents believe that their project would either progress with much reduced scope (due to likely lack of funding) or that they would modify their project by removing some of the novel, high-risk/high-reward aspects to increase their chances of securing alternative funding.

The FWF programmes most commonly mentioned by stakeholders when comparing to EF were the 1000ideas programme and the Special Research Programmes ('Spezialforschungs-bereiche', SFB). The former was frequently noted as being 'related' to EF as it is the only other FWF programme with an overt focus on highly innovative high-risk/high-reward research. However, there appears to be no need to consider any synergies or coordination between 1000ideas and EF. The potential for synergies with the SFBs (which do not share the high novelty aspect but are similar to EF in most other ways) was brought up far more frequently. Many stakeholders also noted that institutions and academics struggled at times fully to understand the difference between SFBs and EF and suggested and/or supported the idea of a fusion of SFBs and EF (with a combined budget).

Recommendations

Recommendations are set out in full at the end of this report. In brief:

- Without reducing budget, there is a case for combining the EF programme with the SFBs. SFBs could then alternate between 'regular' and 'Emerging Fields' rounds, with EF round maintaining the current assessment process and signalling demand for unusual, high-risk/high-reward research ideas

- The FWF should include in the documentation of future calls (both for applicants and assessors) some definitions and illustrative examples of key terms and whether/how they are assessed, including 'novelty, 'interdisciplinarity', 'emergent', 'field' and the role of arts-based and transdisciplinary approaches

- The FWF should explore options at least partially to finalise Jury composition after applications have been received so that the Jury can be better tailored to the portfolio

- The FWF should state more clearly the details around Jury recruitment, allocation of applications to Jury members, and the level of feedback provided at Stage 1

- The Jury meeting at the start of Stage 1 should become a formal part of the EF process

- The FWF should mandate a slightly more defined structure for synopses

- The FWF should continue monitoring diversity markers in future calls, and consider measures to counter bias if outcomes continue as they have in the first call

1 Introduction

This report presents the findings of the evaluation of the selection process of the Austrian Science Fund's (FWF's) first call for its Emerging Fields (EF) programme. This study was commissioned by the FWF and carried out by Technopolis between February 2023 and March 2024. This evaluation was tasked with the following objectives:

- To critically review
  - the embedding of EF into the programme portfolio of the Excellent=Austria initiative and other relevant cooperative funding programmes
  - the EF application guidelines
  - the EF application process
  - the EF decision-making process from the perspective of researchers, the universities / research institutes, the reviewers and panel / Jury members, and the FWF
- To identify strengths and weaknesses of the process design
- To provide evidence-based recommendations to the FWF and its supervisory bodies on how to improve these processes for the next EF call

1.1 Emerging Fields – context and process overview

The EF programme is part of a wider national and international basic research funding portfolio. It is one of the Excellent=Austria initiative's three funding streams. The initiative started with the launch of the Clusters of Excellence programme, supporting large and long-term research collaborations with flexibility in how funding is used. EF is the second step of the initiative and focuses on particularly innovative research projects. The initiative is also intended to support the recruitment and retention of international researchers with the third programme, 'Austria Chairs of Excellence'.

Figure 1   Emerging Fields programme in the wider basic research funding portfolio
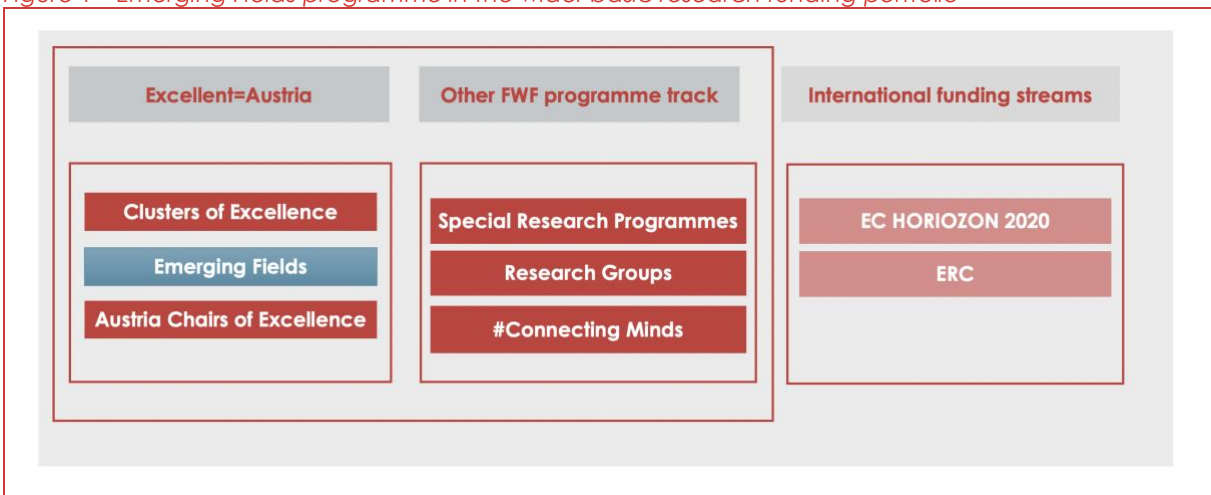
EF seeks to fund particularly original, innovative, paradigm-shifting research. Various terms are used to label this type of research (transformative, disruptive, innovative, original, novel,

frontier, high-risk, etc.). 'High-risk/high-reward' (HRHR) research is an increasingly widely used term internationally and was the subject of a recent OECD policy paper on the topic.[1]

The typical assessment processes in research funding (administrative compliance check, peer review, panel review and sign-off by the funder) have several issues that are especially problematic when the funding programme has a specific objective of supporting HRHR and/or interdisciplinary research. Most notably, peer review tends to be conservative, so proposals based on established knowledge and methods tend to do better than those that include pathbreaking new methods often associated with risk but potentially also higher rewards.[2]

Funders seeking to foster HRHR research therefore often undertake modifications to their standard assessment procedures. Examples include use of partial randomisation (e.g. by the Volkswagen Foundation), complete bypass of peer review (e.g. by the US NSF), presentation events (e.g. by the UK's EPSRC) or double-blind reviewing (e.g. by the UK's ESRC).[3] These innovative assessment methods tend to come with a degree of controversy, and so they are typically associated with relatively small awards sizes. An alternative approach (practised notably at the ERC) is to provide large awards to exceptionally promising scientists, giving them enough resources to spend part of their efforts on HRHR ideas.[4]

For the EF programme, the FWF has combined various elements of both these schools of thought. It involves a three-stage assessment process:

- First, review of short 'synopses', i.e. short outlines of the project idea
- Second, peer review by at least three external reviewers
- Third, a presentation by shortlisted applicants

For Stages 1 and 3, the FWF assembled an international 'Jury' of 16 senior academics (plus one Jury chair). The Jury reviews the synopses in Stage 1, with two Jury members reviewing each application and providing a score from A to C. The Jury also forms the 'audience' for presentations at Stage 3.

The Jury is not involved in Stage 2, where international reviewers provide written assessments of full applications. Notably, the FWF's Scientific Board ('Kuratorium') formally leads the decision-making in all three stages. Following review by the Jury in Stage 1, the Scientific Board discusses the results and determines which applications proceed to Stage 2. Likewise, the Scientific Board decides which applications to advance from Stage 2 to Stage 3 based on the external reviews. In Stage 3, the Jury makes a recommendation on which applications to fund, but the Scientific Board formally makes the final decision.

Additionally, FWF administrative staff are involved in the decision-making process at the start by performing eligibility checks (if an application fails eligibility criteria, there are routes for revision and resubmission, as well as outright rejection). After Stage 1, the heads of the FWF's specialised departments also enter the process, as they lead the identification of suitable external reviewers.

---

[1] OECD (2021), "Effective policies to foster high-risk/high-reward research", *OECD Science, Technology and Industry Policy Papers*, No. 112, OECD Publishing, Paris, https://doi.org/10.1787/06913b3b-en.

[2] Ibid; see also Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. Research evaluation, 15(1), 31-41.

[3] For a more extensive selection of examples, see a review of peer review interventions recently published by UKRI: https://www.ukri.org/publications/review-of-peer-review/

[4] Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. Research Evaluation, 21(1), 48-60; Laudel, G. & Gläser, J. (2014) 'Beyond breakthrough research: Epistemic properties of research and their consequences for research funding.' Research Policy, 43(7): 1204-16.

We provide below a full process map of the EF application assessment process, from application submission to final funding decision. This process map acts as a point of reference throughout our evaluation. To further discuss the approach to assessing the innovative aspects of the applications, the Jury chair organised two[5] informal Jury meetings (we stress that this was not an intended 'formal' part of the process but happened at the Jury chair's request). We include these in the process map.

Figure 2 Emerging fields application and selection process map
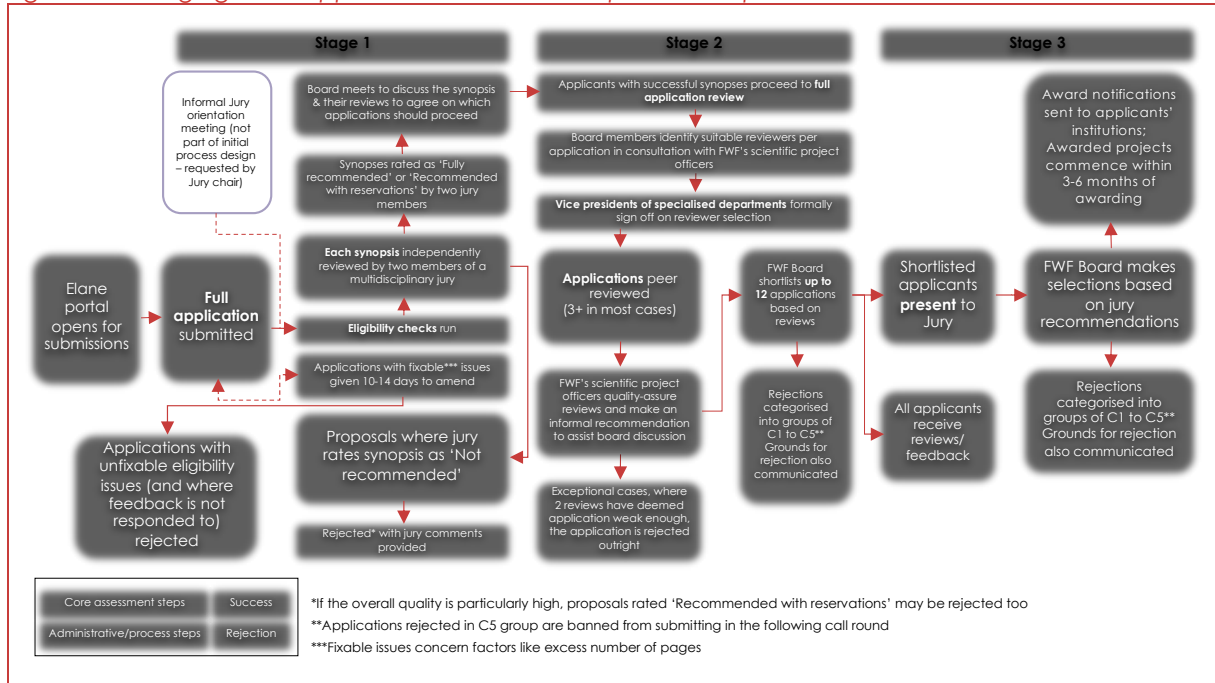
1.2   Method note

We used a mixed methods approach for this evaluation, consisting of the following main components:

- **Desk review of the programme and other documentation:** we received all documentation pertinent to the Emerging Fields scheme from FWF and reviewed these materials. This chiefly provided important context and background information and enabled us to construct a process map

- **Composition analysis of applications, applicants and reviewers:** programme data supplied by FWF allowed us to profile and track numbers and characteristics of applicants and reviewers throughout the scheme's assessment stages

- **Bibliometric and text analysis of applications to construct novelty indicators and assess originality of applications:** this is a novel approach involving analysis of synopsis and application texts and, based on this, constructing various indicators that indicate novelty in a number of different ways

---

[5] Two meetings were held because all Jury members could not join at the same time because of time zone differences.

- **Text analysis of application and synopsis reviews:** this step allowed us to assess the types of criteria that were especially dominant in the reviewing of applications and synopses

- **Observation of the hearings and Jury meetings**: we observed the FWF Scientific Board meeting for each stage of the EF scheme (i.e. where the final decisions on each assessment stage were taken). We also attended the Stage 3 Jury hearings in their entirety, including the Jury discussions after the hearings

- **An online survey of all lead applicants** ('coordinators') of the Emerging Fields call. The survey was conducted after Stage 2 had completed but before the Stage 3 hearings. We had a response rate of 82%

- **An online survey of all individuals who acted as external reviewers** of an Emerging Fields application at Stage 2 of the assessment process. We had a response rate of 43%

- **Semi-structured interviews with Jury members, representatives of applicant institutions, policymakers and FWF staff:** these were variously carried out throughout the study period

All relevant methodological details of the above components are appended to this report, including survey response rate details and raw results, content coding from the observation exercises and details of interviews. Any raw data that may allow identification of individual participants' views are not included as a matter of sound research ethics and data protection. We refer especially to our full descriptions of the textual analysis of applications and reviews (Appendix A), as these method components contain especially novel approaches to evaluative studies of this kind.

1.3   This report

In the first main section of this report, we provide an overall process assessment, focussing on overall facts and figures and overall process satisfaction levels. The second section then deals with each of the three stages of the EF funding process in turn. In the final main section, we discuss the wider strategic significance, placement and added value of the programme. The concluding section addresses directly the full range of evaluation criteria set out in the terms of reference for this study, before ending with a set of recommendations to be considered for future EF calls.

We are happy to answer any questions from readers regarding any substantive or methodological elements pertaining to this report and its underlying research.

2   Headline findings

Launched in 2022 as part of the Excellent=Austria initiative, the first call of the Emerging Fields (EF) Programme attracted 45 applications from a range of multidisciplinary Austrian research teams, of which five were funded via the three-stage selection process. The total funding awarded was €31m.[6]

We begin by presenting some broader findings that provide important context and overall judgement. This includes overall findings on process satisfaction as well as headline figures around diversity and success rates. In the later stages of this report, we then present our findings on the operation and role of the three assessment stages.

2.1   Novelty of Emerging Fields applications – an overall assessment

Scientific 'novelty' is a contested term. It can mean many things both within and between scientific fields. There are criticisms to be levelled at any attempts at creating an absolute and comprehensive definition of 'novelty'. At the same time, leaving judgements of novelty purely at the discretion of peer reviewers with no external definition may quickly become relativistic and render the term meaningless. It is not within the scope of this evaluation to solve these dilemmas.

Mindful of such issues, our research finds ample evidence to indicate **very high levels of scientific novelty across the pool of applications submitted to the Emerging Fields call**. We are confident in this conclusion because it is supported both by multiple stakeholder groups' views, as well as by multiple established and experimental indicators of scientific novelty.

In the first instance, our indicators from our analysis of synopses and applications show that the EF programme attracted highly novel proposals compared to the 'state of the art' more generally, as well as compared to FWF's core funding scheme (Principal Investigator Projects – Einzelprojekte) in particular. By 'state of the art', we refer to the highest general advancement of a scientific field achieved at the time of each proposal. This conclusion holds for all three indicators used (unprecedented combinations of journal references, unprecedented combinations of concepts, and semantic dissimilarity). The underlying methods and data for these findings (including how 'state of the art' is defined) is described in detail in Appendix A.1.

These findings are also reflected by Jury members and reviewers. The Jury members we interviewed often commented on the high level of novelty of the synopses they reviewed (and similarly on the presentations at the Stage 3 Jury hearings).

We also have an assessment of novelty from the external reviewers who conducted peer reviews of full applications in Stage 2 of the assessment process. We asked reviewers to compare the novelty of EF applications to that of other applications to basic research funding schemes they had worked on previously. We note that these survey findings only cover those applications that passed to Stage 2. However, the results are very pronounced: around 60% of respondents judge the EF application they reviewed to have a higher level of novelty than what they would consider the norm, splitting roughly equally between the 'slightly' and 'significantly' answer options.

Figure 3   Stage 2 reviewers' assessment of applications' novelty

How original, innovative and/or paradigm-shifting were research ideas presented in the EF application you reviewed compared with the research ideas of other research grant applications you have reviewed outside the EF scheme?
(n=29)

- Significantly more original, innovative and/or paradigm-shifting: 31%
- Slightly more original, innovative and/or paradigm-shifting: 28%
- About as original, innovative and/or paradigm-shifting: 34%
- Slightly less original, innovative and/or paradigm-shifting: 7%
- Significantly less original, innovative and/or paradigm-shifting: 0%

Source: Survey of EF Stage 2 external reviewers. Survey item included a 'Cannot say/Don't know/Not applicable' option, which was not selected by any respondents. NB: original survey item stated 'Emerging Fields', shortened to 'EF' here for formatting reasons. Answer options above are abridged for presentation. Full answer options in the survey item follow the format, "The research ideas in the Emerging Fields application were significantly more original, innovative and/or paradigm-shifting than the research ideas of other research grant applications I have reviewed in the past". Survey item included the following instruction: "Please feel free to consider for comparison any applications for any basic research grants you may have conducted at any point, for any funder, public or private, in Austria or internationally".

Finally, there is a high level of confidence among applicants themselves that their project presented novel and/or unconventional ideas, perspectives and methods. Likewise, applicants strongly claim their proposed project to present new interdisciplinary ideas. In the survey item below, we also checked for various other possible definitions or characteristics of highly novel research, including whether it challenged established assumptions in their field, whether it carries a strong risk of failure or whether it would have struggled in funding programmes for conventional research. Large majorities of applicants indicate that their project ideas reflected these definitions at least to some extent.

Figure 4   Applicants' self-assessment of their project's novelty

To what extent do you agree with each of the following statements about your proposed project?
(n = 37)

- My project presented novel and/or unconventional ideas, perspectives and methods: 97%
- My project presented new interdisciplinary perspectives: 84%, 16%
- My project challenged core assumptions of the research establishment in my field: 57%, 27%, 11%, 5%
- My project presented a risky idea with significant chance of failure: 38%, 27%, 22%, 11%
- My project would have struggled to be accepted by funding programmes for conventional research: 31%, 44%, 14%, 6%
- My project presented new arts-based perspectives: 19%, 8%, 8%, 53%, 8%
- My project presented new transdisciplinary approaches (involving non-academic participants): 17%, 33%, 19%, 22%, 8%

■ Strongly agree ■ Somewhat agree ■ Neither agree nor disagree ■ Somewhat disagree ■ Strongly disagree ■ Don't know / cannot say

Source: Survey of EF applicants

In addition to the evidence demonstrating the overall high novelty of applications submitted to the EF scheme, we also note that sizeable minorities of applicants note that their project ideas included arts-based and/or transdisciplinary elements. The EF programme documentation notes that such unconventional types of research ideas were sought by the programme, and to an extent this has happened.

Overall, these findings say little about the operation of the EF programme itself (on which more after this section). However, they tell us a lot about the signalling power of the EF scheme. In general, researchers in Austria seem to have understood the EF scheme as a call for explicitly unconventional/novel ideas and they were able to supply these.[7]

Further, the EF scheme did not simply act as a signalling-point for already existent novel and potentially transformative research ideas, but may have been a catalyst for such ideas to be formulated in the first place. Of the applicants who participated in our survey, 46% noted that their project idea was entirely new and motivated by the EF scheme. A further 35% noted that the idea had existed prior to the EF call but that no funding had previously been sought for it. Just 8% report that their idea existed before the call and that it had received prior funding.

Figure 5   The origin of applicants' research ideas
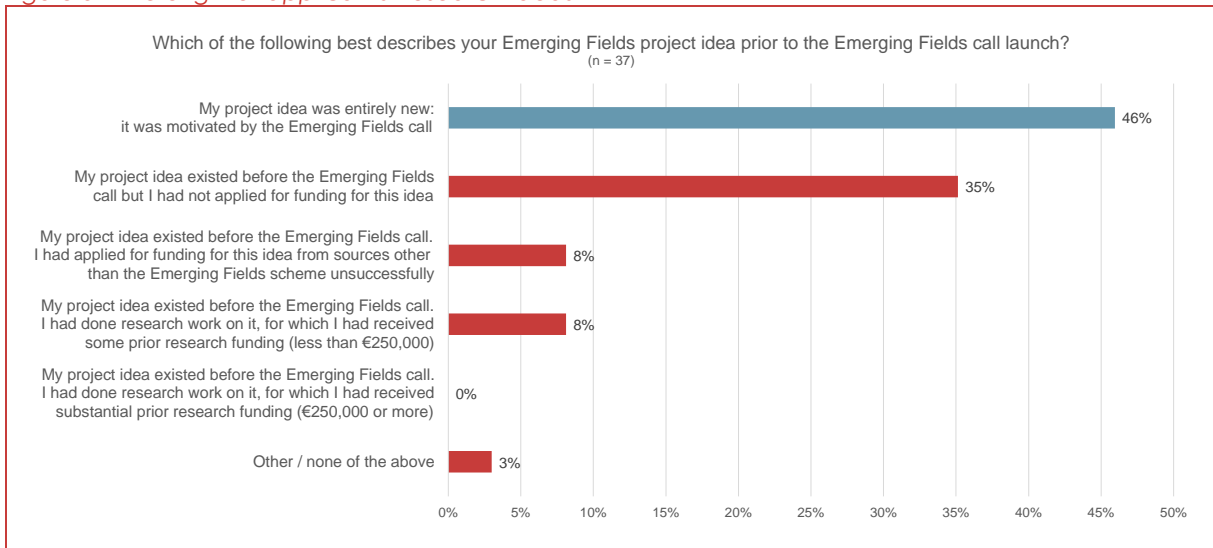
Source: Survey of EF applicants. NB: one answer option above is abridged for presentation. Full answer option in the survey item was: 'My project idea existed before the Emerging Fields call. I had applied for funding for this idea from sources other than the Emerging Fields scheme unsuccessfully'.

In short, independently of how well the EF selection process might operate, we find that creating the EF scheme itself helped not only to attract, but likely also to create, a wide pool of highly novel ideas. This speaks in favour of Austrian researchers' ability to produce such ideas and also for the FWF's signalling-ability in these matters.

---

[7] Our interviews with institutional representatives indicate that there was a lack of clarity on what exactly is meant by 'novelty' in relation to the EF scheme. We return to the issue of unclear definitions later in this report. However, at the general level this does not appear to have been a problem. Even if individual EF applications are novel/ pathbreaking/ transformative in different ways, the result is still an overall pool of applications that appear genuinely different from 'business-as-usual research funding.

2.2 Headline figures

The EF scheme is quite small in terms of the overall number of applications, though the amount of funds awarded is large. While funding schemes with hundreds of awards lend themselves to extensive programme data analysis, it is hard to produce statistically significant results with the small number of applications in the EF scheme. The EF headline data do allow for some important observations and even point to a few potential issues. However, what looks like a trend may at such low numbers simply be coincidence. It will be important to keep looking at these figures over subsequent EF calls to see whether trends remain the same. With this in mind, we make the following observations:

- Of the 45 applications submitted, five were funded, equalling a total success rate of 11%. Though not entirely untypical, we note that this rate is quite low even for such large awards[8]

- The success rate is staggered roughly equally across the three assessment stages: at each stage, roughly half of remaining applications are sifted out, meaning that in terms of numbers alone, each stage has roughly equal significance in the overall decision making

- Applications on average had 5.2 team members and requested on average €5.4m. Eligible applications can range from 3 to 7 team members and from €3-6m. Applications therefore tend towards the permitted upper limits on both these dimensions, but not to an extreme extent, indicating that these award parameters have been set to about the right level by FWF. We also find that the average team and budget size remain fairly constant throughout the three assessment stages, so there is no immediate evidence of bias along these lines

- Of the three disciplinary areas associated with the three main FWF departments, the social sciences and humanities were most strongly represented among the initial 45 applications, with biomedical sciences slightly underrepresented and the natural and technical sciences occupying an intermediary position. These figures remain stable through the first assessment stage. However, applications from the natural and technical sciences are especially unsuccessful in Stage 2 while social sciences and humanities are especially unsuccessful in Stage 3. As noted, the overall numbers are too small to make robust statistical claims and our remaining research finds no direct evidence of bias against particular fields. However, this is a trend worth monitoring in future EF calls and we note later in this report where additional care may be given to differences between these three main areas of science

- Applications submitted had on average 2.4 female and 2.8 male team members and 47% of applications had a female lead coordinator. While this gender balance favours men in absolute terms, these proportions are slight, and likely reflect the gender inequalities that already exist in the wider science system. In fact, these wider inequalities area more pronounced than the gender balance among EF applicants[9]

- The gender balance figures (both for coordinators and overall number of female team members) remain stable through Stage 1 of the assessment process. However, the average number of female team members reduces after Stage 2 and there are no female coordinators among the five winning applications

---

[8] In terms of scale, interviewees frequently compared EF to ERC grants, which have a success rate of around 14-15%. The Human Frontier Science Program (HFSP), which somewhat reflects the EF ambition of funding especially novel research ideas, has in recent years had success rates of between 8% and 15%. Our recent review of the HFSP highlighted that such low figures may begin to discourage application. A comparative review of similar programmes found an average success rate of around 20%: https://www.hfsp.org/node/74873#book/

[9] See e.g. the EC She Figures: https://projects.research-and-innovation.ec.europa.eu/en/knowledge-publications-tools-and-data/interactive-reports/she-figures-2021

Table 1   *Headline application characteristics and success rates*

	All applications*		Applications passed to Stage 2		Applications passed to Stage 3		Awardees	
Total number	45		24		10		5	
Proportion/success rate	100%		53%		22%		11%	
Main associated FWF department								
HumSoc	20	44%	10	42%	5	50%	1	20%
NatTec	14	31%	7	29%	1	10%	1	20%
BioMed	11	24%	7	29%	4	40%	3	60%
Gender balance								
No. female consortium members	2.4		2.5		1.8		2.0	
No. male consortium members	2.8		3.0		2.9		3.4	
Female coordinator	21	47%	9	38%	4	40%	0	0%
Budget requested**	€5,404,249		€5,850,867		€5,607,403		€5,771,841	
Average team size	5.2		5.6		4.8		5.4	

*Includes three applications rejected without review **Including personnel costs adjustment. We also note that inflation over the course of the roughly one-year assessment period means that the final amounts increased, leading to the total sum of €31m. However, we present here the originally requested amounts to ensure stable comparison through the assessment stages.

Regarding gender, we stress again that these are small numbers and we have not found evidence of direct gender discrimination. However, the especially stark gender figures in the final outcomes indicate a strong need to ensure that future EF calls are as gender inclusive as possible and to continue monitoring gender representation. Given the outcomes specifically at Stage 3, there may even be a case to provide unconscious bias training for the Stage 3 Jury hearings.[10]

2.3   Overall process satisfaction

There is a mixed picture in terms of overall process satisfaction. While the next main section delves into the specific issues encountered at various points of the assessment process, the aggregate view from our survey of applicants highlights some important themes that our evaluation work has encountered.

The applicants are most satisfied with the FWF's support during the application process and with the user-friendliness of the application portal. These are important features of the FWF's overall 'service provision' and speak in favour of the FWF as an organisation, as these support and IT functions likely also operate in similar ways across other parts of the FWF's funding portfolio. However, there are three areas where applicants' satisfaction levels are

---

[10] We note that we did not pick up on any gender discrimination or evidence of bias in our observation of the Stage 3 Jury hearings and we also note there is inconclusive evidence on the effectiveness of unconscious bias training. We therefore do not posit this as a formal recommendation at the end of this report but given these outcomes, we note the need for continued attention to this issue.

comparatively low: The effort needed to prepare applications, the overall efficiency of the process, and the clarity and completeness of feedback received.

Figure 6   Applicants' views of application process
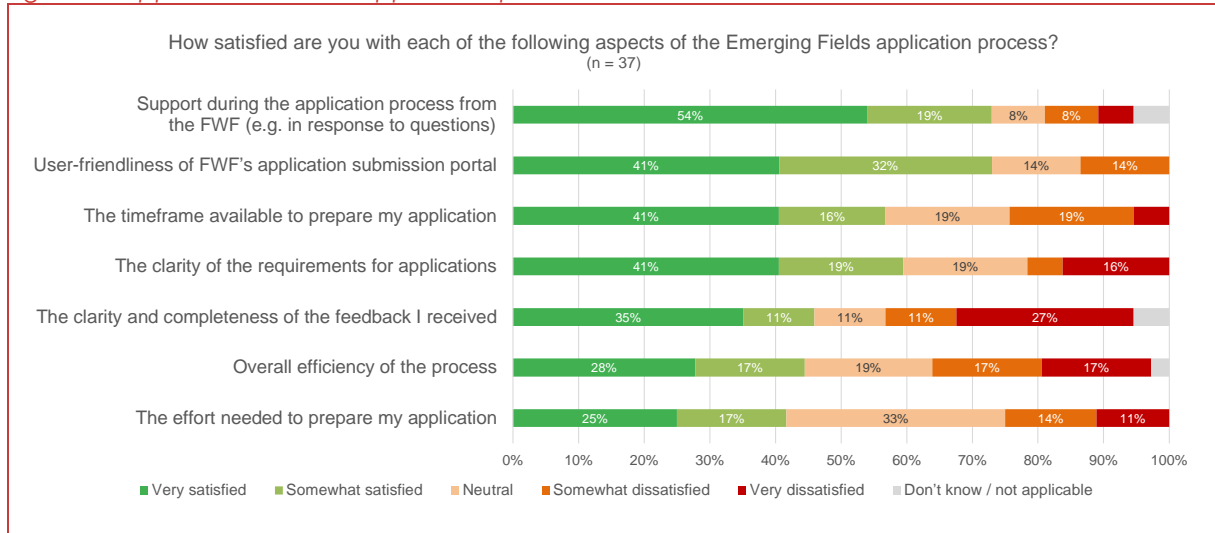
Source: Survey of EF applicants. NB: one question is abridged for presentation in this survey item. The full question read: 'The clarity of the requirements for applications (call documents, criteria, etc.)'

It is worth reflecting on the three items with the lowest satisfaction ratings in turn.

Regarding the effort needed to prepare EF applications, a more granular analysis highlights satisfaction levels to be especially low among applicants who failed at Stage 1. A key issue here is that applicants have to prepare their synopsis and full application for initial submission but if their application fails at the first assessment stage, only the synopsis will be read. In other words: for around half of the initial applicants, no Jury member or reviewer will ever see most of the pages they submitted to the EF scheme.

There is a genuine dilemma here: alternatively, the FWF could request only the 3-page synopses to be submitted at Stage 1, with only successful candidates then invited to submit their full application. The drawback of this is that it significantly lowers barriers to entry, leading most likely to a much larger influx of synopses. These in turn may include many sub-par and poorly conceived research ideas and will require substantial additional reviewing work.

There is unfortunately no 'middle-way' here: either applicants have to submit full applications from the start, or they do not. In the former case there is extra burden for applicants, in the latter for administrators and reviewers due to higher application influx.[11] However, given the already low success rate of the EF scheme, coupled with the overall pool of highly innovative ideas, we judge that the current approach is the right one. The EF scheme already attracts enough of the ideas it is trying to fund, and it cannot afford to grow, so the level of effort required to submit an application is justified as compared with the alternatives.

On the overall efficiency of the process, we note that at around 18 months from the initial call for applications to the final funding decision (and over 12 months for the assessment process itself), the overall process is quite lengthy. 'Time-to-grant' typically takes less than a year for

---

[11] For another example of this dilemma and its possible effects, see our recent review of the Human Frontier Science Program: https://www.hfsp.org/node/74873#book/

most research funding schemes, often as little as six months. Interviewed Jury members also pointed out that the assessment process is lengthy, and as researchers themselves, they suggested seeking ways to make it more rapid to allow for projects to start sooner. Interviewed Jury members think the gaps between the assessment stages are unreasonably long and could be shortened.

Our consultations with FWF staff have shown that this first call is something of an exception and that the overall process took significantly longer than planned. Due to ongoing budget negotiations at the time, the date for approval of the EFP could only be finalised after submission, which extended the total process by around three months. This extension mostly affected Stage 2, which lasted from May to December 2023. Our consultations indicate that this stage should have been about half as long, which would have resulted in a time-to-grant of around nine months. This shorter timeline is the plan for the next EF call.

On the clarity and completeness of feedback, we note that there appears to be a wider problem, which is also reflected elsewhere in our survey results. Applicants generally appear to have low confidence in the attention and expertise of those who reviewed their applications. We stress that this is a perception by applicants rather than a verdict of this evaluation. As we show in later sections of this report, the reviewer selection, allocation of applications, as well as reviewer diversity and other possible factors all appear to have been sound. We also note that these satisfaction and confidence levels are lowest for applicants who were rejected at Stage 1.

We refer back to these observations where relevant during the remainder of this report. However, based on the totality of our research we do not judge there to be an issue with the Jury members or reviewers themselves. Instead, there may be a need for greater transparency around reviewer selection and potentially for more structured and detailed feedback especially at the first assessment stage, where confidence and satisfaction levels are at their lowest.

Figure 7   Applicants' confidence in reviewers
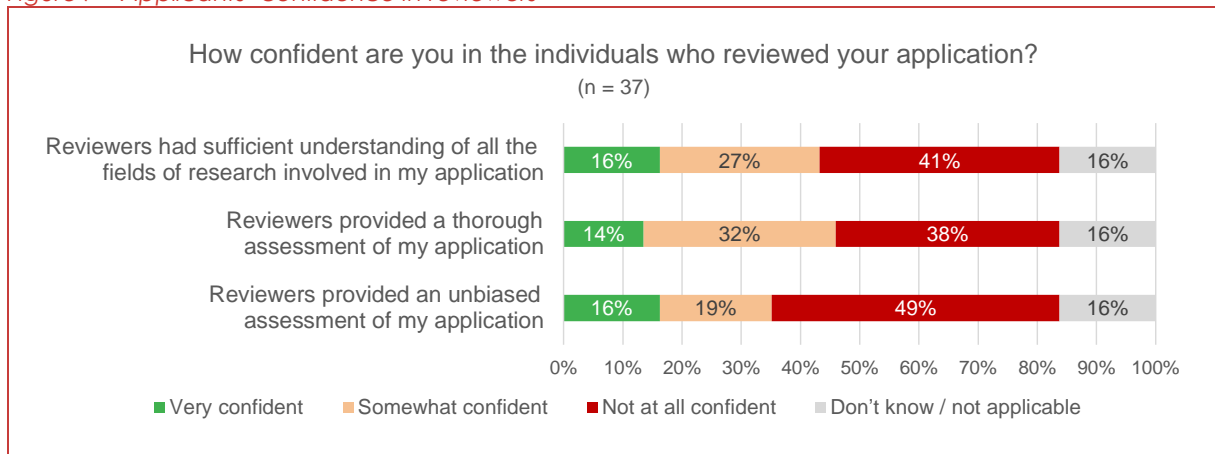
Source: Survey of EF applicants

3 Stage-by-stage

Research funding schemes may have multiple assessment stages for different reasons: the three assessment stages of the EF scheme may be used simply as three separate layers of risk-control and scrutiny. Given the large maximum size of EF awards, it is reasonable to ensure this level of scrutiny exists. However, short pre-proposals and presentation events have also been used specifically to encourage risk-taking by funders and to ensure that assessors' focus lies purely on the innovative potential of research ideas (rather than, for instance, detailed research plans or track records).

Our interviews with FWF staff indicate that the latter intention lies behind the structure of the EF assessment process. Specifically, the intention behind the process design was as follows:

- Stage 1 (Jury review of 3-page synopses) was intended primarily to assess and reward novelty of the proposed research ideas
- Stage 2 (external peer review of full applications) was intended primarily to assess scientific quality in the conventional sense (feasibility, robustness, etc)
- Stage 3 (Jury hearings) was intended primarily to assess team composition

These are not absolute separations. In essence, all stages could scan for each of the criteria noted above (and more), but each stage has an intended primary focus, reflecting its design (respectively: short description of idea, detailed scientific plan, presentation of the team).

Our headline conclusion of our stage-by-stage evaluation is that these different emphases were indeed operationalised in each stage more-or-less as intended. In other words, at the level of overall process-logic, the EF assessment process did what it was supposed to do. We substantiate this main finding for each of the stages in the respective sub sections below, alongside all other evidence pertaining to the operation of the individual process stages.

There is an important meta-question throughout the three stages of whose judgement holds the most sway in the process: who has power? While the FWF Scientific Board has formal decision-making power, the decisions are based on input from Jury members and reviewers. However, the evaluation has also examined whether the Scientific Board mostly 'reproduces' Jury and/or reviewer verdicts or whether and to what extent its decisions deviate from those verdicts. We touch on this question where relevant in the sub-sections below.

3.1 Stage 1: short synopses

The first stage involved a review of 3-page project synopses submitted as part of each EF application. The synopses were reviewed by a 16-person international Jury which was drawn together especially for the EF scheme. Each synopsis was reviewed by two Jury members and each Jury member provided a grade on the following scale:

- Grade 'A': Fully recommended; i.e., the project meets the programme objectives and I fully recommend it for Stage 2 of the review process
- Grade 'B': Recommend with reservation; i.e., I recommend the project for Stage 2 of the review process, but I have some concerns about one or several aspects of the synopsis and/or research team
- Grade 'C': Not recommended; i.e., I have major concerns about the proposal's potential to fulfil the objectives of the Emerging Fields programme

While there was an informal Jury meeting prior to the reviewing work in Stage 1, there was no panel-style review meeting of Jury members to discuss or rank the synopses. Instead, the independent Jury verdicts were then forwarded to the FWF Scientific Board (Kuratorium) for discussion and approval.

3.1.1   Jury allocation, scoring and behaviour

The Jury was put together through recommendation of several FWF staff, most notably the heads of the three main FWF science divisions. This was done before all synopses had been received, so the Jury itself was roughly equally balanced between different discipline-areas, irrespective of the actual disciplinary composition of the applications.

Our findings indicate that the written guidance provided to the Jury members is comprehensive and clear. FWF's online briefing to the Jury members was also helpful. However, to further discuss the approach to assessing the innovative aspects of the applications, the Jury chair organised two[12] informal Jury meetings (we stress that this was not an intended 'formal' part of the process but happened at the Jury chair's request). Interviewed Jury members found these additional meetings very helpful in ensuring consistency in their approach to scoring. They suggested that more attention could be given to this aspect in the FWF briefing.

At 16 members plus the chair, the Jury is too small to go into detail on its characteristics as this would border onto using or revealing personal identifiers. Instead of listing names, we simply represent each Jury member with a number as this is not an evaluation of individuals' performance or behaviour.

At an aggregate level, we note that the Jury consisted of six men and 11 women (including the chair). Members are based at institutions in 11 different countries. Most are located in Europe other than three who are located in the USA and one in Australia. The UK is most strongly represented with four individuals. No institution is represented more than once.

The Jury members conducted a total of 90 reviews (two per application), averaging 5.6 reviews per Jury member. However, there are substantial differences among individual Jury members in terms of reviews per person. Applications are allocated based on expertise, and while most Jury members reviewed around 1-5 applications, three members reviewed 10 or more, indicating a highly uneven spread of the reviewing workload and a predominance of these individuals' judgements in the pool of applications. The Jury chair did not review applications.

The higher number of applications tagged as 'Social Sciences and Humanities' is somewhat reflected in the inclusion of six Jury members from these fields compared with five from 'Biology and Medical Science' and 'Natural Science and Engineering' respectively. There are no significant differences in reviewing workload between the 'Natural Sciences and Engineering' and 'Social Science and Humanities' Jury members, though the 'Biology and Medical Science' Jury members conducted fewer reviews.

Our evaluation did not encounter any evidence of insufficient expertise in the Jury to review each synopsis. However, the uneven distribution of workload highlights that there is at least a slight mismatch between the Jury composition and the overall thematic/disciplinary mix of the EF synopses. This is a natural result of the Jury being selected prior to application submission, so there is a case in future EF calls to consider assembling or finalising the Jury at a slightly later stage, when the mix of applications is already known.

Table 2   Jury members' review load in Stage 1

Reviewer No.	Specialist Department FWF	Number of project evaluations	Total per department	Average per department
1	Biology and Medical Science	6	22	4.4
2	Biology and Medical Science	5		

---

[12] Two meetings were held because all Jury members could not join at the same time because of time zone differences.

Reviewer No.	Specialist Department FWF	Number of project evaluations	Total per department	Average per department
3	Biology and Medical Science	4		
4	Biology and Medical Science	4		
5	Biology and Medical Science	3		
6	Natural Science and Engineering	12		
7	Natural Science and Engineering	11		
8	Natural Science and Engineering	4	32	6.4
9	Natural Science and Engineering	4		
10	Natural Science and Engineering	1		
11	Social Science and Humanities	10		
12	Social Science and Humanities	7		
13	Social Science and Humanities	7		
14	Social Science and Humanities	6	36	6
15	Social Science and Humanities	3		
16	Social Science and Humanities	3		

Source: data provided by FWF

Jury members' grading ranges differ significantly. Many Jury members gave 'A' grades in 50% of their reviews or even more frequently. Others have a far higher propensity to give 'B' or 'C' grades. This may of course simply be a reflection of some Jury members receiving predominantly high-quality applications to review while others received generally lower-quality ones. Indeed, initial interviews suggest that Jury members did have applications of different quality levels, while also engaging in informal discussion about how to consistently apply the A-C grading scale (organised by the Jury chair as noted above).

However, the figures below raise the possibility that the A-C grading scale may have been applied differently by different Jury members. We also note that only a small number of Jury members opted for mixed grades (AB or BC), so it is unclear whether all Jury members perceived these as equally viable grading options.

Table 3   Scores by Jury members in Stage 1

Reviewer No.	Dept	Total reviews	A		AB		B		BC		C	
1	NatTec	12	0	0%	0	0%	6	50%	0	0%	6	50%
2	NatTec	10	4	40%	0	0%	4	40%	0	0%	2	20%
3	HumSoc	10	2	20%	0	0%	4	40%	0	0%	4	40%
4	HumSoc	7	1	14%	0	0%	3	43%	0	0%	3	43%
5	HumSoc	6	1	17%	1	17%	1	17%	2	33%	1	17%
6	BioMed	6	3	50%	0	0%	1	17%	0	0%	2	33%
7	BioMed	5	1	20%	0	0%	1	20%	1	20%	2	40%

Reviewer No.	Dept	Total reviews	A		AB		B		BC		C	
8	HumSoc	5	3	60%	0	0%	1	20%	0	0%	1	20%
9	BioMed	4	3	75%	0	0%	0	0%	0	0%	1	25%
10	BioMed	4	2	50%	0	0%	2	50%	0	0%	0	0%
11	NatTec	4	2	50%	0	0%	1	25%	0	0%	1	25%
12	NatTec	4	2	50%	0	0%	1	25%	0	0%	1	25%
13	HumSoc	3	2	67%	0	0%	0	0%	0	0%	1	33%
14	BioMed	3	1	33%	0	0%	1	33%	0	0%	1	33%
15	HumSoc	2	1	50%	0	0%	1	50%	0	0%	0	0%
16	NatTec	1	0	0%	0	0%	1	100%	0	0%	0	0%

Source: data provided by FWF. Top and bottom three (four if equal) by percentage highlighted green/red. NB: the Jury members here are listed in order of number of reviews completed. The data in the lower rows of this table should therefore be disregarded as it is based on a very small number of reviews easily skewed by individual cases.

3.1.2   Stage 1 decision-making

At the FWF Scientific Board meeting immediately following the synopsis reviews, only those synopses that had two significantly diverging Jury scores were discussed by the Board. All that scored highly with both Jury members were automatically passed to Stage 2 while all scoring poorly were rejected without further discussion. This left only eight applications that were discussed. In other words, the Board intervened in just 18% of all applications submitted to the EF scheme. For the remainder, the decision to pass or fail was entirely based on Jury grades. Since the Board only discussed applications where the two Jury verdicts were genuinely 'poles apart', there is no sense of the Board overriding the Jury.

Of the eight applications discussed, three were passed to Stage 2 and five were rejected. For most of the applications, most of the discussion focused on the transformative/innovative nature of applications (or lack thereof) and levels of risk in terms of feasibility. Generally, those applications that were discussed positively in these two categories were passed. This indicates that the main aims of the EF programme were operationalised well in the Scientific Board discussion. Moreover, at the level of the Scientific Board, the novelty of research ideas appears to have been the main criterion determining progression from Stage 1.

The picture of actual Stage 1 outcomes is a little more nuanced in relation to applications' characteristics. Our novelty indicators (see Appendix section A.1) do not show evidence of bias against novel ideas in the Stage 1 review process, though they also do not suggest any drastic novelty appetite within the process either. In other words: when we compare the novelty-levels of successful and unsuccessful synopses at Stage 1, there are no significant differences evident from our analysis.

This may not be a problem, as our bibliometric and textual analyses and preliminary findings from the interviews with Jury members indicate that the EF programme attracted highly novel applications in general. Two of the Jury members interviewed had previous experience assessing applications to FWF programmes and concluded that applications they reviewed for the EF programme were significantly more novel than those they had reviewed for other FWF

programmes. The EF programme was able to convince the scientific community to think 'outside the box' and to submit novel ideas. The overall lack of variation in novelty indicators may therefore be explained by the fact that Stage 1 also needed to fail some research ideas that were 'too' novel, i.e. entirely unfeasible or totally unrealistic (though we cannot confirm this).

The second part of our textual analysis, namely our analysis of reviews, strongly suggests that novelty played an important role in Stage 1 verdicts:

- In terms of priority, our analysis suggests that written reviews of synopses focus much more on the criterion of novelty than on scientific quality, with risk/feasibility and team characteristics occupying intermediate positions in terms of the space dedicated to these criteria in the reviews

- Moreover, successful synopses are associated with a slightly greater focus in reviews on novelty while unsuccessful synopses are associated with a greater share of reviews dedicated to quality issues

- In terms of review sentiment (whether reviews are supportive or critical on each criterion, rather than the space devoted to each criterion), we find that all four criteria are rewarded by reviewers. Successful synopses had much more positive reviewer sentiments for novelty, risk, scientific quality and team qualifications than unsuccessful ones. This indicates that while novelty may have been a focal point at Stage 1, Jury members were also considering these other criteria and they almost certainly played a large part in the allocated grades. In combination with the overall high novelty-levels, this goes some way to explaining the lack of increased novelty-levels as we move from Stage 1 to Stage 2

The full analysis of reviews can be found in Appendix section A.2. We present more detailed results of this analysis in the next main section when we contrast reviewer behaviour in Stage 2 with Jury member behaviour in Stage 1.

3.1.3   Applicants' experience and views on Stage 1

From the applicants' perspective, we find no major issues with the operation of Stage 1. In the first main section of this report, we presented aggregate satisfaction data, so the submission and support processes appear to be in good health.

Before presenting applicant views on some specific issues, it is worth noting some additional comments from Jury members. Several Jury members we interviewed observed that FWF guidance to the applicants about the contents of the synopses could be improved. Some noted significant heterogeneity in how applicants presented the team. While some applicants provided detailed and justified team presentations, others placed much less emphasis on this aspect, making the assessment challenging. Given the programme's focus on transformative ideas, some applicants may have perceived the team description as less relevant, though the programme documentation notes that the team will be considered at Stage 1. There is a case therefore to provide clearer guidance or prescribed structure to the synopses.

While there may be some uncertainty regarding how to structure synopsis content, applicants themselves seem to be broadly satisfied with the maximum permitted length. Clearly it is a challenge to condense a large research project idea into just three pages, but three quarters of applicant survey respondents note that the length was about right, with one quarter saying the synopses should have a longer maximum permitted length.

Figure 8 Applicants' assessment of synopsis length

The first stage of the assessment process was based on the short synopsis of your application. What is your view on the permitted 3-page maximum length of the synopsis?
(n = 37)

- The maximum synopsis length was too short and it should be made longer — 24%
- The maximum synopsis length was of roughly the right length to crystallise the central idea(s) of the project — 73%
- The maximum synopsis length was too long: an even shorter format would have been better — 3%

Source: Survey of EF applicants. Survey item included the option: 'Don't know / no opinion', which was not selected by any respondents. NB: answer options above are abridged for presentation. Full answer options in the survey item were: The maximum synopsis length was too short to suitably describe the central idea(s) of the project and it should be made longer; The maximum synopsis length was of roughly the right length to crystallise the central idea(s) of the project; The maximum synopsis length was too long: an even shorter format to suitably describe the central idea(s) of the project would have been better.

The overall popularity of short synopses is mixed, especially in terms of wider applicability. Of our survey respondents, 36% say it works for the EF scheme but should not be implemented elsewhere, while around 30% respectively say either that it may be helpful in some other funding schemes or that it should even be used for EF.

Figure 9 Applicants' assessment of the appropriateness of synopses in peer review process

Do you think synopses for the first stage of the assessment process should be expanded to other research funding schemes?
(n = 37)

- Yes, short summaries for the first assessment stage should be used as widely as possible — 6%
- Yes, in some other research funding programmes it would be helpful — 31%
- No, it works for the FWF Emerging Fields programme, but should not be used elsewhere — 36%
- No, and it should not be part of the Emerging Fields assessment process — 28%
- Don't know / no opinion — 3%

Source: Survey of EF applicants. NB: graph title is abridged for presentation. Full question in the survey item asked: 'Do you think short proposal summaries (synopses) for the first stage of the assessment process should be expanded to other research funding schemes?'

As suggested briefly in the first main section of this report, one major issue from the point of view of applicants lies with feedback. We present below our survey data on satisfaction with feedback, disaggregated into applicants rejected at Stage 1, rejected at Stage 2, and invited to the Stage 3 hearings. Dissatisfaction with feedback is strongly associated with Stage 1.

The clarity and completeness of the feedback I received

	Very satisfied	Somewhat satisfied	Neutral	Somewhat dissatisfied	Very dissatisfied
Invited to hearing (n = 9)	78%	11%	11%		
Rejected at stage 2 (n = 11)	36%	18%	9%	27%	9%
Rejected at stage 1 (n = 13)	15%	8%	15%	8%	54%

■ Very satisfied   ■ Somewhat satisfied   ■ Neutral   ■ Somewhat dissatisfied   ■ Very dissatisfied

In a similar vein, applicants rejected at Stage 1 are far more likely to have low confidence in the thoroughness of their application assessment.

Reviewers provided a thorough assessment of my application

	Very confident	Somewhat confident	Not at all confident	Don't know / not applicable
Invited to hearing (n = 10)	20%	30%	10%	40%
Rejected at stage 2 (n = 11)	18%	55%	27%	
Rejected at stage 1 (n = 14)	7%	21%	71%	

■ Very confident   ■ Somewhat confident   ■ Not at all confident   ■ Don't know / not applicable

As mentioned, we find no evidence of bias or of 'poor fit' between application topics and Jury member expertise.[13] The feedback meanwhile was relatively short but certainly had substance. Based on our interviews, document review and supplementary written answers of our survey respondents, we judge there to be issues around transparency, which may have affected applicants' perceptions: to the applicants, it would have been unclear how the Jury was composed, how applications were allocated to Jury members, how the different criteria were weighed or translated into the eventual rankings, or indeed how these rankings translated into an actual pass/fail verdict. Such lack of clarity is particularly significant for two reasons: first, the 3-page synopsis stage departs significantly from the 'standard' peer review process typical in ordinary research funding (Stage 2 of the EF process is a far closer reflection of the 'norm' in

---

[13] As noted, any mismatch that existed at the aggregate level between the total pool of applications and the Jury composition appears to have been balanced out by allocating applications very unevenly among Jury members.

research funding processes). Therefore, there was likely a greater need to build trust among the researcher community.[14] Second, there is the issue that applicants put in a substantial amount of work on their full application, much of which will never be read if they fail at Stage 1. This may substantially increase sensitivity to any perceived bias or unclear process elements.

In short, we find that Stage 1 is broadly functioning well, especially in terms of its prioritisation of novelty as a focus of attention. However, there is a need to better define and outwardly explain several details of the process. Several of our recommendations we set out at the end of this report therefore relate specifically to Stage 1.

3.2   Stage 2: external review

Stage 2 of the EF assessment process reflects most closely the standard process used in most research funding across the world. The 24 applications passed from Stage 1 to Stage 2 were each sent to three international academic experts for external peer review. The experts provide full written reviews, including scoring on a standard 5-point scale also used across other FWF funding schemes (Excellent/ Very good/ Good/ Average/ Poor).

External reviewers are identified and selected by the FWF Scientific Board with assistance from the heads of the FWF scientific departments. Each application has a main responsible Board member (and a deputy). Following completion of all external reviews, the Board meets to take decisions, based on reviews and each responsible Board member's assessment, which applications proceed to Stage 3. In our understanding, this is a standard procedure used throughout much of the FWF's funding processes.

3.2.1   Function and priorities of Stage 2

Our research highlights from several points of view that, while novelty was a key focus of the assessment and decision-making in Stage 1, Stage 2 focussed strongly on the scientific quality of applications in the conventional sense. This applies both to the external reviews themselves as well as to the Scientific Board discussion.

Our analysis of review texts highlights a striking contrast between reviews of Stage 1 synopses and reviews of full applications at Stage 2. As noted above, novelty was a main thematic focus of reviews from Jury members at Stage 1, while scientific quality considerations did not occupy much space. At Stage 2, we see a near-reversal of this: having been the lowest priority at Stage 1 (i.e. the least amount of space devoted to it in the written reviews), scientific quality has by far the highest priority at Stage 2. As with Stage 1, the priority of focus on scientific quality is not absolute: risk/feasibility, team characteristics and indeed novelty all have some space devoted to them. However, none of these come close to the priority given to scientific quality.

---

[14] Technopolis' recent Review of Peer Review for UKRI discussing the literature on attitudes to traditional review processes, including the high level of trust many researchers have in the 'standard' process of 2-3 external reviews followed by panel ranking, while more unorthodox approaches may solve many problems but also require careful consultation, piloting and trust-building: https://www.ukri.org/publications/review-of-peer-review/

Figure 12 Priority areas in reviews: Stage 1 and Stage 2 reviews compared
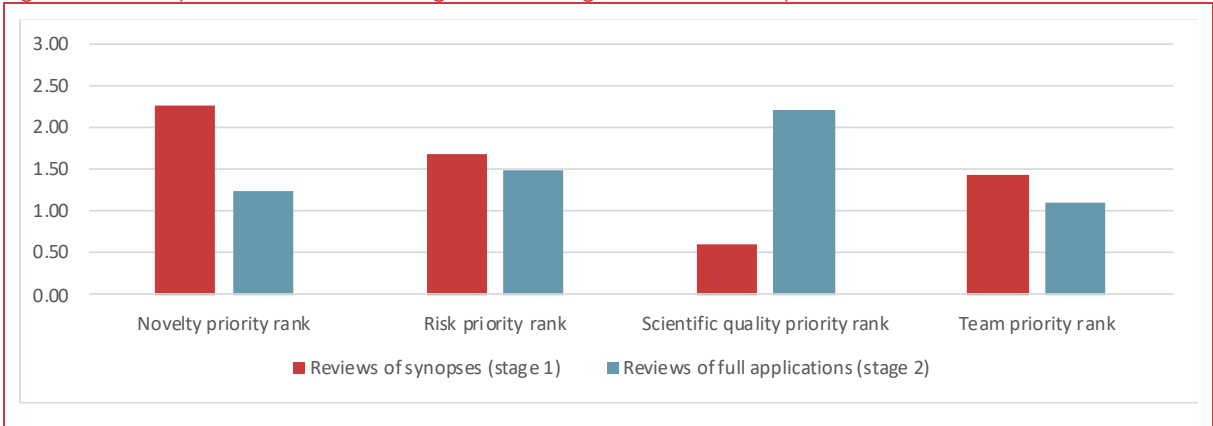
This headline observation from our analysis of reviews is also supported by views from the reviewers themselves. We asked external reviewers to rank how strongly various criteria influenced their overall judgement of EF applications. We were able to take a more granular approach here, distinguishing between 10 different criteria rather than the four used in the analysis of reviews. The results show that the criteria most conventionally associated with scientific quality (rigour and robustness, significance of the topic) played the greatest role, while our proxy for novelty ('the extent to which the proposed work departs from established or conventional methods, approaches or perspectives) has comparatively smaller influence. We note nevertheless that all criteria we surveyed for do appear to have played at least some part in reviewers' verdicts.

Figure 13 Influence of different criteria on reviewers' judgement
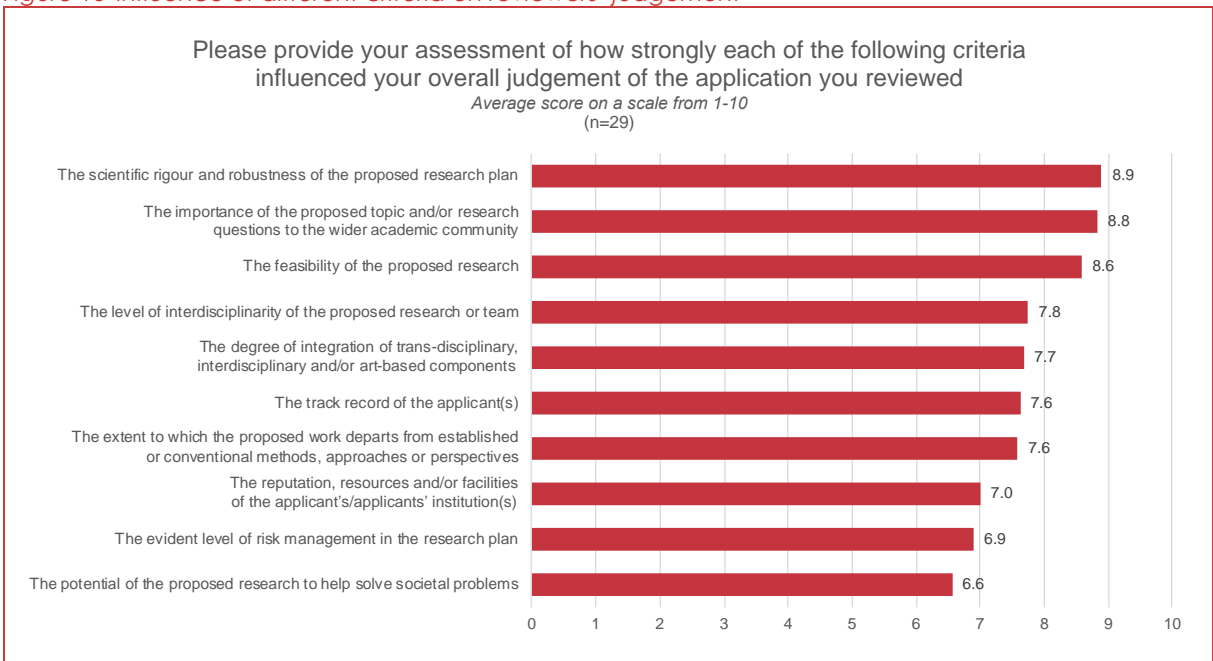
Source: Survey of EF Stage 2 external reviewers. Survey item included a 'Don't know/' option for each criterion. which was selected a total of three times across the entire item. NB: Survey item included the following instruction: "Please rank each criterion on a scale from 1 to 10, where '1' means 'no influence' and 10 means 'extremely strong influence'".

Finally, we note that this emphasis on scientific quality also extended to the Stage 2 FWF Scientific Board meeting. Our overarching conclusion from our observation exercises (see Appendix D) is that the FWF Scientific Board placed the greatest emphasis on synopses' innovativeness at Stage 1, and on scientific quality of proposals at Stage 2. At both stages, other factors were also considered, and it is worth noting in particular the presence of many comments on proposals' novelty as a source of support at Stage 2. In other words, as with the reviews themselves, the Board discussions focused on a range of criteria but with scientific quality taking distinct priority.

In this context, we note that the role of the Scientific Board was much expanded compared with Stage 1. Unlike in Stage 1, there were no 'automatic' passes or rejections of applications based solely on external reviewers' scores. Every application was discussed, often in some detail, regardless of the overall level or divergence among external reviewers' scores.

3.2.2 Views from the external reviewers

We can report an almost entirely positive picture in terms of external reviewers' own views and perceptions. Aggregate process satisfaction scores in our survey of reviewers are consistently high on all aspects we surveyed for. The highest proportion of dissatisfaction is on the appropriateness of feedback required from reviewers, though even here, only a small minority express dissatisfaction. We note it nevertheless as we have touched previously on the need to systematise feedback, given the low satisfaction rates among applicants.

Figure 14 External reviewers' satisfaction with administrative aspects of the EF process
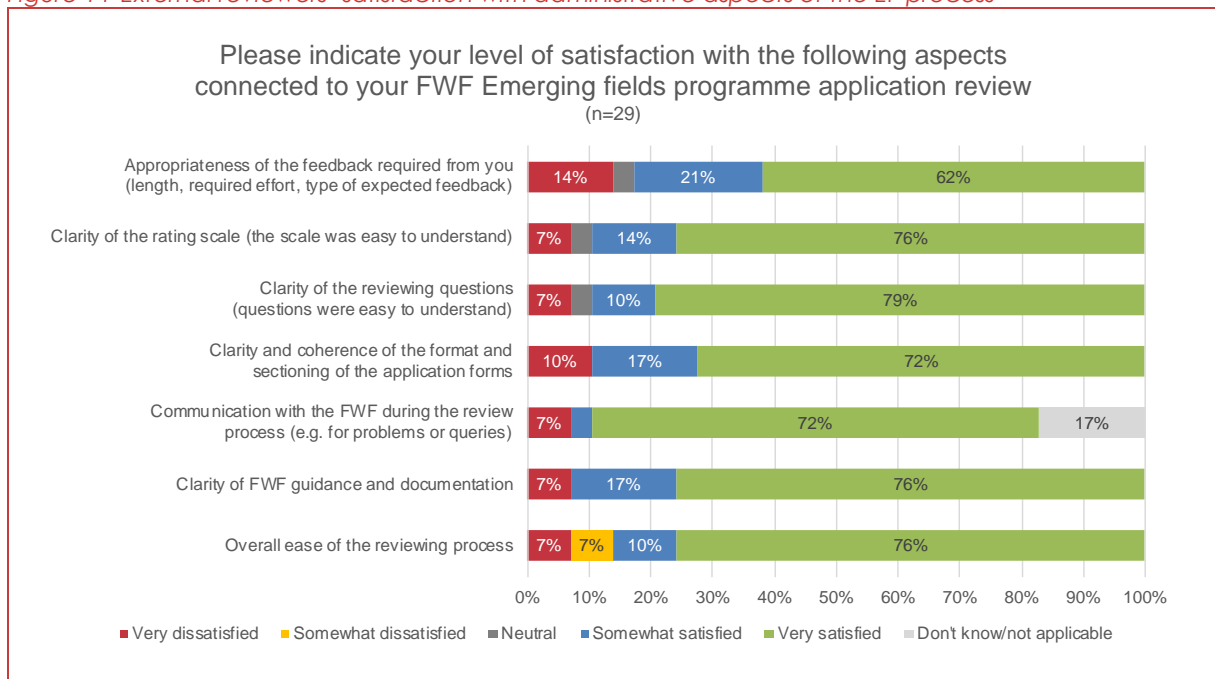
Source: Survey of EF Stage 2 external reviewers

We asked reviewers to provide views on various additional aspects besides the process satisfaction elements listed above. In short, we find that reviewers were almost entirely mid-to-late career academics and that the subject matter of the EF application they received matched their expertise in large part. For the most part, reviewers also judge the applications to have been of suitable length and to have had sufficient time to prepare their review.

Figure 15 Survey of reviewers: data on experience, fit, time and application length

How many years of full-time professional research experience (excluding PhD study) did you have when you reviewed the application for the FWF Emerging Fields programme?
(n=30)

0-4 years	5-9 years	10-14 years	15-19 years	20-24 years	25-29 years	30 years or more
0%	7%	20%	17%	20%	13%	23%

Reflecting on the overall length of the application form, please indicate whether the length of the application form was sufficient, or too short, or longer than necessary for you to make a robust and reliable assessment
(n=29)

The application form was much too short	The application form was a little too short	The application form was about the right length	The application form was a little longer than was necessary	The application form was much longer than was necessary
0%	3%	69%	21%	7%

Reflecting on the time window available for you to conduct the review, how appropriate was it?
(n=29)

A much shorter time-window would have been sufficient	A slightly shorter time-window would have been sufficient	The time-window for me to conduct the review was about right	The amount of time should have been slightly longer	The amount of time should have been a lot longer
0%	0%	69%	24%	7%

To what extent did the application that the FWF provided you for review reflect your research expertise?
(n=29)

The application I was given closely or entirely matched my expertise	The application I was given mostly matched my expertise	The application I was given only partially matched my expertise	The application I was given hardly matched my expertise	The application I was given did not match my expertise at all
52%	45%	3%	0%	0%

Source: Survey of EF Stage 2 external reviewers. Each survey item included a 'Don't know/Not applicable' option, which was not selected by any respondents in any of the above items. NB: answer options for the question on available time-window are abridged for presentation. Full answer options in the survey item follow the format, "A much shorter time-window for me to conduct the review would have been sufficient".

While there are signs of some minor issues around time-window and application-fit, our observations and interviews highlight that these may be explained by special cases. There were two instances where reviewers noted that they had insufficient expertise to provide reviews meaningfully and so additional reviews were sought for these cases. Further, some

reviewers were only contacted quite late in the process due to difficulty finding reviewers for some applications. These likely explain much of the 30% of responses noting that the time window could have been longer.

3.3   Stage 3: Jury hearings

> **"I like both [applications]. One is an apple; the other is an orange."**
>
> *Anonymous. Jury member in discussion after the Stage 3 hearings*
>
> **"This one seems more like a project than a 'field'."**
>
> *Anonymous. Jury member in discussion after the Stage 3 hearings*

Stage 3 was the shortest part of the assessment process, spanning less than two months from the communication of Stage 2 decisions and culminating in the Jury hearings in Vienna. Ten applicant teams were invited, of which five were funded. The hearings themselves took place over a day and a half (six on day one, four on day two) and were followed by Jury discussion and decision-making on the afternoon of day two, continuing the following morning.

Due to the very small application numbers, we do not provide quantitative analysis for Stage 3. However, we observed the hearings and subsequent discussions in full and conducted several interviews afterwards. Based on these sources, we do not find any significant problems relating to the hearings.

Across the hearings and especially in subsequent discussion, we observed that there was much need among Jury members for debate around the core terms of the EF scheme. There was considerable discussion about what exactly novelty means, and it was frequently evident that it means different things in different contexts, including in different disciplines. There was also discussion about what it meant to be 'emergent': what if, for example, a team had already had a large grant (e.g. ERC) for a closely related topic? The required benchmark for overall significance also came up: at what point can something be described as a 'field'? The importance of interdisciplinarity was also a point of discussion, as it was unclear whether interdisciplinarity itself should be rewarded.

These were all important questions that informed decision-making and the hearings and subsequent discussions acted as fora to address them. However, elements of the discussion could likely have been sped up if some of these concepts had been more clearly defined from the beginning. In detail, our main findings are as follows:

- Interviewees felt that the hearings were well organised and excellently supported by FWF staff. Time keeping was very precise throughout the hearings with all components taking up precisely the amount of time they were supposed to. The event met high standards of accessibility (e.g. large clock and presentation display, sound optimisation, wheelchair access), although one presentation encountered some technical difficulties with the equipment provided

- Jury interviewees said that having the third assessment stage in this form was very important. They brought up issues such as the ability to probe into any unclear methodological aspects. However, they consistently identified meeting the project teams as the most important aspect. This included both the ability to get a clearer sense of who was doing what, but also to observe the team dynamics, collegiality and hierarchy

- There was a wide range in how teams chose to present: some had an evident lead presenter, others shared the presenting equally. It was generally viewed critically when a

presentation was to focused on one individual, so it may be worth in future calls to instruct that all team members should be involved in the presenting

- Presentations were structured in similar ways, each focusing on wider significance of the topic, more detailed methodological aspects, and presentation of the team, though in varying degrees of order and emphasis

- Jury questions and comments to the teams (as well as discussion after presenters had left) related to a wide range of criteria. Questions and discussion points covered technical methodological points, novelty, ethics and risks, as well as qualifications and resources available among the team

- While each application had lead Jury members assigned to it (based on closest match of field expertise), and these Jury members began the questioning, we observed substantial involvement of other Jury members in both questioning and subsequent discussion. Interviews indicate that this was viewed as being very helpful: for awards of this size, Jury members felt that teams should be able to make a case for the importance of their topic beyond experts from the relevant fields. Some Jury members even noted that the amount of question time should be capped for the 'leading' Jury members so that the remaining non field-expert members would have enough time to ask their questions. In a small number of cases, all or almost all the question time was taken up by the lead Jury members

- While there was extensive discussion after the presentations, there was in fact a clear consensus almost immediately about the rough order of applications from best to worst. From initial scores (which formed the basis of discussion) to the finalised list, there were almost no changes in ranks. This prompted some Jury members to wonder whether a shorter amount of time for Jury discussion would have been sufficient

- Several details of the discussions were shaped by the chair, who had substantial discretion in how to run the proceedings. It was made clear by FWF staff that the initial Jury grades were only intended as a starting point, and from this point onwards, the chair was able to structure the discussion, and even propose additional ranking exercises to arrive at a definitive verdict. This means that the selection of Jury chair is important, and they should certainly be someone with substantial prior experience in chairing events of this type (which was the indeed the case and it should remain so in future EF calls)

- While we find no evidence of overt bias or discrimination, we note that of all of the five eventual winning applications had male coordinators, and no applications from the humanities were successful. This was acknowledged by the Jury and a degree of self-reflection took place. However, the discussion concluded that while these outcomes were unfortunate, they were fully attributable to differences in quality

- A couple of applications received quite poor verdicts from the Jury and there was some discussion about how this was possible, given that all applications had gone through the Stage 1 assessments and full scientific peer review in Stage 2. Some Jury members noted that they recalled far more promising-sounding outlines from Stage 1 which had not made it to Stage 3

4 Relevance and added value

In this section we focus on wider issues around the EF scheme to draw conclusions about its overall relevance, function and added value. We look first at the overall size and demand for the scheme before considering what would happen in its absence. Finally, we consider the position of the EF scheme in the wider FWF portfolio and any possible synergies.

4.1 Size and demand

As noted at the outset, the EF scheme attracted a pool of highly novel research ideas, many of which had previously been nascent or had been specifically thought up in response to the EF call. To an extent, the scheme thereby created its own demand. In doing so, it certainly managed to unleash creative potential in the Austrian research base that otherwise may have remained at least partially undiscovered.

As mentioned earlier, the overall success rate is rather low at 11%. Beyond the top five, no other application received a 'C1' grade (fully fundable if budget were available) but there were several 'C2' verdicts (fundable if budget were available, with minor amendments). The overall quality-profile therefore suggests that the available budget is roughly right. However, in future calls there may be a risk that there is too little money for otherwise fundable applications if all figures and quality profiles remain roughly stable.

If the overall scheme budget cannot be changed, we note that a sizeable minority of applicants would support a slightly smaller maximum award size, if this would enable a slightly higher number of awards to be funded.

Figure 16 Applicants' views on the potential to adjust award size
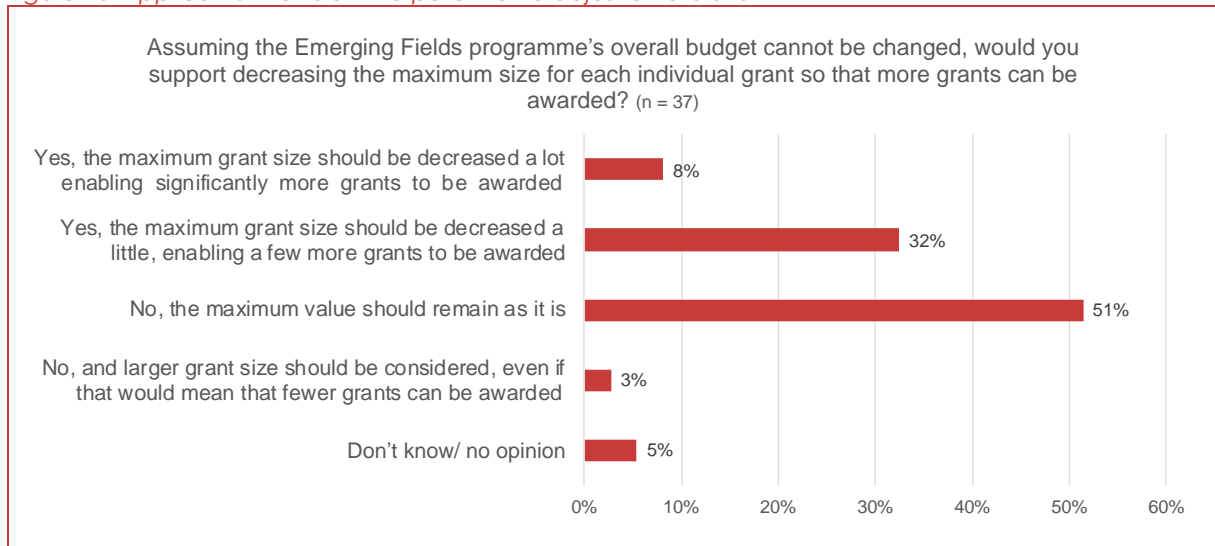
Source: Survey of EF applicants. NB: NB: answer options above are abridged for presentation. Full answer options in the survey item were: 'Yes, the maximum grant size should be decreased a lot (to less than half of the current maximum value), enabling significantly more grants to be awarded', Yes, the maximum grant size should be decreased a little (to half of the current maximum value or more), enabling a few more grants to be awarded', 'No, the maximum value should remain as it is', 'No, and larger grant size should be considered, even if that would mean that fewer grants can be awarded', and 'Don't know/ no opinion'

4.2 Survey-based counterfactual

In a process evaluation of such a new and specialised scheme, we cannot provide a meaningful counterfactual or control group in the strict sense. However, we used our survey of applicants to ask them what they think would happen to their project idea in the absence of EF funding (either through unsuccessful application or if the EF scheme did not exist). This is of course an imperfect measure, but these results begin to provide an idea of the scheme's added value.

Only 11% of respondents believe that their EF project idea would end completely without the EF funding. However, almost three quarters of respondents believe that their project would either progress with much reduced scope (due to likely lack of funding) or that they would modify their project by removing some of the novel, high-risk/high-reward aspects to increase their chances of securing alternative funding.

Figure 17 Projected view of project ideas in lieu of the Emerging Fields funding
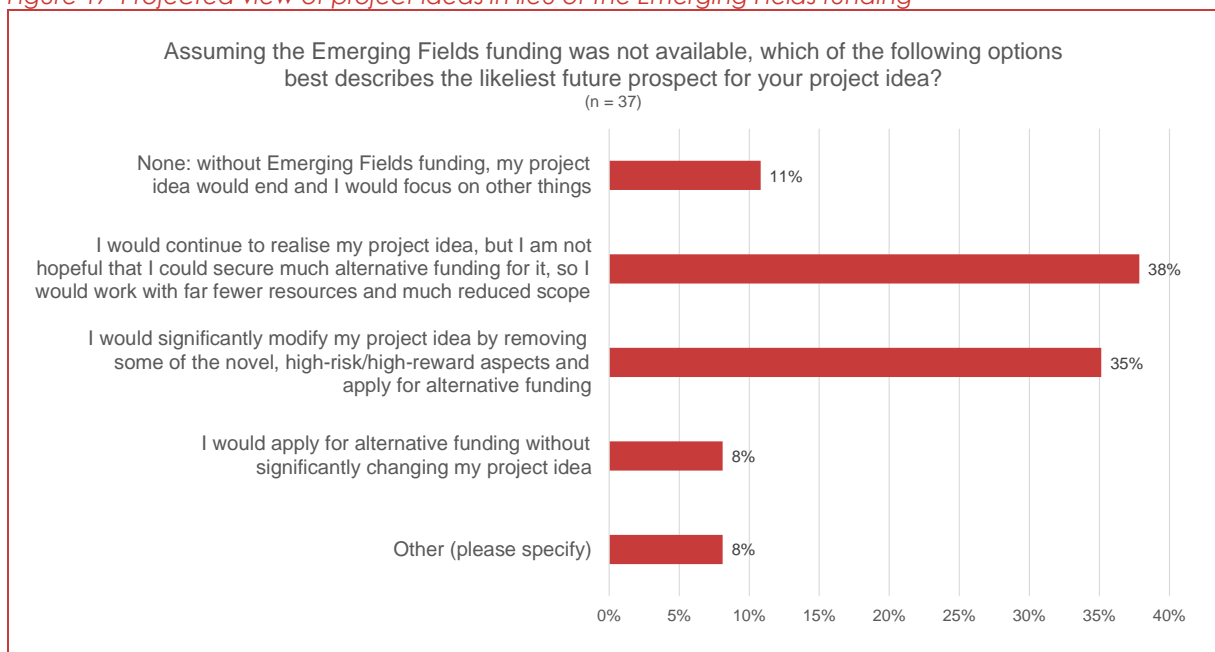
Source: Survey of EF applicants. NB: graph title is abridged for presentation. Full question in the survey item asked: "Assuming the Emerging Fields funding was not available (either because your application is rejected or because the programme simply did not exist), which of the following options best describes the likeliest future prospect for your project idea?". NB: Two applicants selecting 'Other' reported applying for alternative funding (e.g., an ERC Synergy Grant) but did not specify the extent or manner of modifications they were prepared to apply to their ideas.

These results strongly suggest that the true added value of the EF scheme lies in its provision of funding for exceptionally novel ideas. Indeed, our interviews and document review suggest that there are certainly other large research awards available to researchers in Austria. However, these are focused much more on 'ordinary' basic research. We return to this issue in the next sub-section.

In addition to the counterfactual survey question above, we also asked applicants which alternative funding sources they might consider if EF were unavailable. As the major public research funder in Austria, we would expect other FWF schemes to be the main 'address' for this. However, greater numbers of respondents indicated they would probably seek funding from outside Austria, including from EU programmes like ERC. ERC grants are about the same size as EF awards. However, while ERC grants were traditionally understood as being vehicles

for highly novel, high-risk/high-reward research ideas, recently doubts have been raised about the strength of this connection.[15]

While these results indicate that at least some of the EF project ideas would not be fully lost without the programme, they do indicate that within the Austrian system itself there are few perceived alternative funding sources and applicants would look further afield in the absence of EF.

Figure 18 Alternative funding sources in lieu of the Emerging Fields funding

Assuming the Emerging Fields funding was not available, how likely would you be to seek funding for the same project idea from each of the following sources?
(n = 37)

	Very likely or certain	Somewhat likely	Unlikely or not at all
A different FWF grant funding scheme	25%	53%	22%
A funding scheme from an Austrian funder other than FWF	15%	15%	71%
A funding scheme from a public funder based outside Austria (incl. EU Horizon/ERC)	43%	41%	16%
Internal funding from my institution		13%	87%

Source: Survey of EF applicants

4.3   The programme in wider context

The EF programme was designed as part of the Excellent=Austria initiative, which sought to provide much-needed additional funding for basic research in Austria. Parts of the initiative have an emphasis on attracting and retaining the best international talent, while other parts of the initiative are focused more on fostering ideas among researchers already based in Austria. EF firmly addresses the latter of these aims, while also adding the element of fostering highly novel, high-risk/high-reward research ideas. The programme was also explicitly intended to be open to interdisciplinary, transdisciplinary and arts-based research proposals.

Our interviews of all stakeholder groups consistently suggest that a programme with these characteristics was indeed needed, and the observed demand and success rate further confirm this. Considering the highly novel profile of applications and the substantial share of interdisciplinary, transdisciplinary and arts-based applications, we also conclude that the programme managed to attract the breadth of ideas that was intended.

In terms of the overall logic and ambition, we therefore do not find any significant problems or issues. However, it is worth contextualising EF directly alongside related FWF programmes.

---

[15] Veugelers, R., Wang, J., & Stephan, P. (2022) 'Do funding agencies select and enable risky research: Evidence from ERC using novelty as a proxy of risk taking' (No. w30320). National Bureau of Economic Research; Luukkonen, T. (2012) 'Conservatism and risk-taking in peer review: Emerging ERC practices.' Research Evaluation, 21(1): 48–60.

The FWF programmes commonly mentioned by stakeholders as comparators to EF were the 1000ideas programme and the special research programmes ('Spezialforschungsbereiche', SFB).

The 1000ideas programme was frequently noted as being 'related' to EF as it is the only other FWF programme with an overt focus on highly innovative high-risk/high-reward research. However, it provides awards at a much smaller scale (€50,000-€175,000), is targeted at very early-stage exploratory research, and generally attracts early career researchers (though it is open to all). We explored with several interviewees, whether there should be any explicit synergy between 1000ideas and EF, e.g. an automatic recommendation or even 'fast-track' to EF if a 1000ideas award has fruitful results, possibly via an intermediary programme (such as a hypothetical EF-style award but at a smaller scale comparable to the 'Forschungsgruppen'). However, there was no appetite or perceived need for such explicit connections or 'funding-ladders'.

The SFBs by contrast do not share the EF emphasis on novel research but were judged to be the EF programme's closest 'relative' in terms of type and size of award (collaborative teams working over several years with several million Euros at a scale better described as a field than a project). Here, the potential for synergies was brought up far more frequently. Many stakeholders also noted that institutions and academics struggled at times to fully understand the difference between SFBs and EF.

In fact, several stakeholders suggested and/or supported the idea of a fusion of SFBs and EF (with a combined budget). This might involve either a dual track where teams can apply each year either for a 'regular' or for a 'high novelty' SFB (with separate assessment processes), or alternating between regular and high-novelty SFBs year-on-year. This would have a number of benefits. First, it would create a small degree of streamlining in the FWF portfolio (some interviewees noted that funding offers should not be over-complicated by having too many distinct programmes). Secondly, there would be a consistent understanding of the expected scale of projects (what is a 'field'?) and of the status of interdisciplinary, transdisciplinary and arts-based approaches. It could be clarified that across FWF's large-consortium funding for basic research, such approaches are welcome.

In short, at a strategic/portfolio level, the EF programme fulfils an important role both in terms of providing additional funding for large basic research initiatives and through signalling the FWF's interest in funding highly novel research ideas. However, EF does not necessarily need to be a fully separate stand-alone programme.

5 Conclusion and recommendations

Our headline conclusion is that the Emerging Fields (EF) programme has fulfilled a critical signalling and galvanising function. By launching the programme, the FWF provided an opportunity for Austria-based researchers to formulate a wide range of highly novel research ideas. All the strands of our research indicate that the overall pool of 45 submitted ideas had very high levels of novelty, and that many of these ideas were either nascent or were fully formulated only in response to the EF call.

In the wider context of the Excellent=Austria initiative, the EF call has thereby fulfilled an important function by adding funding for basic research and by providing a vehicle for innovative ideas to take shape.

Further, we find that the three-stage EF assessment process itself is well designed and broadly in good health. In line with the original design intention, Stage 1 has a strong focus on innovative aspects of the research ideas. By focussing on short synopses, this stage also ensured that the burden of extensive peer reviewing of full applications was minimised. Stage 2 was a full formal external expert peer review focused primarily on scientific quality, while the Stage 3 Jury hearings allowed for a broad range of questions about many criteria but also could provide insight into team composition, dynamics and hierarchy. Though it was formally the designated decision-making body, the FWF Scientific Board played only minimal roles at Stages 1 and 3 but a major role at Stage 2. The Board's decision-making style also reflected the intended process logic at each stage.

We find a small number of challenges. First, there are low satisfaction rates with feedback and evident issues around process transparency, especially at Stage 1. Second, while the initial pool of applications and applicants is diverse, there are inequitable outcomes in terms of gender and discipline, with no humanities/arts applications ultimately successful and no female coordinators among the five winning awards. While we find no overt evidence of bias or discrimination, the FWF needs to take note and continue to monitor these figures in future calls. Finally, there was much need throughout the process for defining several key terms and how to treat them (novel, interdisciplinary, field, what it means to be 'emergent', etc). This extended to applicants and institutions being able to contextualise the scheme and what exactly was expected, extending to the difference or similarity to the FWF's SFBs.

The terms of reference for this study set out a long-list of evaluation criteria. We provide a direct reflection on each of these in the table below before setting out our recommendations in the final sub-section.

Table 4   Headline conclusions on each evaluation criterion

Evaluation criterion	Headline conclusion
Synopsis and proposals	
Satisfaction with requirements and timeline	The timeline of the assessment process was longer than anticipated and became too long. In future calls it is assumed that time-to-grant will be around 9-10 months, which appears appropriate. We find no other issues about timelines for submission and individual reviews, other than for a small number of cases where reviewers were identified unusually late in the process.
	The submission requirements appear disproportionate especially to applicants unsuccessful at Stage 1 as the majority of their submitted work is never assessed. However, as the alternative would likely lead to a massive increase in application influx, we judge the application requirements to be about right. The only caveat is a small share of reviewers who believe the total maximum application length could be slightly shorter, but this is a very minor point.

Evaluation criterion	Headline conclusion
Originality, innovation, and risk-related aspects of the applications	The EF call attracted a pool of highly novel research ideas, some of which were specifically developed in response to the EF call. In this respect, the EF programme has been a particular success, galvanising the Austrian research system to put forward such research ideas which likely otherwise would have remained un-tapped or been funded (potentially by funders outside Austria such as ERC) with modified, more risk-averse scope.
Quality of the proposals	All proposals deemed fully fundable without caveats at the end of Stage 3 could be funded. The quality profile by the end of the process therefore fully justified use of the budget. Several stakeholders commented that many other applications also had significant promise (e.g. those ranked 'C2' indicating the ability to fund with very minor amendments, were the budget available).
Characteristics of the proposals in terms of disciplines, inter- and transdisciplinarity	Most applicants characterised their own research as containing new interdisciplinary perspectives and several also noted that their proposed work included arts-based or transdisciplinary dimensions.
Composition of the teams (gender, academic age, institutions, disciplines etc.)	At the point of application submission, we find a good balance in terms of gender and headline discipline. The age profile is also somewhat varied though most coordinators are in the 40-49 age bracket. The diversity indicators remain stable through Stage 1 of the assessment process, however, of the five winning applications, all coordinators are male and none of the winners are from an arts/humanities background. We do not however find any evidence of overt discrimination or bias.
Review and selection process	
Objectivity and quality of the reviews as experienced by the applicants	While we find no evidence of bias, many of the applicants appear to have a poor opinion of the thoroughness and objectivity of the reviews. This applies above all to those rejected at Stage 1 of the assessment process. We judge that the main issue here is not the quality of the reviewing itself, but the transparency of the process, including processes of Jury selection and how different criteria are operationalised.
Composition of the Jury (applicants' perspectives)	The Jury (including the chair) consisted of six men and 11 women. Members are based at institutions in 11 different countries. Most are located in Europe other than three who are located in the USA and one in Australia. The UK is most strongly represented with four individuals. No institution is represented more than once.
	While the Jury is essentially suitably diverse, the uneven distribution of workload (ranging from 1 to 12 synopses reviewed per member) highlights that there is at least a slight mismatch between the Jury composition and the overall mix of the EF applications. This is a natural result of the Jury being selected prior to application submission, so there is a case in future EF calls to consider assembling or finalising the Jury at a slightly later stage, when the mix of applications is already known.
Assessment of the hearings and the relevant FWF Board Meetings	The FWF board meeting for Stage 1 had a minimal role, involving discussion only of those synopses that had received a significantly split score from Jury members. the Board focused largely on novelty of research ideas in its decisions to pass or fail. The FWF Board meeting for Stage 2 was far more substantive, with all Stage 2 applications discussed at length, and a far greater focus on scientific quality in the conventional sense. The FWF Board meeting at Stage 3 did not make any modifications or extensions to the Jury decisions. Particularly at the Stage 1 board meeting, there were many questions around the scheme design, indicating that there was some scrutiny from the board being exercised as the process took place.
	The Stage 3 Jury hearings were generally well organised and there is widespread consensus that this assessment step is necessary for a programme of this type. It was deemed especially important in order to observe the proposed teams, their interaction, collegiality and hierarchy.
Satisfaction with the timeline (all stakeholders)	As above, see 'Satisfaction with requirements and timeline'.
Quality and usefulness of different materials provided to the Jury	We do not find evidence that the materials provided to the Jury were in any form deficient. However, it is worth noting that the Jury chair decided that an informal Jury meeting was necessary in order to discuss how Stage 1 reviews should be conducted. This meeting was considered helpful by Jury members interviewed, so it should become a normal part of the process. There may also be a need to expand the briefing materials

Evaluation criterion	Headline conclusion
	for Jury members. We return to this issue when we discuss our recommendations about better defining key terms and criteria for the scheme.
Ease / difficulty in finding reviewers and Jury members	There do not appear to have been any major difficulties finding Jury members or reviewers. We note that the Jury was put together before the application deadline, and therefore did not reflect the pool of applications, leading to different levels of workload among different Jury members. in a small number of cases, there were also some difficulties finding external reviewers for Stage 2 but this does not appear to have been any more severe than what might be expected in other funding programmes.
Appropriateness of decision-making procedures and criteria	We judge the decision-making process to be appropriate. The intention was for Stage 1 to prioritise novelty, Stage 2 to prioritise scientific quality and Stage 3 to prioritise team-related aspects and in large part each stage was well-designed for these respective purposes. The criteria themselves were also appropriate, though there is a lack of clarity about how some of them should be operationalised (notably interdisciplinarity)
EF Programme	
Satisfaction with the design	Overall satisfaction levels are high in relation to design aspects. Elements such as proposal and synopsis length, available time windows, usefulness of each assessment step and other factors are consistently deemed appropriate.
Programme implementation and management	Support from FWF staff and functionality of the application submission and processing system are consistently among the most well-regarded elements of all the aspects covered by our evaluation. The EF experience showcases that FWF has an outstanding service provision, administrative support in implementation and management.
Grant preparation (coaching, information events, and consultation services)	As above, support from FWF in cases of questions or clarification requirements have met with high satisfaction levels. However, while there were two information webinars for potential applicants in September/October 2022, there were no grant preparation mechanisms assisting with application as such that we are aware of. At this relatively early stage (with the first iteration only having finished) we cannot comment on whether such activities would be necessary. The high novelty profile of applications suggest that no further pre-submission consultation services were needed as the research base evidently understood the purpose of the programme well.
Cost and efficiency of the selection process	We find no issues in terms of cost-efficiency. We note in particular the use of Stage 1 as a mechanism to limit demand and minimise the volume of external reviews required. While the Stage 3 hearings are an in-person event involving international attendees, the overall expenditure is not significant as compared with the scale of funding awarded by the programme.
Evaluation of the selection criteria	The criteria are largely appropriate. However, there was at times a lack of clarity whether or not, for instance, interdisciplinarity is a criterion to be directly rewarded. Issues around scale and significance (what is a 'field'?) as well as prior funding (what does it mean to be 'emerging'?) were also not always suitably clearly defined. Further, 'novelty' might mean very different things in different contexts. It likely required a Jury of international, multidisciplinary stature and seniority to navigate these terms.
Programme level	
Embedding in Excellent=Austria initiative	The EF scheme addresses an important ambition of the Excellent=Austria initiative in that it provides additional funding for basic research and galvanised the Austrian research base to formulate a substantial pool of highly novel research ideas, resulting in five fully fundable consortia and likely many more that may come to fruition after some modification and development.
Complementarity to relevant FWF funding tracks (SFBs, Research Groups, #ConnectingMinds)	Alongside 1000ideas (which funds much smaller grants), EF is the only other FWF tool with an explicit focus on highly novel, high-risk/high-reward research. However, the closest 'relative' within the FWF portfolio are the SFBs. In fact, there is some lack of clarity why EF and SFBs need to be fully separate. Part of the EF scope and requirements could be clarified by re-thinking EF as a special track of SFBs, sharing most characteristics, other than the modified award process and the emphasis on highly novel ideas.

5.1   Recommendations

Based on our findings, we make a small set of recommendations to consider for future EF calls. One of these is of a broader strategic nature, the remaining ones are more technical process elements.

- The additional funds provided for basic research in Austria through the EF scheme (and the Excellent=Austria initiative more broadly) is clearly much needed in stakeholders' views and has been money well spent. However, without reducing budget, there is a case for combining the EF programme into the Spezialforschungsbereiche (SFBs). This would mean much clearer 'signalling' in terms of the scale and scope expected of projects (implicit definitions around what is a 'field' already culturally existing around SFBs would then automatically apply to EF), with only the element of high novelty as a distinguishing feature. SFBs could then alternate between 'regular' and 'Emerging Fields' rounds, with the EF round maintaining the current assessment process and signalling demand for unusual, high-risk/high-reward research

- The FWF should include in the documentation of future calls (both for applicants and assessors) some definitions and illustrative examples of key terms. This could also include a distinction between actual assessment criteria and eligibility, i.e. which key terms will actually be assessed and which key terms simply signal what kinds of applications are welcome. Specifically, this should include:

    - A non-exhaustive list of definitions of what is meant by 'highly novel/innovative' or 'high-risk/high-reward'. This may helpfully include any differing definitions and understandings that might exist between natural/technical sciences, biomedicine, social sciences and arts/humanities. The winning awards of the current call may even be used as illustrative examples. Novelty should be clearly stated as a central assessment criterion

    - A definition of scale (what is a field?). This may lean heavily on definitions included in the SFBs, especially if the programmes are combined as recommended above. Minimum definition of a 'field' relates to the criterion of overall significance, which should affect decision-making

    - There may be a need to define what it means to be 'emergent', especially in relation to prior funding for proposed research ideas. In cases of substantial prior funding (e.g. ERC grants for closely related topics or strongly overlapping consortia), the FWF might want to consider that a field is no longer 'emergent'

    - For interdisciplinarity, transdisciplinarity and arts-based approaches, the documentation should emphasise that such approaches are welcome and that there will be experts able to evaluate such approaches, but that they are not in themselves assessment criteria and will not in themselves be rewarded

- While there is an administrative need to assemble most of the Jury before applications have been submitted (simply to ensure adherence to timelines), the FWF should explore options at least partially to finalise Jury composition after applications have been received so that the Jury can be better tailored to the portfolio. This might involve having a slightly larger 'roster' of potential Jury members, from which a smaller subset may then be chosen to best reflect applications' range of topics and fields

- The FWF should state in greater detail to applicants (i.e. in the guidelines) how the Jury is recruited, how allocations of Jury members to applications are undertaken, what main criteria the Jury members will emphasise in their reviews of synopses, what elements of synopses they will provide feedback on, and what kind of feedback will not be provided at Stage 1

- The Jury meeting at the start of Stage 1 should become a formal part of the assessment process. It should be led by the Jury chair though they may request the presence of FWF staff if they wish. The meeting should provide opportunity for exchange among Jury members, most notably on how to apply the A-C rating scale and how to weigh up different aspects of the synopses

- The FWF should also mandate a slightly more defined structure for synopses. While there needs to be space for discretion, there is a need to clarify whether academic references, bibliographies and previous research awards should or should not be included, and what kind of information about the proposed project team should be included

- The FWF needs to continue monitoring diversity markers in future calls. Should the inequitable outcomes be repeated in future calls, the FWF may need to consider unconscious bias training or briefings for the Jury, external reviewers and FWF Scientific Board. We do not posit such provision as a recommendation at this time because the overall numbers and supporting observations cannot confirm the existence of actual bias. The observed outcomes may at this point simply have been coincidence

Appendix A   AI-assisted analysis of applications and reviews

A.1. Quantitative indicators of scientific novelty

In this section, we describe our quantitative analysis of the programme's proposals, focusing on assessing their levels of scientific novelty. Our indicators show that the EF programme attracted highly novel proposals compared to the 'state of the art' more generally, as well as compared to FWF's core funding scheme (Principal Investigator Projects – Einzelprojekte) in particular. By 'state of the art', we refer to the highest general advancement of a scientific field achieved at the time of each proposal.

As documented in the OECD High-risk/high-reward Science Funding,[16] one central challenge science policymakers face is applicants' fear that their ideas are too ambitious and risky. Scientists' fear of following risky ideas is the initial challenge in promoting breakthrough science because, regardless of other potential sources of bias against novelty, breakthrough discoveries are unlikely to emerge without a pool of novel scientific ideas to choose from. Our results suggest that the EF programme was able to convince the scientific community to think outside of the box and submit highly novel ideas for funding.

Our indicators do not show evidence of bias against novel ideas in the initial review process, though they also do not suggest any drastic novelty appetite within the process either. Being able to attract highly novel applications does not eliminate other potential sources of bias against novelty, for example, in the project selection process. However, as most applications' novelty levels are already excellent, other obstacles become less problematic. For example, even if peer-reviewers were highly biased against choosing novel proposals for funding, funded projects would still be highly novel, given the novelty-levels in the overall pool of applications.

Some of our indicators suggest that applications that proceeded to the following round are slightly more novel than those that did not proceed, but other indicators suggest otherwise (i.e., applications that moved forward being slightly less novel). This inconclusive result, combined with the minor novelty differences between the proceeding and non-proceeding applications, indicates that this initial review round faced the challenge of selecting among highly novel proposals and did not present a bias against novel ideas.

A.1.1. Quantitative methodology

We proposed focusing on indicators of scientific novelty, which are particularly well-suited to capture the essence of risk and reward the EF programme aims to promote. Scientific breakthroughs significantly advancing the knowledge frontier are typically extraordinarily novel compared to prior art. Thus, high novelty relates to high reward. However, despite the potential for radical impact, novel ideas face more uncertainty as scientists attempt to explore untested processes and face delayed recognition or resistance from incumbent paradigms. Machado (2021) presents evidence of this effect, showing that highly novel articles have, on average, higher citation impact but also significantly higher levels of variation by covering a more comprehensive range of both low and highly cited articles.[17] Consequently, novelty relates both to high-risk and high-reward, representing the nature of the EF programme.

We computed three types of novelty indicators exploring the proposals' text, including:

---

1. **Unprecedented combinations of journal references (journal pairs):** Exploring journal references is the most common approach to measuring novelty [see e.g. Uzzi et al. (2013), Wang et al. (2017) or Veugelers et al. (2022)].[18] Journals typically represent scientific knowledge spaces. Thus, if articles combine pairs of journals never combined before in their reference lists, such papers will likely make novel combinations of scientific ideas. We apply this strategy by exploring proposals' references. Computing this indicator required 1) extracting the proposals' reference list, 2) identifying and matching journals to our bibliometric repository, and 3) counting the number of journal pairs that never occurred in the state-of-the-art. For this indicator, we considered state-of-the-art as all the articles referenced in each EF proposal. Next, we augmented this pool with scientific publications that are semantically similar to the references (and published before the outline proposal) and the references of these semantically related publications. Therefore, each outline has its own proxy for relevant state-of-the-art.

2. **Unprecedented combinations of concepts (concept pairs):** Journals are imperfect representations of knowledge spaces because new journals often derive from publishers' business decisions and not from the natural dynamics of scientific evolution. We address this caveat by identifying a list of core concepts/ideas in the proposals' text and assessing how many were never combined in the prior art. Computing this indicator required reading the text of the outlines and applying the technique of "Named entity recognition" (a Natural Language Processing operation), which consists of identifying concepts. We only consider concepts that occurred at least once in our bibliometric database of past titles and abstracts (including not only articles but also, e.g., conference proceedings, reviews or books. Finally, the indicator consists of counting the number of times pairs of concepts never occurred in the state-of-the-art / occurred for the first time in the text of each outline. For this indicator, we consider the state-of-the-art to be all the scientific publications in the Open-Alex bibliometric database published before the outline proposal.

3. **Semantic dissimilarity:** The previous indicators do not distinguish between the meaning of different journals or concepts in an accurate approach. Consider the terms "Deep learning" and "artificial neural networks". Deep learning relies on artificial neural networks and so these concepts are practically synonyms. However, indicators based on unprecedented combinations ignore that reality. For example, if a scientist combines "deep learning" and "immunotherapy" for the first time and later another scientist uses the terms "artificial neural networks" and "immunotherapy", the indicator will count both as equally novel. The same applies for journals that represent extremely similar knowledge spaces. We propose an indicator that assesses "scientific meaning" to address this shortcoming. The text of highly novel ideas combines textual references of concepts that prior art articles never connected. Therefore, the text of highly novel ideas is more semantically dissimilar than prior art. We measure proposals' semantic meaning by applying a recent development in Natural Language Processing that transforms text into a mathematical representation (document/word embeddings).[19] In this mathematical representation, vectors capture the meaning of documents, and we can apply mathematical operations over such vectors, including the calculation of distances. We

[18] Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. Science, 342(6157), 468-472. doi:10.1126/science.1240474; Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy. doi:10.1016/j.respol.2017.06.006; Veugelers, R., Wang, J., & Stephan, P. (2022). Do funding agencies select and enable risky research: Evidence from ERC using novelty as a proxy of risk taking (No. w30320). National Bureau of Economic Research.

[19] We use the deep learning framework "BERT" (Bidirectional Encoder Representations from Transformers) and the pre-trained embeddings model SciBert trained on the corpus of semanticscholar.org.

compute a common (dis)similarity metric – cosine distance – to measure the dissimilarity between each proposal and its state-of-the-art. The dissimilarity metrics consist of the maximum distances between these vectors (focal proposals and prior art) and variation measurements, e.g., variance, standard deviation, range and interquartile range. We consider that the larger the maximum distance and variation between proposals' vectors and the vectors of their prior art, the more novel the focal proposals are. For each outline, we construct each pool of state-of-the-art as in indicator 1.

In addition to computing novelty indicators for EF outlines, we also measure the novelty of submissions to other FWF schemes – the core funding Einzelprojekte scheme – as a benchmark. For each EF outline, the benchmark consists of the ten most semantically similar Einzelprojekte proposals, and we compute all the novelty indicators (unprecedented combinations of journals and concepts and semantic dissimilarity) for all these benchmark submissions. Next, we map the novelty scores of each EF outline against their own benchmark. Both the outlines and benchmark Einzelprojekte proposals have their own specific proxies for state-of-the-art as specified in the description of the indicators. The state-of-the-art differs by submission dates and uses scientific publications from the OpenAlex database.

A.1.2. Results

All the indicators under analysis point in the direction that EF proposals present high levels of scientific novelty.

Regarding journal references (indicator 1), EF outlines present, on average, 105 novel combinations, whereas the respective benchmark for each submission averages 38 (Figure 19). As for concepts, EF outlines combine, on average, seven unprecedented novel pairs, whereas their respective benchmarks present four new concept pairs (Figure 20). In all figures we denote EF outlines as focal proposals.

The differences between rejected outlines and those advancing to stage two are not substantial. Approved submissions combine, on average, 119 unprecedented journal pairs; rejected incorporate 93 (see Table 5). Both groups present higher average levels against their benchmarks (80 and 55 more combinations, respectively). In contrast, novel term combinations (indicator 2) depict the opposite tendency. Approved submissions combine six novel concept pairs, while rejected combine 8. Still, both levels present higher averages than their benchmarks (four and two more, respectively).

Please note, to preserve anonymity of applicants, we have randomised the order of applications and assigned them identifiers following the style 'randXX'. These new identifiers do not relate to the original application identifiers, i.e. 'rand01' is <u>not</u> application number EFP-01. We use this convention throughout the graphs below.

Figure 19 Novel combinations of journal references
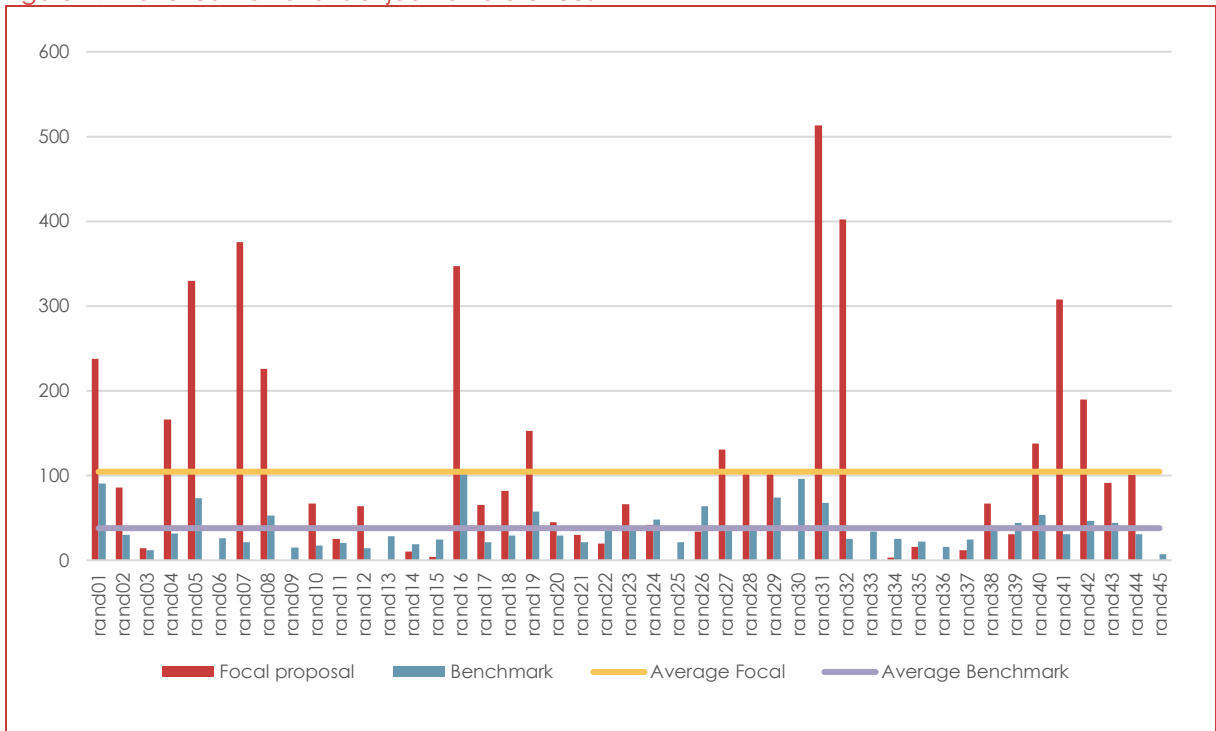
Figure 20 Novel concepts pair combinations
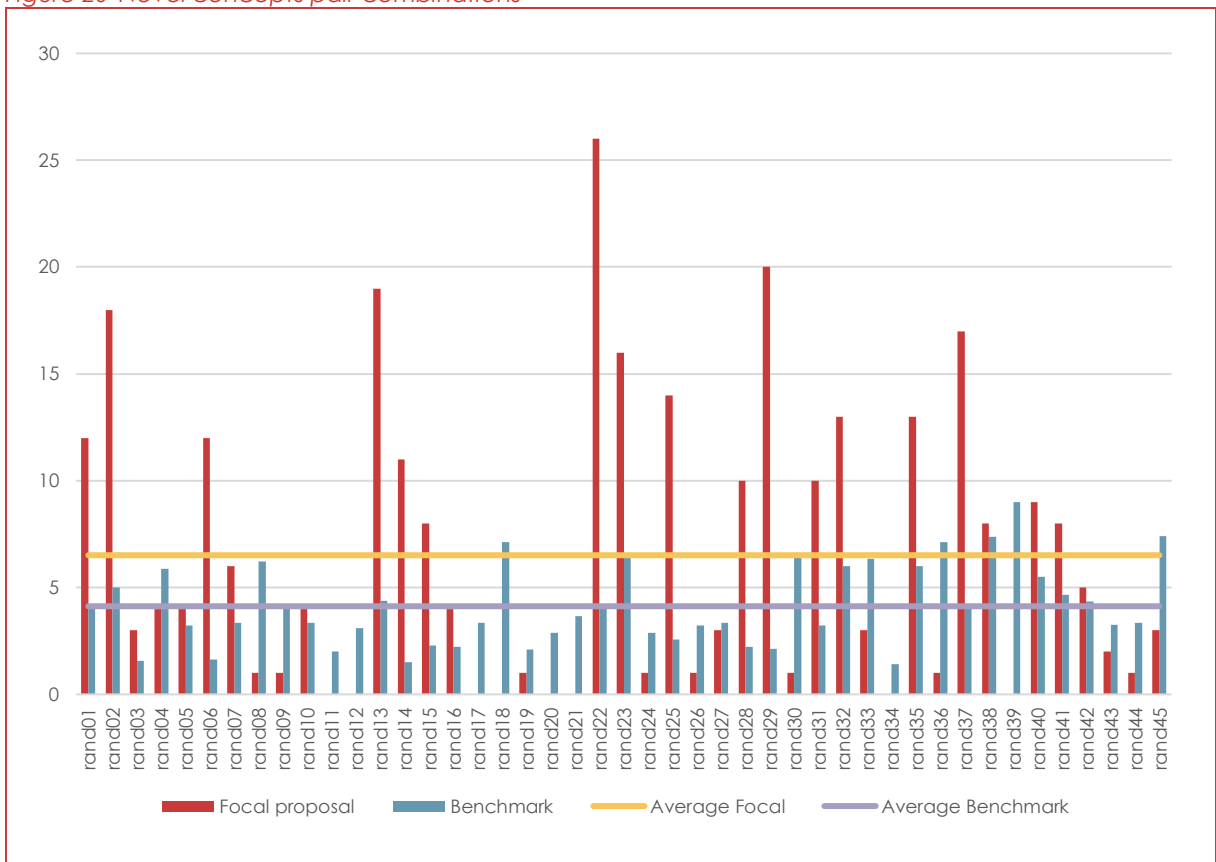
Table 5   Breakdown of novel concept and journal pairs per decision type (rounded value)

Decision	Novel journal pairs	Journal differences between focal and benchmark	Novel concept pairs	Concepts differences between focal and benchmark
Rejection	93	55	8	4
Approval Stage 2	119	80	6	2
Return without review (Reassumierungsantrag nach der 92 KS wurde abgelehnt)	78	49	1	-2
Return without review	10	-9	11	10

All the metrics used to reflect semantic dissimilarity present results in the same direction as novel journal and concept pairs. Focal proposals systematically present higher levels of maximum semantic distances than prior art, higher variance, standard deviation and range (full and interquartile). Regarding the breakdown per decision type, projects that proceeded to stage two present slightly lower values. However, the differences for both groups with the respective benchmarks are positive, meaning that both groups present higher levels of novelty than their prior art. The following figures and table illustrate these results.

Figure 21 Semantic distance: Max
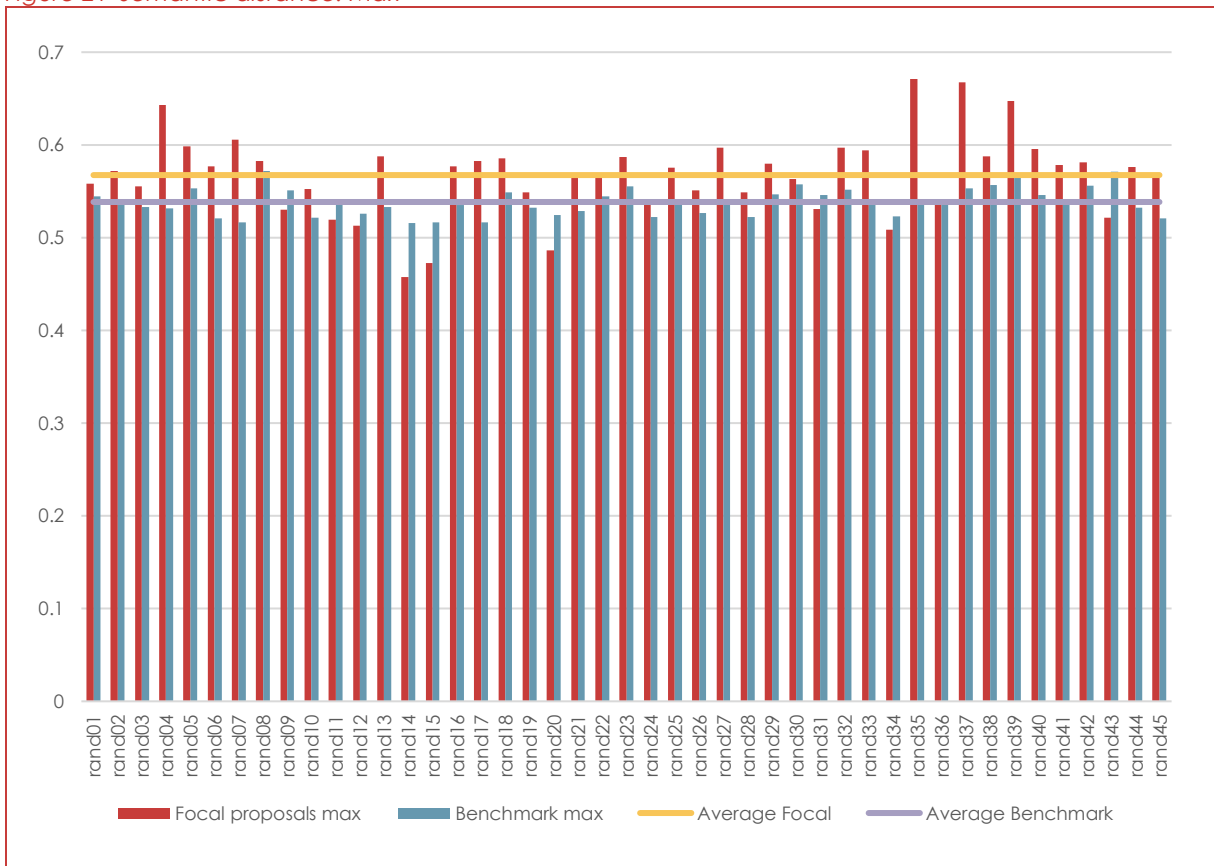
Figure 22 Semantic distance: Variance
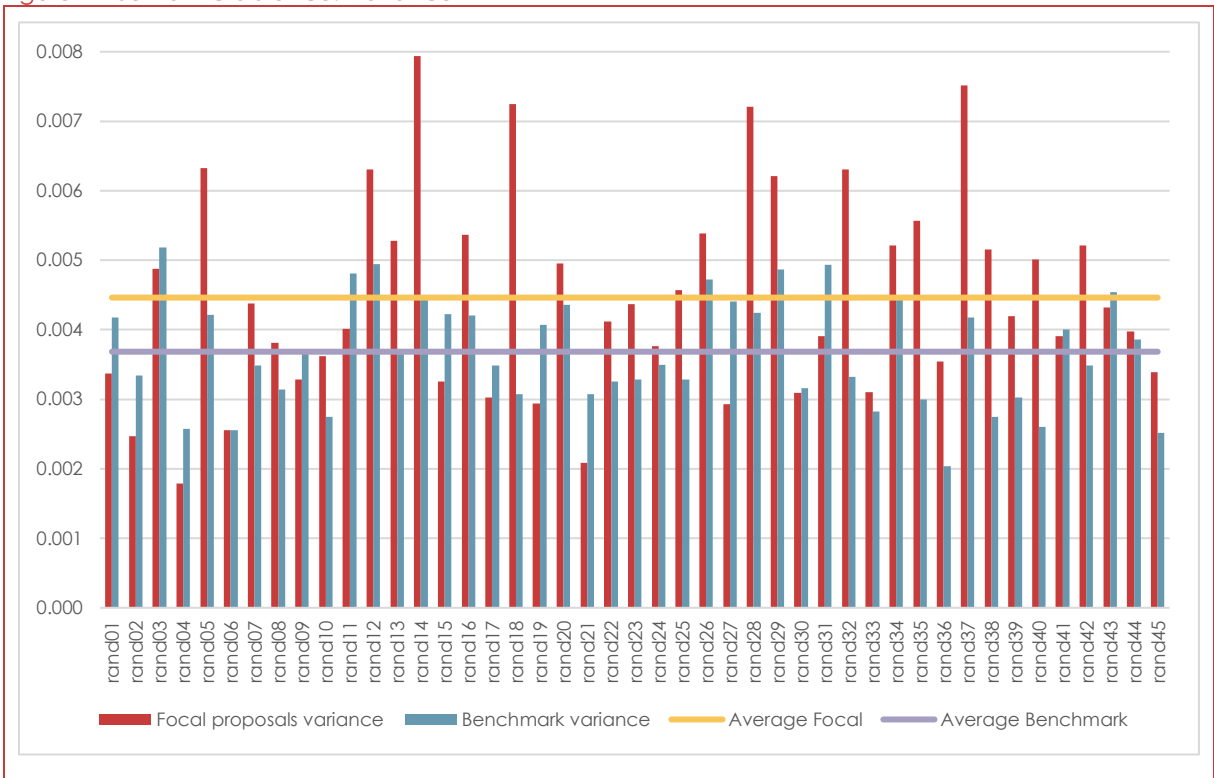
Figure 23 Semantic distance: Standard Deviation
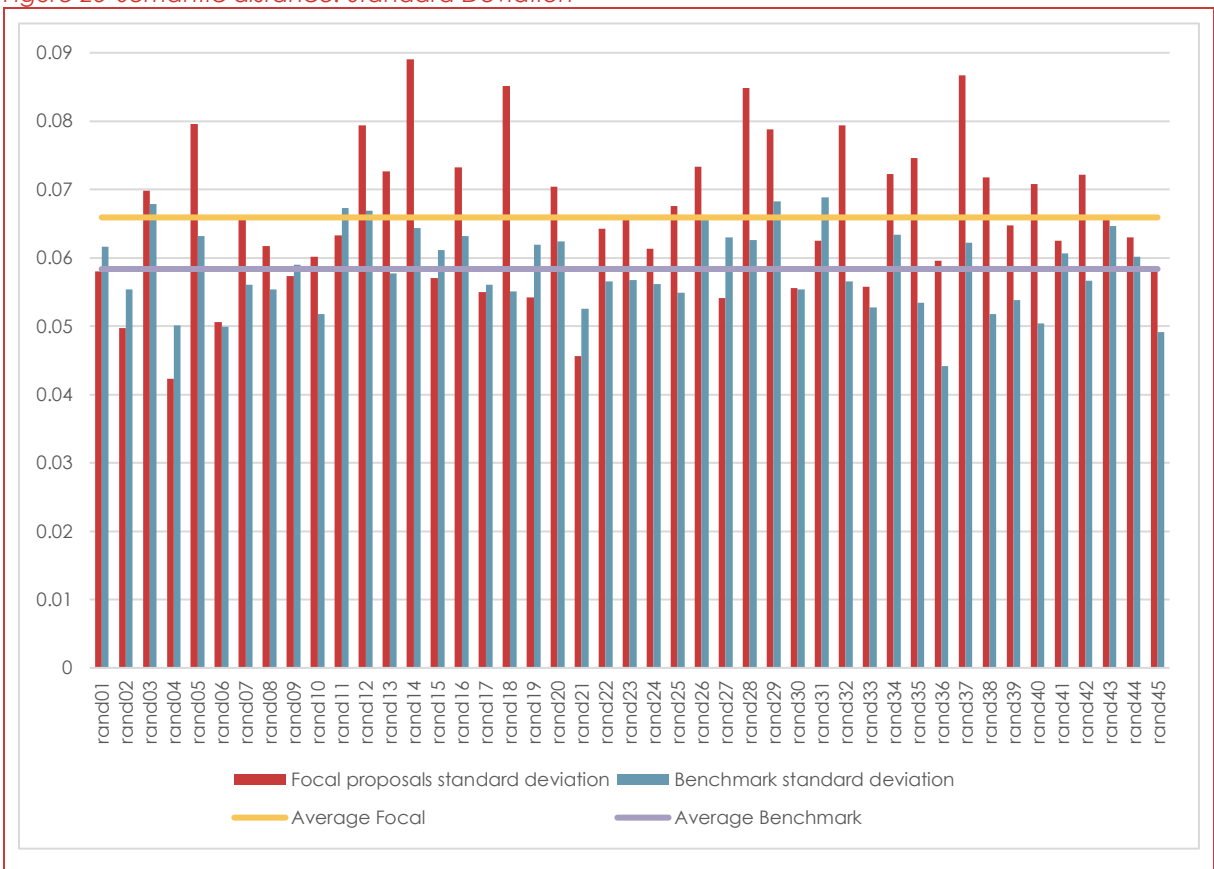
Figure 24 Semantic distance: Total range
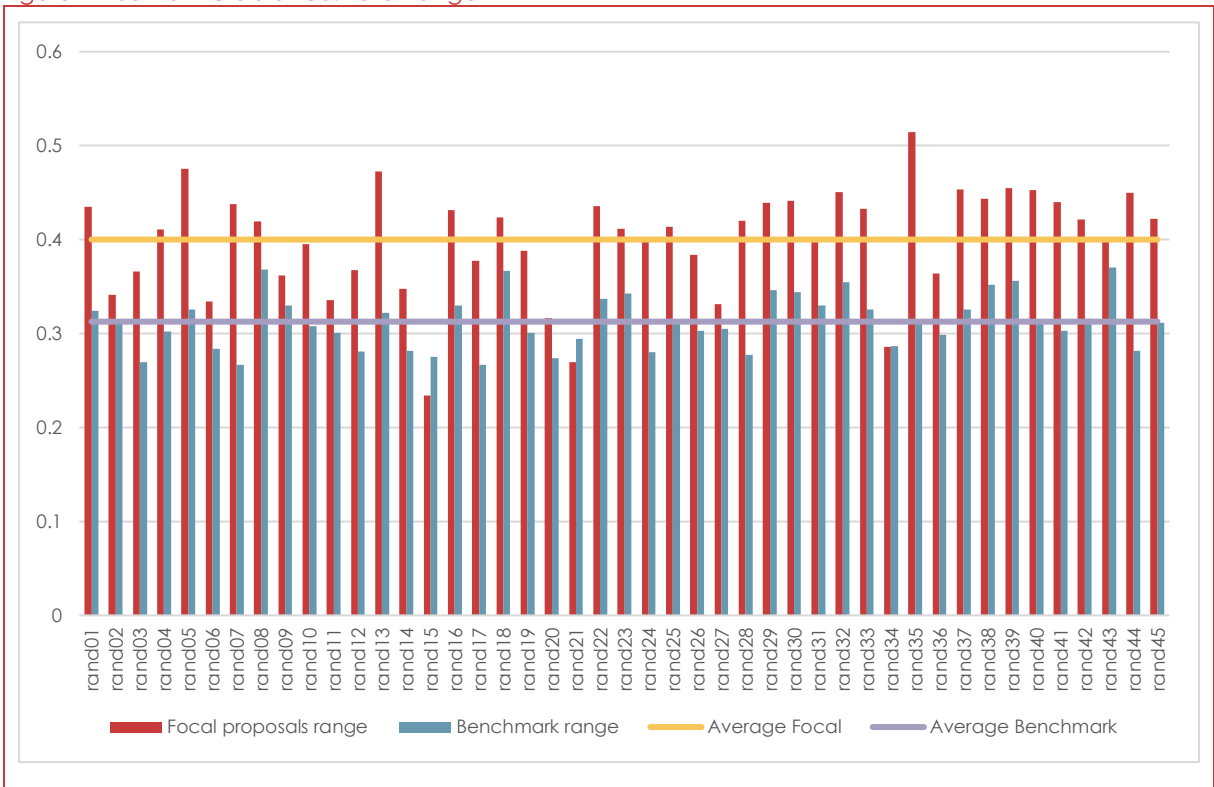
Figure 25 Semantic distance: Interquartile Range
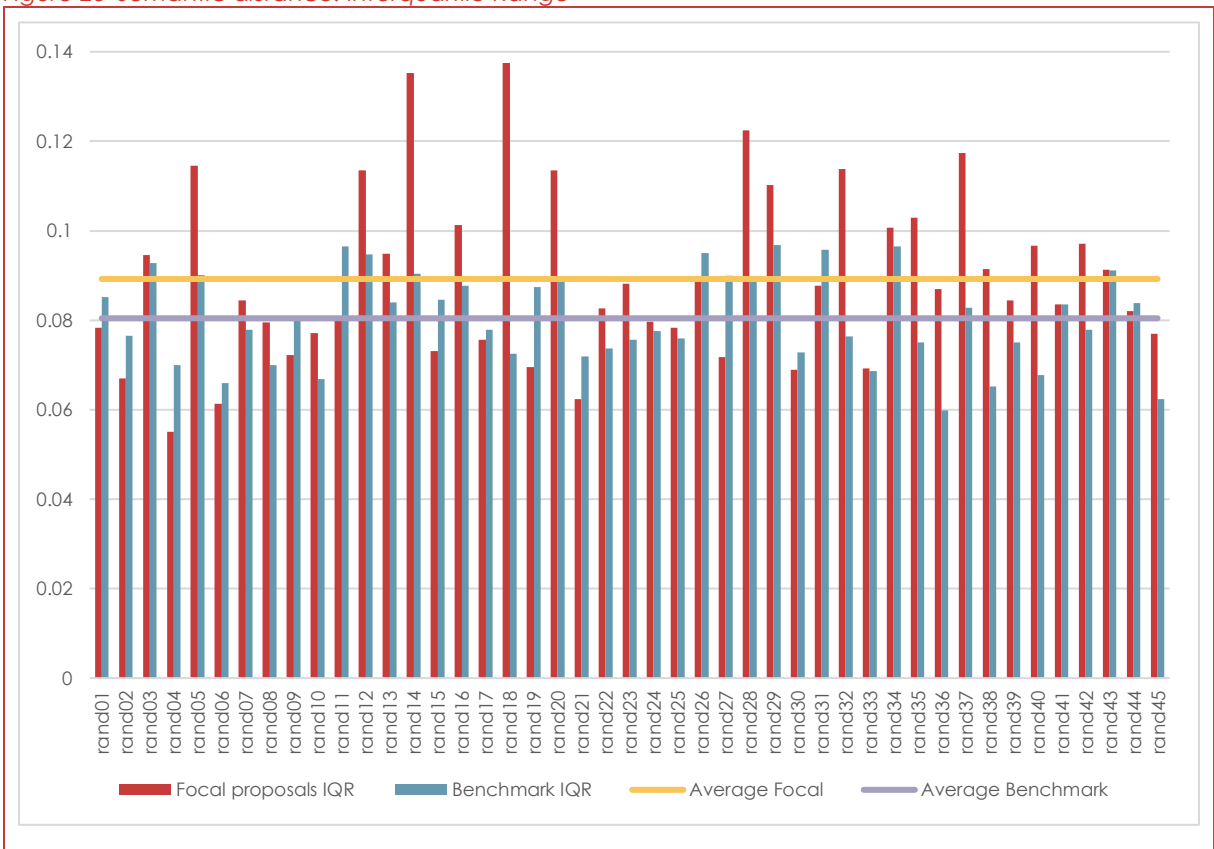
Table 6   Differences between focal and benchmark dissimilarity per decision type

Decision 93. KS	Max	Variance	Standard Deviation	Total range	IQR range
rejection	0,0397	0,0011	0,0098	0,0944	0,0133
approval Stage 2	0,0297	0,0004	0,0052	0,0854	0,0039
return without review (Reassumierungsantrag nach der 92 KS wurde abgelehnt)	-0,0316	0,0006	0,0068	0,0584	0,0095
return without review	-0,0585	0,0034	0,0247	0,0658	0,0449

A.2. Using Generative AI to assess peer-review evaluation documents

This section describes our use of generative AI to uncover insights about the EF project selection process, considering the dimensions of scientific novelty, risk, quality and team qualifications. Prominent international scientists reviewed the synopses and proposals for the EF programme and produced evaluation documents detailing the peer-review exercises. In total, we had access to 156 records, 87 peer-reviews of synopses and 69 reviews of full proposals.[20] The peer-reviewers detail their views about the potential of the scientific ideas from the synopses and proposals in terms of their levels of novelty, risk, scientific quality and suitability of the project team to undertake the research. Therefore, these documents are a rich source of information about the scientific potential of the forthcoming EF-funded projects, but also about the dimensions receiving more attention at the different stages of the selection process.

Given the number, heterogeneity and complexity of the review documents, we found generative AI particularly useful in facilitating a systematic assessment. The multidisciplinary nature of the EF programme and the fact that EF applications are at the frontier of science make the review documents far from digestible for a general audience. They are heavy on scientific jargon and technical details, which are hard to understand for someone without deep expertise in each topic. We used a rich and comprehensive generative AI model that managed to navigate these technical details to find relevant individual insights about each review document and stylised facts about the selection process in its two stages (synopsis and proposal steps). The model was OPENAI's latest GPT4 large language model accessed programmatically via a dedicated API. This access mode ensures the privacy and confidentiality of the underlying data. Also, this mode of access enabled us to explore the capabilities of the GPT4 model in large-scale automation (querying all the documents programmatically without manually inputting and querying each review document individually). For more information about our approach, check the following box.

Generative AI approach and prompt engineering

Generative AI is an advanced form of autocomplete. A large language model trained on large amounts of text predicts the words that follow a particular input text based on its trained data. When the input text consists of a question, the expected prediction or output from the model should be an answer. We explore this question-and-answer capability to systematically assess the review documents by asking the same questions to all documents and expecting answers following a similar structure.

Hallucinations are one of the main concerns when using large language models. These consist of model predictions that are wrong or highly biased, even if exceptionally well articulated. No silver bullet exists to avoid hallucinations completely, just strategies to 1) allow detecting them more transparently with answers that include an explanation, 2) limit the model's knowledge to the provided input, and 3) require justifications linked with the input text. We proceeded with these strategies in a small pilot. We iteratively tested different variations of our questions on a random sample of documents (this being the prompt engineering process), until reaching robust results.

The following are the prompts we used in this exercise:

- **System prompt (included in all prompts to improve general understanding of the task and format of the input):**
  - "role": "system", "content": "Users will provide you text from scientific reviews. The reviews have a common section template and different answers for each section. Answer questions only using information from those reviews. Expect users to use the following format:\nReview:|||text from the review here|||\nQuestion:|||user question here|||"
- **Sentiment prompts:**

- **Is the reviewer positive, negative or neutral regarding the proposal's novelty? Answer in the following format: Positive/Negative/Neutral/Not discussed. Explanation (max 100 words and always include quotes from the reviewer's text).**
  - Is the reviewer positive, negative or neutral regarding the proposal's risk/feasibility? Answer in the following format: Positive/Negative/Neutral/Not discussed. Explanation (max 100 words and always include quotes from the reviewer's text).
  - Is the reviewer positive, negative or neutral regarding the proposal's Scientific quality/rigour? Answer in the following format: Positive/Negative/Neutral/Not discussed. Explanation (max 100 words and always include quotes from the reviewer's text).
  - Is the reviewer positive, negative or neutral regarding the qualifications/suitability of the team? Answer in the following format: Positive/Negative/Neutral/Not discussed. Explanation (max 100 words and always include quotes from the reviewer's text).
- **Priority ranking prompt:**
  - Sort the following topics from the most important to the least important in terms of number of words dedicated to each topic by the reviewer and provide an estimate of the total number of words dedicated to each topic. The topics are A. Novelty, B. Risk/feasibility, C. Scientific quality/rigour, D. Qualifications/suitability of the team. Your answer should follow strictly the format of the following example of output in brackets:[A/C/B/D (104/51/50/20)]. In the example, the letters provide the rank of topics from most to least important and in parenthesis the estimated number of words dedicated to each topic. Do not add any additional text, just the output in brackets.

Before deploying the generative AI tool, we started by triangulating results from previous tasks of this project to derive the central hypothesis for testing. The programme's fundamental goal is to attract and select high-risk and highly novel ideas for funding. In our quantitative analysis, we produced novelty indicators suggesting that, overall, applications to the EF programme reveal high levels of novelty when compared with other FWF programmes. These insights suggest that the EF programme attracted a worthy pool of novel ideas. Regarding selection, the EF programme implemented a multi-step process, which differs from other FWF programmes. From our interviews and review of existing programme documents, we understand that this procedural innovation aimed to incentivise peer reviewers to focus predominantly on the novelty potential in initial steps and more on scientific quality in later steps. The synopsis review stage potentially enables reviewers to focus more on the creative and out-of-the-box content of the scientific ideas instead of being immediately caught by the risks, technical hurdles and caveats that most ground-breaking discoveries have to "breakthrough". Still, we should expect that reviewers do their job in recommending the ideas with the highest overall potential for funding, ensuring that risky ideas get their chance but still undergo a rigorous check from a scientific quality perspective.

Based on the triangulation of results, our hypotheses for testing were the following:

- Successful synopses (those approved to Stage 2 and full proposal review) receive more positive comments from reviewers across all dimensions when compared with unsuccessful ones
- Successful proposals (those invited to proceed) receive more positive comments from reviewers across all dimensions when compared with unsuccessful ones
- At the synopsis review stage, reviewers give more prominence to novelty and risk than scientific quality, regardless of a positive or negative sentiment
- At the proposal review stage, reviewers give more prominence to scientific quality, regardless of a positive or negative sentiment

To test these hypotheses, we performed topic detection, sentiment and priority detection analyses, exploring the text of the review documents with generative AI. The topic detection

consisted of distinguishing the parts of the reviews focusing on novelty, risk, scientific quality and team suitability dimensions. The sentiment analysis assessed whether the reviewers were positive, negative or neutral on their considerations about each topic. Finally, the priority detection measured the text length or number of words dedicated to each topic as a proxy for how much effort/time the reviewers devoted to each dimension. Next, we transformed the topic-level sentiment and priority indicators into quantitative scores. For the sentiment variables, we generate one sentiment variable per topic/dimension and for each document. Each variable is equal to one when the reviewer is positive about that dimension and zero otherwise. For the priority variables, we generated a rank per document, where the dimension with the highest prominence receives the value of 3, followed by the next dimensions with the values of 2, 1 and zero, respectively.

Our results provide suggestive evidence in favour of all the hypotheses. Overall, reviewers were more positive about all dimensions for successful applications (synopses approved to submit the full proposal and proposals invited to the next stage). On average, successful synopses have a positive sentiment of 0.72 (the range being 0 to 1), whereas unsuccessful scored 0.34. Equally, the sentiment of the successful synopses was higher in 0.31, 0.43 and 0.28 scores than the unsuccessful for team, quality and risk, respectively (see Figure 26). The patterns between successful and unsuccessful full proposals were similar (Figure 27).

Figure 26 Average sentiment scores for synopses' evaluation
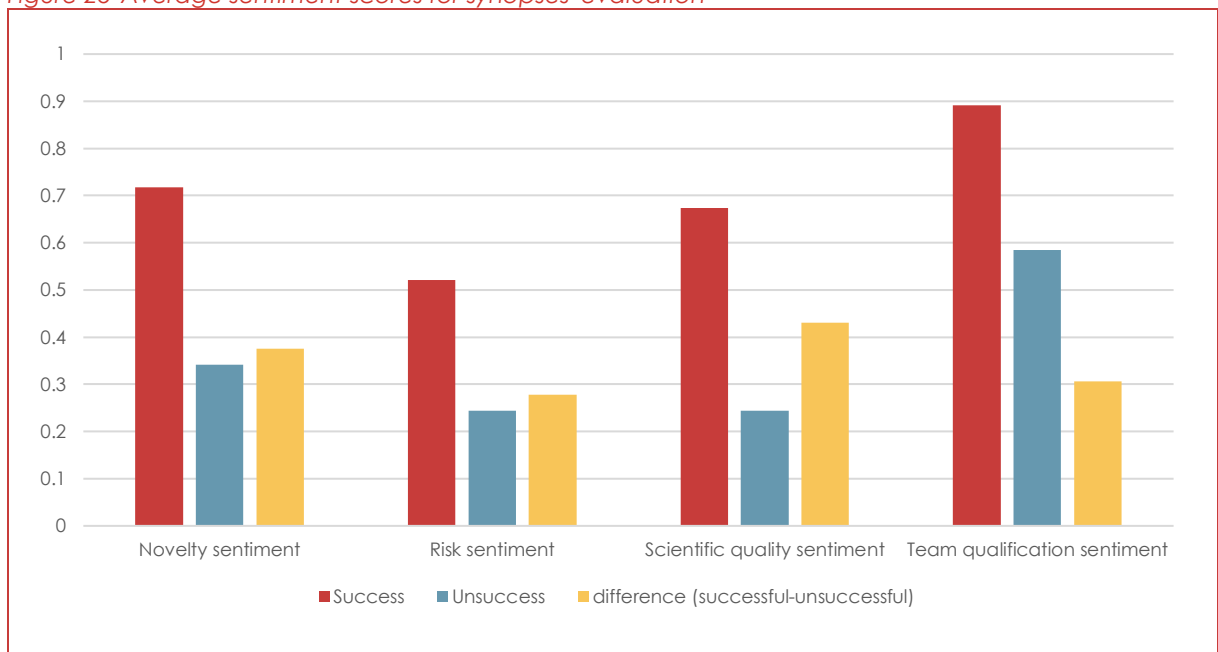
Source: Technopolis-Group analysis based on peer-review documents

Figure 27 Average sentiment scores for proposals' evaluation
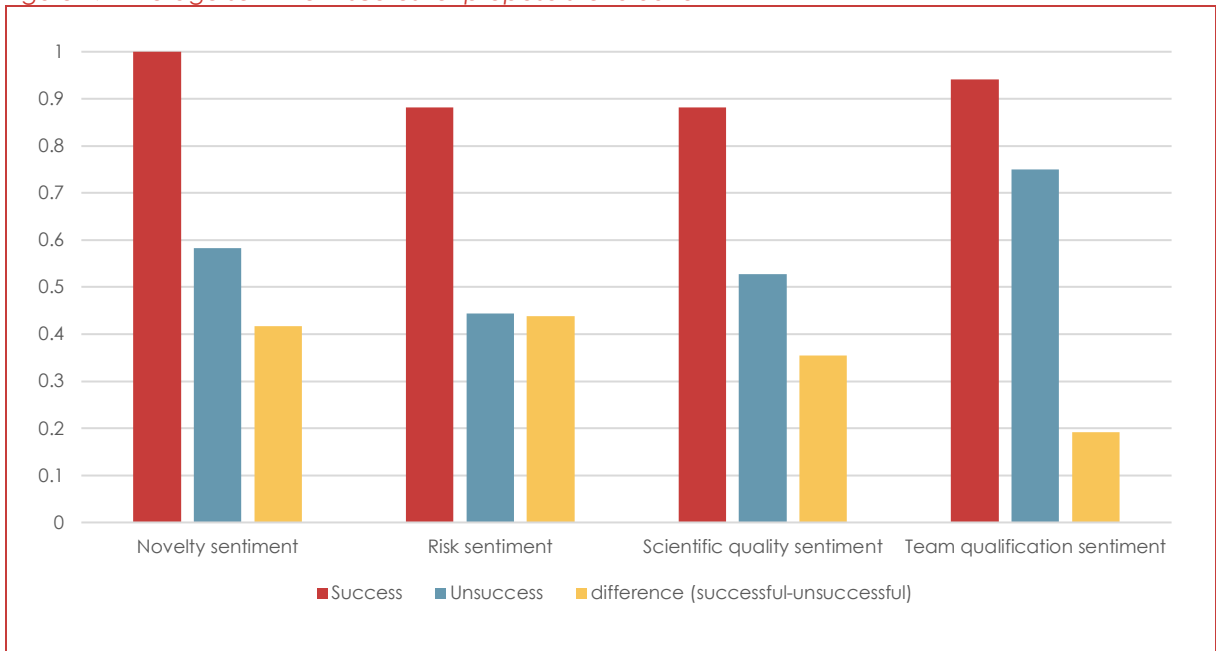
Source: Technopolis-Group analysis based on peer-review documents

The analysis of priorities confirms that at the synopses review stage, reviewers focus more on the novelty dimension and less on considerations about scientific quality. The ranking indicator ranges from 0 to 3, with novelty scoring, on average, 2.26 for synopses and only 1.3 for proposals. The risk and team dimensions also receive more attention at the synopses stage than proposals, but with less pronounced differences. The differences in terms of dedication to reviewing scientific quality are more noticeable. This dimension has the lowest rank at the synopsis stage, scoring on average 0.61. In contrast, scientific quality considerations rank the highest in the proposal evaluation stage, scoring 2.17 (see Figure 28).

Figure 28 Average priority ranking scores for synopses and proposals
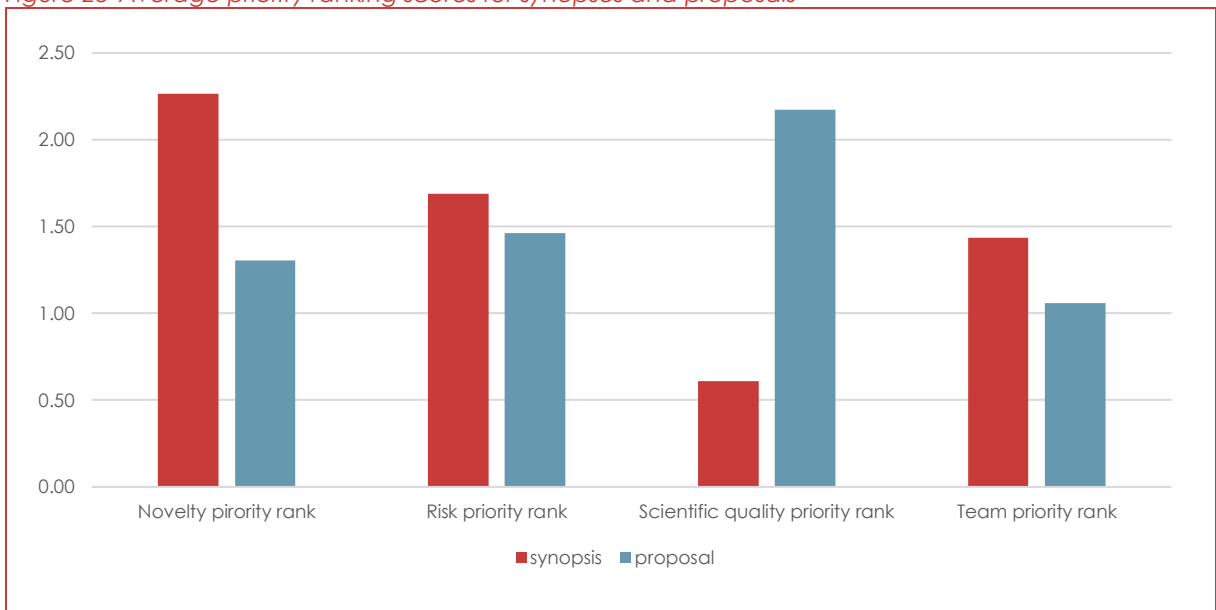
When assessing the breakdown of priority ranks per successful and unsuccessful applications, the pattern remains unchanged regarding differences in focus (see Figure 29). However, the breakdown reveals that in the case of successful proposals, reviewers dedicate, on average, slightly more effort to novelty considerations in comparison to unsuccessful applications. In contrast, for unsuccessful applications, reviewers allocate more effort to review scientific considerations, potentially with a more comprehensive review of existing pitfalls.

Figure 29 Average priority ranking scores for successful/unsuccessful synopses and proposals
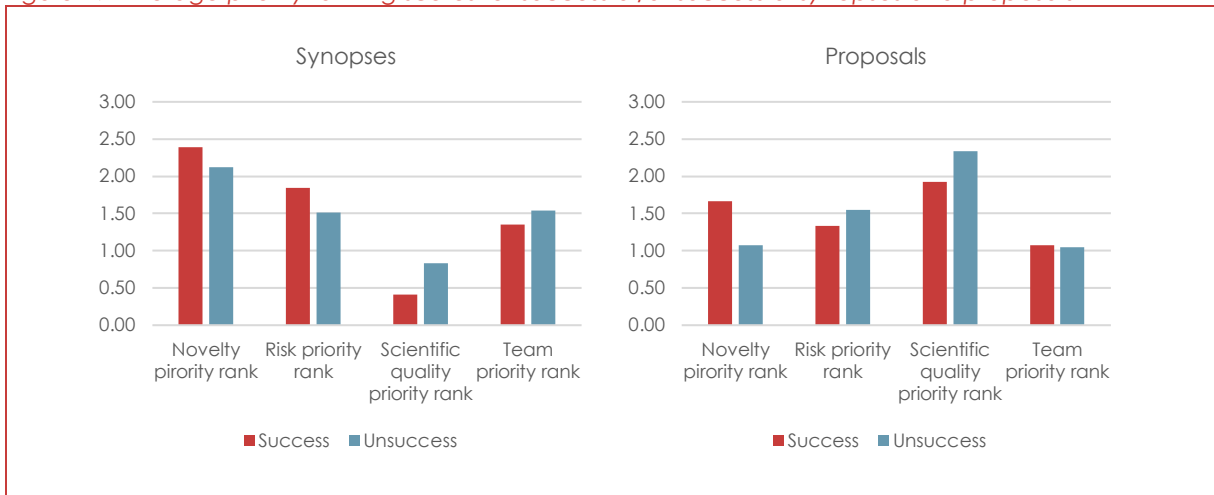
Technopolis-Group analysis based on peer-review documents

Appendix B   Survey details

B.1. Survey of applicants

B.1.1. Response rate details

Table 7   Survey of EF applicants – headline response rate

Total Population:	45	Survey responses:	37	Response rate:	82%
Population notes: Population includes the named coordinators on all 45 applications submitted to the 2022/23 Emerging Fields call, regardless of application outcome. 2 invitations bounced and no respondents had opted out of receiving surveys, so 43 could be invited to take the survey.		Response notes: Responses were collected between 16/01/2024 and 14/02/2024, Involving one initial invite and three reminders (see graph below). Stages 1 and 2 of the EF application process had been completed by the time of the survey launch and all Stage 2 applicants had been informed whether or not they were invited to a Stage 3 hearing.		Response rate notes: For this population size, we cannot meaningfully denote a confidence interval or margin of error. However, the response rate is very high and largely representative along lines of all characteristics we can control for. This means that our results should provide a strongly indicative reflection of reality.	

Characteristic	Total population (N=45)		Responses (n=37)	
FWF top-level field classification				
HumSoc	20	44%	16	43%
NatTec	14	31%	12	32%
BioMed	11	24%	9	24%
Application status				
Rejected at Stage 1 without review	3	7%	2	5%
Invited to hearing	10	22%	10	27%
Rejected at Stage 2	14	31%	11	30%
Rejected at Stage 1	18	40%	14	38%

Figure 30 Survey of applicants – Timeline of response submissions
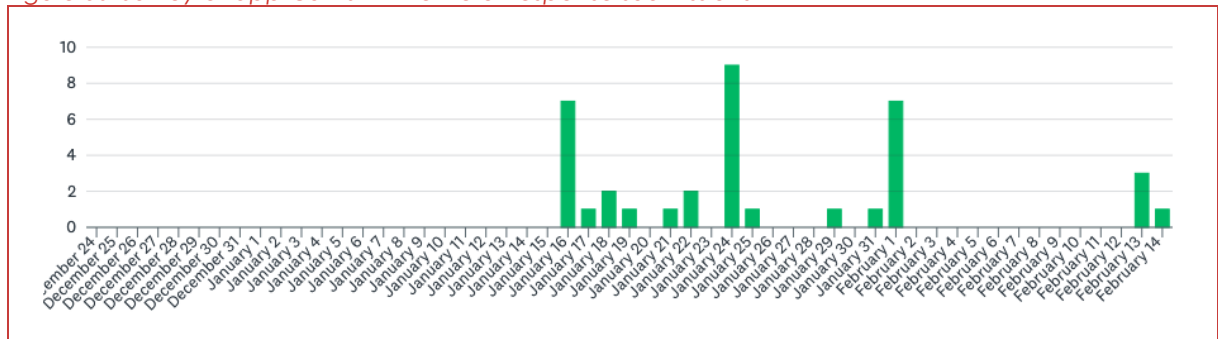
Source: Surveymonkey.

B.1.2. Raw response data

What is your gender?

Answer Choices	Responses	
Woman	45.95%	17
Man	48.65%	18
Prefer not to say	5.41%	2
Non-binary/other	0.00%	0
	Answered	**37**
	Skipped	**0**

What is your age?

Answer Choices	Responses	
29 or under	2.70%	1
30-39	5.41%	2
40-49	48.65%	18
50-59	32.43%	12
60 or over	5.41%	2
Prefer not to say	5.41%	2
	Answered	**37**
	Skipped	**0**

With which of the following areas of science is your work most closely aligned?

Answer Choices	Responses	
Biomedical sciences	18.92%	7
Natural sciences and engineering	40.54%	15
Social sciences, humanities and arts	32.43%	12
Other/cannot say	8.11%	3
	Answered	**37**
	Skipped	**0**

Which of the following best describes your Emerging Fields project idea prior to the Emerging Fields call launch? (Please tick one)

Answer Choices	Responses	
My project idea was entirely new: it was motivated by the Emerging Fields call	45.95%	17
My project idea existed before the Emerging Fields call but I had not applied for funding for this idea	35.14%	13
My project idea existed before the Emerging Fields call. I had applied for funding for this idea from sources other than the Emerging Fields scheme but had been unsuccessful	8.11%	3
My project idea existed before the Emerging Fields call. I had done research work on it, for which I had received some prior research funding (less than €250,000)	8.11%	3
My project idea existed before the Emerging Fields call. I had done research work on it, for which I had received substantial prior research funding (€250,000 or more)	0.00%	0
Other / none of the above (please specify)	2.70%	1
	Answered	**37**

	Skipped	0

If you had applied for funding for your project idea prior to applying to the Emerging Fields scheme, please select which of the following sources you had applied to and whether you were successful or not. (Please tick all that apply)

Answer Choices	Responses	
Not applicable: I had not applied for any prior funding for my project idea before applying to Emerging Fields	72.00%	18
I had applied to another FWF programme, successfully	8.00%	2
I had applied to another FWF programme, but was unsuccessful	0.00%	0
I had applied to an Austrian funder other than FWF, successfully	4.00%	1
I had applied to an Austrian funder other than FWF, but was unsuccessful	0.00%	0
I had applied for internal funding at my institution, successfully	0.00%	0
I had applied for internal funding at my institution, but was unsuccessful	4.00%	1
I had applied to a funder from outside Austria (including EU Horizon/ERC), successfully	8.00%	2
I had applied to a funder from outside Austria (including EU Horizon/ERC), but was unsuccessful	8.00%	2
Other successful or unsuccessful funding requests for your project idea prior to Emerging Fields application (please specify)	0.00%	0
Other (please specify)		0
	Answered	25
	Skipped	12

Why did you apply to the Emerging Fields programme? Please indicate whether each of the following factors were a major reason, a minor reason or not a reason at all for your decision to apply.

	Not a reason		A minor reason		A major reason		Don't know / cannot say		Total
I wanted to conduct more unconventional/high risk-high reward research than I had done previously	0.00%	0	18.92%	7	78.38%	29	2.70%	1	37
It was an opportunity to address interdisciplinary research ideas	0.00%	0	8.11%	3	91.89%	34	0.00%	0	37
It was an opportunity to address arts-based research ideas	69.44%	25	0.00%	0	27.78%	10	2.78%	1	36
It was an opportunity to address transdisciplinary approaches (involving non-academic participants from outside the scientific community)	47.22%	17	27.78%	10	22.22%	8	2.78%	1	36
My project idea would have had lower success chances on other available schemes	16.67%	6	27.78%	10	41.67%	15	13.89%	5	36
There are limited opportunities for funding of novel, high-risk research ideas	8.33%	3	16.67%	6	75.00%	27	0.00%	0	36
My project idea required a lot of resources and there are limited funding opportunities of this size	5.41%	2	35.14%	13	59.46%	22	0.00%	0	37
I wanted to build a collaborative team of researchers	0.00%	0	5.41%	2	94.59%	35	0.00%	0	37
Winning an Emerging Fields award would mean prestige and/or career advancement	22.22%	8	41.67%	15	33.33%	12	2.78%	1	36
Other major reasons, please specify:									1
Answered									**37**
Skipped									**0**
Item included an additional freetext-box: "Feel free to comment on any of the aspects above, or any other administrative aspects around the review process, particularly if you had any noteworthy positive or negative experiences:" [answers redacted to protect anonymity]									**1**

To what extent do you agree with each of the following statements about your proposed project?

	Strongly disagree		Somewhat disagree		Neither agree nor disagree		Somewhat agree		Strongly agree		Don't know / cannot say		Total
My project presented novel and/or unconventional ideas, perspectives and methods	0.00%	0	0.00%	0	0.00%	0	2.70%	1	97.30%	36	0.00%	0	37
My project presented new interdisciplinary perspectives	0.00%	0	0.00%	0	0.00%	0	16.22%	6	83.78%	31	0.00%	0	37
My project presented new arts-based perspectives	52.78%	19	2.78%	1	8.33%	3	8.33%	3	19.44%	7	8.33%	3	36

My project presented new transdisciplinary approaches (involving non-academic participants from outside the scientific community)	22.22%	8	19.44%	7	0.00%	0	33.33%	12	16.67%	6	8.33%	3	36
My project challenged core assumptions of the research establishment in my field	0.00%	0	5.41%	2	10.81%	4	27.03%	10	56.76%	21	0.00%	0	37
My project presented a risky idea with significant chance of failure	2.70%	1	10.81%	4	21.62%	8	27.03%	10	37.84%	14	0.00%	0	37
My project would have struggled to be accepted by funding programmes for conventional research	2.78%	1	5.56%	2	13.89%	5	44.44%	16	30.56%	11	2.78%	1	36
Answered													**37**
Skipped													**0**

The first stage of the assessment process was based on the short synopsis of your application. What is your view on the permitted 3-page maximum length of the synopsis?		
Answer Choices	Responses	
The maximum synopsis length was too short to suitably describe the central idea(s) of the project and it should be made longer	24.32%	9
The maximum synopsis length was of roughly the right length to crystallise the central idea(s) of the project	72.97%	27
The maximum synopsis length was too long: an even shorter format to suitably describe the central idea(s) of the project would have been better	2.70%	1
Don't know / no opinion	0.00%	0
Answered		**37**
Skipped		**0**

Do you think short proposal summaries (synopses) for the first stage of the assessment process should be expanded to other research funding schemes?		
Answer Choices	Responses	
No, and it should not be part of the Emerging Fields assessment process	27.03%	10
No, it works for the FWF Emerging Fields programme, but should not be used elsewhere	35.14%	13
Yes, in some other research funding programmes it would be helpful	29.73%	11
Yes, short summaries for the first assessment stage should be used as widely as possible	5.41%	2
Don't know / no opinion	2.70%	1
Answered		**37**
Skipped		**0**

How satisfied are you with each of the following aspects of the Emerging Fields application process?														
	Very satisfied		Somewhat satisfied		Neutral		Somewhat dissatisfied		Very dissatisfied		Don't know / not applicable		Total	
The clarity of the requirements for applications (call documents, criteria, etc.)	40.54%	15	18.92%	7	18.92%	7	5.41%	2	16.22%	6	0.00%	0	37	
Support during the application process from the FWF (e.g. in response to questions)	54.05%	20	18.92%	7	8.11%	3	8.11%	3	5.41%	2	5.41%	2	37	
User-friendliness of FWF's application submission portal	40.54%	15	32.43%	12	13.51%	5	13.51%	5	0.00%	0	0.00%	0	37	
The clarity and completeness of the feedback I received	35.14%	13	10.81%	4	10.81%	4	10.81%	4	27.03%	10	5.41%	2	37	
The timeframe available to prepare my application	40.54%	15	16.22%	6	18.92%	7	18.92%	7	5.41%	2	0.00%	0	37	
The effort needed to prepare my application	25.00%	9	16.67%	6	33.33%	12	13.89%	5	11.11%	4	0.00%	0	36	
Overall efficiency of the process	27.78%	10	16.67%	6	19.44%	7	16.67%	6	16.67%	6	2.78%	1	36	
Please feel free to note below any major administrative problems or challenges you have encountered that are not covered by the items above.													15	
Answered													**37**	

Skipped	0
Item included an additional freetext-box: "Feel free to comment on any of the aspects above, or any other administrative aspects around the review process, particularly if you had any noteworthy positive or negative experiences:" [answers redacted to protect anonymity]	15

How confident are you in the individuals who reviewed your application in each of the following ways? (Please consider reviews of your synopsis if you were unsuccessful at Stage 1, and consider review of both your synopsis and full application if your application passed to Stage 2)	Very confident		Somewhat confident		Not at all confident		Don't know / not applicable		Total
Reviewers had sufficient understanding of all the fields of research involved in my application	16.22%	6	27.03%	10	40.54%	15	16.22%	6	37
Reviewers provided an unbiased assessment of my application	16.22%	6	18.92%	7	48.65%	18	16.22%	6	37
Reviewers provided a thorough assessment of my application	13.51%	5	32.43%	12	37.84%	14	16.22%	6	37
Answered									37
Skipped									0

Assuming the Emerging Fields programme's overall budget cannot be changed, would you support decreasing the maximum size (currently €6m) for each individual grant so that more grants can be awarded?		
Answer Choices	Responses	
Yes, the maximum grant size should be decreased a lot (to less than half of the current maximum value), enabling significantly more grants to be awarded	8.11%	3
Yes, the maximum grant size should be decreased a little (to half of the current maximum value or more), enabling a few more grants to be awarded	32.43%	12
No, the maximum value should remain as it is	51.35%	19
No, and larger grant size should be considered, even if that would mean that fewer grants can be awarded	2.70%	1
Don't know/ no opinion	5.41%	2
Answered		37
Skipped		0

Assuming the Emerging Fields funding was not available (either because your application is rejected or because the programme simply did not exist), which of the following options best describes the likeliest future prospect for your project idea?		
Answer Choices	Responses	
None: without Emerging Fields funding, my project idea would end and I would focus on other things	10.81%	4
I would continue to realise my project idea, but I am not hopeful that I could secure much alternative funding for it, so I would work with far fewer resources and much reduced scope	37.84%	14
I would significantly modify my project idea by removing some of the novel, high-risk/high-reward aspects and apply for alternative funding	35.14%	13
I would apply for alternative funding without significantly changing my project idea	8.11%	3
Other (please specify)	8.11%	3
Answered		37
Skipped		0
Item included an additional freetext-box: "Feel free to comment on any of the aspects above, or any other administrative aspects around the review process, particularly if you had any noteworthy positive or negative experiences:" [answers redacted to protect anonymity]		3

Assuming the Emerging Fields funding was not available (either because your application is rejected or because the programme simply did not exist), how likely would you be to seek funding for the same project idea from each of the following sources?	Very likely or certain		Somewhat likely		Unlikely or not at all		Don't know/not applicable		Total
A different FWF grant funding scheme	24.32%	9	51.35%	19	21.62%	8	2.70%	1	37
A funding scheme from an Austrian funder other than FWF	13.89%	5	13.89%	5	66.67%	24	5.56%	2	36
A funding scheme from a public funder based outside Austria (incl. EU Horizon/ERC)	43.24%	16	40.54%	15	16.22%	6	0.00%	0	37
Internal funding from my institution	0.00%	0	11.76%	4	79.41%	27	8.82%	3	34
Answered									37
Skipped									0

Please feel free to enter any further comments about your experience of applying for the FWF Emerging Fields programme in the box below.	
[freetext box; answers redacted to protect anonymity]	
Answered	17
Skipped	20

Please feel free to enter any suggestions about how to improve the Emerging Fields programme and review process in the box below.	
[freetext box; answers redacted to protect anonymity]	
Answered	20
Skipped	117

B.2.   Survey of Stage 2 external reviewers

B.2.1. Response rate details

Table 8   Survey of Stage 2 external reviewers – headline response rate

Total Population:	70	Survey responses:	30	Response rate:	43%
Population notes: Population includes all individuals who acted as external reviewers in Stage 2 of the 2022/23 Emerging Fields call. 2 invitations bounced and 2 respondents had opted out of receiving surveys, so 66 could be invited to take the survey.		Response notes: Responses were collected between 11/12/2023 and 16/01/2024, Involving one initial invite and three reminders (see graph below).		Response rate notes: For this population size, we cannot meaningfully denote a confidence interval or margin of error. However, the response rate is relatively high and we identify no evidence of any particular respondent types being significantly under- or over represented. This means that our results should provide an indicative reflection of reality.	

Figure 31 Survey of applicants – Timeline of response submissions
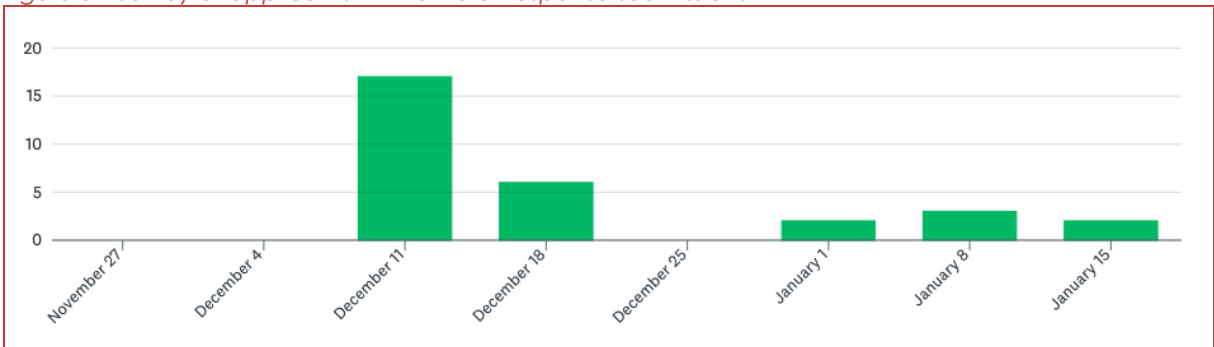
Source: Surveymonkey.

B.2.2. Raw response data

With which of the following areas of science is your work most closely aligned?	
Answer Choices	Responses

Biomedical sciences	13%	4
Natural sciences and Engineering	33%	10
Social sciences, humanities and arts	53%	16
Other/cannot say	0%	0
	Answered	**30**
	Skipped	**0**

How many years of full-time professional research experience (excluding PhD study) did you have when you reviewed the application for the FWF Emerging Fields programme?

Answer Choices	Responses	
0-4 years	0%	0
5-9 years	7%	2
10-14 years	20%	6
15-19 years	17%	5
20-24 years	20%	6
25-29 years	13%	4
30 years or more	23%	7
	Answered	**30**
	Skipped	**0**

Please indicate your level of satisfaction with the following aspects connected to your FWF Emerging fields programme application review

	Very dissatisfied		Somewhat dissatisfied		Neutral		Somewhat satisfied		Very satisfied		Don't know/not applicable		Total
Overall ease of the reviewing process	7%	2	7%	2	0%	0	10%	3	76%	22	0%	0	29
Clarity of FWF guidance and documentation	7%	2	0%	0	0%	0	17%	5	76%	22	0%	0	29
Communication with the FWF during the review process (e.g. for problems or queries)	7%	2	0%	0	0%	0	3%	1	72%	21	17%	5	29
Clarity and coherence of the format and sectioning of the application forms	10%	3	0%	0	0%	0	17%	5	72%	21	0%	0	29
Clarity of the reviewing questions (questions were easy to understand)	7%	2	0%	0	3%	1	10%	3	79%	23	0%	0	29
Clarity of the rating scale (the scale was easy to understand)	7%	2	0%	0	3%	1	14%	4	76%	22	0%	0	29
Appropriateness of the feedback required from you (length, required effort, type of expected feedback)	14%	4	0%	0	3%	1	21%	6	62%	18	0%	0	29
Feel free to comment on any of the aspects above, or any other administrative aspects around the review process, particularly if you had any noteworthy positive or negative experiences:													3
Answered													**29**
Skipped													**1**
Item included an additional freetext-box: "Feel free to comment on any of the aspects above, or any other administrative aspects around the review process, particularly if you had any noteworthy positive or negative experiences:" [answers redacted to protect anonymity]													3

Reflecting on the overall length of the application form, please indicate whether the length of the application form was sufficient, or too short, or longer than necessary for you to make a robust and reliable assessment

Answer Choices	Responses	
The application form was much too short	0%	0

The application form was a little too short	3%	1
The application form was about the right length	69%	20
The application form was a little longer than was necessary	21%	6
The application form was much longer than was necessary	7%	2
Don't know/not applicable	0%	0
	Answered	**29**
	Skipped	**1**

Reflecting on the time window available for you to conduct the review, how appropriate was it?		
Answer Choices	Responses	
A much shorter time-window for me to conduct the review would have been sufficient	0%	0
A slightly shorter time-window for me to conduct the review would have been sufficient	0%	0
The time-window for me to conduct the review was about right	69%	20
The amount of time for me to conduct the review should have been slightly longer	24%	7
The amount of time for me to conduct the review should have been a lot longer	7%	2
Don't know/not applicable	0%	0
	Answered	**29**
	Skipped	**1**

To what extent did the application that the FWF provided you for review reflect your research expertise?		
Answer Choices	Responses	
The application I was given closely or entirely matched my expertise	52%	15
The application I was given mostly matched my expertise	45%	13
The application I was given only partially matched my expertise	3%	1
The application I was given hardly matched my expertise	0%	0
The application I was given did not match my expertise at all	0%	0
Cannot say / no opinion	0%	0
	Answered	**29**
	Skipped	**1**

How original, innovative and/or paradigm-shifting were research ideas presented in the Emerging Fields application you reviewed compared with the research ideas of other research grant applications you have reviewed outside the Emerging Fields scheme? Please feel free to consider for comparison any applications for any basic research grants you may have conducted at any point, for any funder, public or private, in Austria or internationally		
Answer Choices	Responses	
The research ideas in the Emerging Fields application were significantly more original, innovative and/or paradigm-shifting than the research ideas of other research grant applications I have reviewed in the past	31%	9
The research ideas in the Emerging Fields application were slightly more original, innovative and/or paradigm-shifting than the research ideas of other research grant applications I have reviewed in the past	28%	8
The research ideas in the Emerging Fields application were about as original, innovative and/or paradigm-shifting as research ideas submitted for other research grant applications I have reviewed in the past	34%	10
The research ideas in the Emerging Fields application were slightly less original, innovative and/or paradigm-shifting than the research ideas of other research grant applications I have reviewed in the past	7%	2
The research ideas in the Emerging Fields application were significantly less original, innovative and/or paradigm-shifting than other the research ideas of other research grant applications I have reviewed in the past	0%	0
Cannot say / don't know / not applicable	0%	0
	Answered	**29**
	Skipped	**1**

Please provide your assessment of how strongly each of the following criteria influenced your overall judgement of the application you reviewed. Please rank each criterion on a scale from 1 to 10, where '1' means 'no influence' and 10 means 'extremely strong influence'												
	1	2	3	4	5	6	7	8	9	10	Don't know	Total

The feasibility of the proposed research	0%	0	0%	0	0%	0	0%	0	0%	0	14%	4	10%	3	17%	5	21%	6	38%	11	0%	0	29
The extent to which the proposed work departs from established or conventional methods, approaches or perspectives	0%	0	3%	1	0%	0	0%	0	7%	2	14%	4	24%	7	17%	5	17%	5	17%	5	0%	0	29
The level of interdisciplinarity of the proposed research or team	0%	0	0%	0	0%	0	0%	0	7%	2	4%	1	29%	8	32%	9	25%	7	4%	1	0%	0	28
The scientific rigour and robustness of the proposed research plan	0%	0	0%	0	0%	0	0%	0	3%	1	7%	2	3%	1	14%	4	28%	8	45%	13	0%	0	29
The evident level of risk management in the research plan	0%	0	3%	1	0%	0	7%	2	10%	3	10%	3	28%	8	24%	7	17%	5	0%	0	0%	0	29
The importance of the proposed topic and/or research questions to the wider academic community	0%	0	0%	0	0%	0	3%	1	3%	1	0%	0	3%	1	21%	6	28%	8	41%	12	0%	0	29
The track record of the applicant(s)	0%	0	0%	0	3%	1	0%	0	10%	3	7%	2	21%	6	28%	8	17%	5	14%	4	0%	0	29
The reputation, resources and/or facilities of the applicant's/applicants' institution(s)	3%	1	0%	0	7%	2	7%	2	10%	3	7%	2	17%	5	14%	4	21%	6	14%	4	0%	0	29
The degree of integration of trans-disciplinary, interdisciplinary and/or art-based components	0%	0	0%	0	0%	0	0%	0	7%	2	14%	4	21%	6	24%	7	24%	7	7%	2	3%	1	29
The potential of the proposed research to help solve societal problems	10%	3	0%	0	0%	0	7%	2	14%	4	7%	2	14%	4	17%	5	14%	4	10%	3	7%	2	29

Answered	**29**
Skipped	**1**

Please feel free to enter any further comments about your experience as a reviewer for the FWF Emerging Fields programme in the box below. Suggestions about how to improve the programme and review process are also welcome.
[freetext box; answers redacted to protect anonymity]

Answered	**9**
Skipped	**21**

Appendix C   Interview details

Interviews took place throughout the study period, with the final interviews occurring after the Stage 3 Jury hearings. Response rates were consistently very high (70-90% depending on interviewee group) so there is very little risk of self-selection bias. Jury members in particular were interviewed at different stages of the EF process so different sets of questions were asked depending on the timing of each Jury member interview (i.e. early Jury member interviews did not include questions about the Stage 3 hearings). FWF staff members were interviewed throughout the study, often for clarification purposes or to validate various deliverables (e.g. the EF process map presented in this report) so there is no pre-defined set of interview questions for this group. We present below our full list of interviews and the interview tool and questions used throughout.

C.1. List of interviews

Please note, the names of Jury member interviewees have been redacted as the identities of the Emerging Fields Jury were not made public.

Table 9   List of interviews

Name	Interviewee type	Organisation	Position	Interview date	Interviewer
[Redacted]	Jury chair	*[Redacted]*	Professor of Modern History	05/09/2023	Anete Vingre
[Redacted] (re-interview post-Stage 3 hearings)	Jury chair	*[Redacted]*	Professor of Modern History	24/02/2024	Peter Kolarz
Alexander Goritschnig	Applicant institution	University of Graz	Research Manager and Coordinator of the Field of Excellence, 'Complexity of Life in Basic Research and Innovation (COLIBRI)'	19/12/2023	Anete Vingre & Peter Kolarz
Anton Graschopf	Policymaker	Rat für Forschung, Wissenschaft, Innovation und Technologie-entwicklung	Senior Policy Advisor	12/04/2023	Anete Vingre & Peter Kolarz
Barbara Leitner	Applicant institution	University of Vienna	Research Services and Career Development	20/12/2023	Anete Vingre
[Redacted]	Jury member	*[Redacted]*	Professor of Mathematics	24/08/2023	Anete Vingre
[Redacted]	Jury member	*[Redacted]*	Professor of Cell And Gene Therapy	29/09/2023	Anete Vingre
[Redacted]	Jury member	*[Redacted]*	Professor of English	06/09/2023	Anete Vingre
Christof Gattringer	FWF Staff	FWF	President	26/03/2024	Peter Kolarz
Elisabeth Schludermann	Applicant institution	Vienna University of Technology	Senior advisor, Research	19/01/2024	Anete Vingre

Name	Interviewee type	Organisation	Position	Interview date	Interviewer
Eva Gottmann David Müller Wolfgang Neurath Sandra Mukherjee	Policymaker	Austrian Federal Ministry of Education, Science and Research	Various positions	19/12/2023	Anete Vingre & Peter Kolarz
[Redacted]	Jury member	*[Redacted]*	Professor, Deputy Director of Center, and Head of PhD School	24/02/2024	Peter Kolarz
[Redacted]	Jury member	*[Redacted]*	Professor of Greek Art and Archaeology	24/02/2024	Peter Kolarz
Michaela Glanz	Applicant institution	Academy of Fine Arts Vienna	Head of the Art Research Support Department	07/03/2024	Anete Vingre
Nikolaus Possanner	Policymaker	Austrian Science Council	Head of Office	01/12/2024	Anete Vingre
Petra Biberhofer	FWF Staff	FWF	Programme manager, Strategy – National programmes	Several*	Anete Vingre & Peter Kolarz
Tina Olteanu	FWF Staff	FWF	Unit head – Programme management coordination	Several*	Anete Vingre & Peter Kolarz
Uwe von Ahsen	FWF Staff	FWF	Department head, Strategy – National programmes	27/03/2024	Peter Kolarz

*Petra Biberhofer and Tina Olteanu were the main project contacts for this study. In addition to regular update meetings, we also interviewed them on several occasions (usually together) to clarify elements of the EF process, validate process maps, or to probe emerging findings and conclusions where helpful.

C.2. Interview tools

Interviewee Name:	
Position (organisation)	
Role	
Interview date/time	dd-mm-yyyy; xxxxhrs UK time
Interviewer	[Interviewer name]

Points to make prior to interview start

- The FWF has commissioned Technopolis to carry out this evaluation
- The purpose of this interview is to hear about your experience in applying for the programme
- What you say in this interview will only be reported in aggregate non-attributable form, and the notes to this interview will not be shared with anyone

• However, we would like to note the names of all our interviewees in the method annex to the final report. In other words: we'd like to report that we spoke to you, but not what you specifically said. Is this ok with you?

Questions for interviewees from applicant universities:

1. Please briefly introduce your role in your organisation and engagement with FWF and Emerging Fields programme.

2. What was the overall interest from researchers in your university about the EF programme? And what were the key motivations for applying?

3. How EF programme is distinct, complementary or overlapping with other programmes?

4. Does the programme meet the funding needs?

5. What was the overall experience with assessment processes as experienced by submitting research groups? Was the guidance clear, processes transparent, etc?

6. The first stage of the assessment process was based on the short synopsis of your application. What is your/submitting research groups view on the permitted 3-page maximum length of the synopsis?

7. What is your/research groups view on the programme requirements?

8. What is your/research groups view on the support during the application process?

9. What is your/research groups view on the programme timeline?

10. Assuming the Emerging Fields funding was not available, how likely would research groups seek funding for the same project ideas from other funding sources? Which sources?

11. Any other points?

Questions for Jury members interviewed after the first review stage:

1. Please can you describe your overall experience working on the Jury so far? Did you participate in the informal Jury meeting? Did you interact with FWF, the Jury chair, or other Jury members? How many applications did you review in the first stage (pre-selection, based solely on the synopsis)?

2. Have you served on a FWF Jury before for other programmes?

3. Did you find the FWF Jury briefing document clear and useful? Was there any information you found particularly relevant or anything that was unclear?

4. *(ask if the interviewee took part in the informal Jury meeting)* What was the purpose of the informal Jury meeting and how that meeting was relevant/helped with the assessment process? Do you think a similar meeting should be a process of the assessment process, why?

5. To what extent the applications that the FWF provided you for review reflect your research expertise?

6. The first review stage is based solely on the synopsis, and you had to assess the project and the research team. Were there any challenges in arriving at a decision based on the synopsis? Did you have any challenges in assessing the applications based on the evaluation criteria? If so, which evaluation criteria were hard to assess?

7. Do you have any comments about the overall appropriateness and length, format and sectioning of the synopsis? Would you change anything for future funding rounds?

8. The Emerging Fields programme aims to fund non-conventional/original research (transformative, multi- and interdisciplinary, including arts-based research). Did you experience any challenges or difficulties in assessing these elements when reviewing the synopsis?

9. Do you have any comments about the feedback you were required to provide to unsuccessful applicants? Was it appropriate in terms of content requested and length?

10. Did you have enough time to assess the synopsis (it was 2 months)? Do you think you could complete the assessment faster?

11. Overall, how would you rate the quality of the applications you reviewed and to what extent did the applications reflect the programme objectives?

12. Do you have any suggestions for the first review stage (review of synopsis) of the next funding round? Would you change anything?

13. Do you have any suggestions for the upcoming review stages of the current funding round, especially, the presentation to the Jury?

14. Have any other challenges arisen that we have not yet discussed?

15. Any other points?

Questions for Jury members interviewed after the third selection stage (Jury hearings):

1. Let's start with these Jury hearings. Can you give me some reflections on what you feel went well or anything that went not so well?

2. Do you think it was important to have a major presentation event like this (instead of, for instance, just regular remote peer review)? If so, why?

3. Have you been a juror at similar hearings, either at FWF or elsewhere? If so, how do the experiences compare?

4. Did you participate in the informal Jury meeting in March last year? What was the purpose of that meeting? Did it help with the assessment process?
    a. Do you think a similar meeting should be a process of the assessment process?
    b. Did you have any other interactions with, the Jury chair, or other Jury members?

5. Do you have any views on the overall Emerging Fields assessment process, short synopsis reviews, followed by external review of full applications, followed by these hearings? Do you see particular strengths or weaknesses?

6. The first review stage is based solely on the synopsis, and you had to assess the project and the research team. Were there any challenges in arriving at a decision based on the synopsis?

7. Do you feel the purpose of the scheme and your role within the assessment process were communicated to you clearly?

8. What are you views on The Emerging Fields scheme? Do you agree/support this kind of tool? Is it the right tool for the right job?

9. Have any other challenges arisen that we have not yet discussed? Any other points?

Questions for interviews with policymakers:

1. Please briefly introduce your organisation, your role in your organisation and engagement with FWF and Emerging Fields programme?
2. Could you please briefly summarise the main strengths, weaknesses and challenges of the Austrian R&D&I system?
3. Could you please tell us about the Excellent=Austria initiative:
   - how the initiative came about
   - what is the rationale and evidence for introduction of this initiative
   - what are the main objectives and expectations
   - what is the role for the Emerging Fields programme in this context?
4. The Emerging Fields programme aims to fund non-conventional/original research (transformative, multi- and interdisciplinary, including arts-based research). What challenges or difficulties you anticipate in implementing a programme with such focus?
5. Do you have any comments about the overall appropriateness of the Emerging Fields design and assessment process?
6. How is Emerging Fields programme distinct, complementary or overlapping with other programmes available to Austrian researchers?
7. What do you expect from this evaluation?
8. Any other points?

Questions for interviews with FWF staff:

For FWF staff, we opted for a bespoke approach, as different staff members had very different functions within the scheme. We also interviewed FWF staff at various points in the process and the needs and questions changed substantially throughout.

Appendix D   Meeting observation details

As part of the evaluation, we observed all three FWF Scientific Board meetings were decisions were made on the three assessment stages of the Emerging Fields programme. We also observed the entirety of the Stage 3 Jury hearings and subsequent discussion.

For the Stage 1 and Stage 2 Board meetings, we took freeform notes on the discussion and also used a systematic coding frame to categorise discussion points (see below). For the hearings and the final Board meeting, we used freeform notes only. We omit our freeform notes here to preserve the anonymity and confidentiality of Board members. attendance from our study team members was as follows:

- Study team members present at the Stage 1 FWF Board meeting: Peter Kolarz (videolink), Tobias Dudenbostel (in person)
- Study team members present at the Stage 2 FWF Board meeting: Peter Kolarz (videolink), Tobias Dudenbostel (in person)
- Study team members present at the Jury hearings: Peter Kolarz, Tobias Dudenbostel, Erik Arnold (all in person)
- Study team members present at the Stage 3 FWF Board meeting: Peter Kolarz, Tobias Dudenbostel (both via videolink)

D.1. Observation data coding for the Stage 1 and 2 meetings

D.1.1. Data collection instructions for observers

Data collection template for meeting observations

We expect that the FWF board discussion will go through several applications individually. For each one, there should be an assessor (and possibly a co-assessor) who is familiar with the application and has the reviews/recommendations from the Jury. They will likely lead the discussion but other board members may be able to respond and make comments.

We will live-code this discussion of individual applications (or of application synopses, to be exact, hereafter applications nevertheless). The long table below gives space for coding of each individual application discussion.

Each statement made as part of the discussion will be coded by criteria-domain (transformative nature, scientific quality, risk, team, etc) and whether it is positive/supportive or negative/critical of the application (or neutral).

Where a 'statement' begins and ends is a matter for the observers' interpretation. A board member may for instance talk positively about the transformative potential for several minutes, and then say a couple of sentences expressing concern about the suitability of the team. Each part of this pontification is one statement. There will inevitably be 'grey areas' and observers will need to use their experience and interpretation to make sense of these.

To compensate for this, and indeed for the 'weight' of different statements, we have a five-point scale: ++, +, -, -- and 0 (0 being 'neutral). A double-symbol (++ or --) should be used to denote major points of praise or criticism, especially if they are presented as central grounds for shortlisting or rejection.

Our scale is described on the next page, and thereafter is our table we will use for coding.

- For each application, ensure there is enough information entered into the first column to identify it (this can be application number, title, whatever gets us through the day, we can clean up later)

- For each application, we distinguish between the main presenting board members ('assessors') and any other board members who may make further comments.

- For each statement made, enter the appropriate symbol in the appropriate place and hit enter (an assessor might make several positive claims about different aspects of an applications scientific quality – in this way we can capture that)

- Enter comments as you see fit. If this coding frame works well, we might not need many comments. It's only if needed, no obligation to comment on anything

- It is unclear at this point whether we will hear the definitive pass/fail (i.e. shortlist/reject) verdict in the actual meeting, but the final column allows us to note the result if it is announced in the meeting

Our coding scale. Please use these symbols as explained below:

Symbol	Meaning	Explanation
++	Major positive/supportive comment	Board member's statement presents the application in a positive light, typically making a case for why it should be shortlisted, or at least highlighting a positive aspect of the application. Major (++) should be used if the statement is a central point, possibly discussed at length or presented as a main merit of the application, potentially as a single-issue reason for shortlisting
+	Minor positive/supportive comment	
--	Major negative/critical comment	Board member's statement presents the application in a negative light, either making a case for why it should not be shortlisted or highlight some kind of flaw, concern or dissatisfaction. Major (--) should be used if the statement is a central point, possibly discussed at length or presented as a main weakness of the application, potentially as a single-issue reason for rejection
-	Minor negative/critical comment	
0	Neutral/ both at the same time/ cannot classify	Board member's statement cannot readily be classified as either of the above

There is space below to enter data for all 45 applications received by FWF. However, it is almost certain that not all 45 will be discussed – potentially even just a small sub-set. Likely, those where the Jury reviews are all negative will be rejected outright and perhaps there are some with universally positive reviews which will not be subject to discussion either. It is possible however, that either of these 'groups' of applications will still be discussed 'in bulk' in some form. At the end of this document there is a simplified version of the main coding matrix, which can be used to log such discussions. It is unlikely however. As a rule: any individual application that is discussed should be coded in the main table below.

Outside of discussion of individual applications, we of course also want to take notes on the content of any general discussion that might occur. There is no way to prepare a coding frame for this. So there is simple a space on the penultimate page of this document to add notes on those aspects of the meeting. Please use it.

Application	Jury rankings		Innovativeness/ Transformative scope of the project	Scientific quality and related scholarly aspects of proposal	Risk in terms of feasibility	Wider impact and societal importance of the proposed research/ topic	Team's expertise and/or suitability	Interdisciplinary/ arts-based inclusion	Other dimensions (specify in comments if possible)	Comments	Result
App number or title	A, B or C x2		INNO/TRANS	SCI QUAL	RISK	IMPACT	TEAM	IDR/ART	OTHER		Pass/fail
EXAMPLE App 001666	A/B	Assessors' comments	++ -	+ ++ -	–		0 ++			One board member didn't like the proposed use of Zebras	fail
		Others' comments	-	0		-	0 +		-		
		Assessors' comments									
		Others' comments									
		Assessors' comments									
		Others' comments									
		Assessors' comments									
		Others' comments									
		Assessors' comments									
		Others' comments									
		Assessors' comments									
		Others' comments									

D.2. Aggregate findings from the Stage 1 and 2 observation coding analysis

The raw item-by-item coding of the FWF Scientific Board discussions cannot be reproduced here in full as it would breach confidentiality and would risk revealing named individuals' comments on specific applications/synopses.

However, we present below the aggregate figures from the two coded observation exercises. In both the EF Stage 1 and EF Stage 2 Scientific Board meetings, both Technopolis observers coded each comment made on each application/synopsis discussed. Subsequent review of the coding showed that the two observers' had coded very similarly, with around two thirds of Scientific Board members' comments coded identically or near-identically by each observer. Further, there were almost no cases of directly contradictorily coded comments.

In the figures below, we include comments coded equally by both observers and comments coded only by one observer (comments coded equally by both still are only counted as one comment in our analysis).

In short, the analysis indicates:

- For those EF Stage 1 synopses discussed by the Scientific Board, the majority of comments pertained to the innovative/transformative character of proposed ideas, and to levels of risk, while scientific quality considerations were less often the focus of attention

- Of the comments focused on innovative/transformative potential, 33% were of a positive nature (i.e. praising innovativeness) and 67% were negative or critical (i.e. synopses were insufficiently innovative)

- Comments about risk and scientific quality were more overwhelmingly of a negative/critical nature

- For EF Stage 2 proposals, Scientific Board discussion comments were most often about the scientific quality of proposals, and comments in this domain were overwhelmingly (79%) of a critical/negative nature

- Innovative/transformative potential was less often a discussion point, but of the comments in this domain, the great majority (86%) were of a positive nature.

- Of the lesser-discussed categories of comments, the suitability/qualifications of the proposed team carried some significance, with 60% of comments positive at Stage 1 and 73% negative at Stage 2

- Scope for non-academic impact and considerations around interdisciplinary and arts-based approaches played only minimal roles in discussions at both stages.

Our overarching conclusion from these exercises is that the FWF Scientific Board placed the greatest emphasis on synopses' innovativeness at Stage 1, and on scientific quality of proposals at Stage 2. However, at both stages, other factors were also considered. In particular, it is worth noting the presence of many comments on proposals' innovativeness as a source of support at Stage 2.

Table 10 Summary of coded observations: FWF Scientific Board discussion of Stage 1 synopses

	INNO/ TRANS	SCI QUAL	RISK	IMPACT	TEAM	IDR/ART	OTHER	Total	%
Positive comments	5	1	2	3	3	3	4	21	39%
Negative comments	10	8	11	0	2	0	1	32	59%
Neutral comments	0	0	1	0	0	0	0	1	2%
Total comments	**15**	**9**	**14**	**3**	**5**	**3**	**5**	**54**	
% positive comments per category	33%	11%	14%	100%	60%	100%	80%	39%	
% negative comments per category	67%	89%	79%	0%	40%	0%	20%	59%	
% of total comments	**28%**	**17%**	**26%**	**6%**	**9%**	**6%**	**9%**		

NB: analysis includes only the eight synopses that were discussed at the meeting

Table 11 Summary of coded observations: FWF Scientific Board discussion of Stage 2 proposals

	INNO/ TRANS	SCI QUAL	RISK	IMPACT	TEAM	IDR/ART	OTHER	Total	%
Positive comments	24	1	6	3	6	7	4	51	37%
Negative comments	4	30	10	0	16	5	10	75	54%
Neutral comments	0	7	3	0	0	1	1	12	9%
Total comments	**28**	**38**	**19**	**3**	**22**	**13**	**15**	**138**	
% positive comments per category	86%	3%	32%	100%	27%	54%	27%	37%	
% negative comments per category	14%	79%	53%	0%	73%	38%	67%	54%	
% of total comments	**20%**	**28%**	**14%**	**2%**	**16%**	**9%**	**11%**		

technopolis |group

www.technopolis-group.com