SENSITIVE

OBSERVER REPORT

CALL	
Call:	**HORIZON-JU-ER-2023-01**
Topic(s):	HORIZON-ER-JU-2023-FA1-SESAR HORIZON-ER-JU-2023-ExpIR-01 HORIZON-ER-JU-2023-ExpIR-02 HORIZON-ER-JU-2023-ExpIR-03 HORIZON-ER-JU-2023-ExpIR-04 HORIZON-ER-JU-2023-ExpIR-05 HORIZON-ER-JU-2023-ExpIR-06
Type(s) of action:	RIA, IA, CSA
Service:	EU- RAIL
Call deadline:	07 February 2024, 17:00:00
Submission model:	Single

EVALUATION	
Evaluation model:	Single
Panel(s):	Panel 1 RIA, Panel 1 IA, Panel 2 CSA, Panel 2 RIA, Panel 3 CSA, Panel 4 RIA
Observer(s):	Kristin Oxley

TABLE OF CONTENTS

1. BACKGROUND AND SCOPE

Background and scope

This report describes the observer's assessment of the evaluation of the following call:

Call for proposals: HORIZON-ER-JU-2023-01
Deadline: 7 February 2024

This call covers the following topics:

- HORIZON-ER-JU-2023-FA1-SESAR
- HORIZON-ER-JU-2023-ExplR-01
- HORIZON-ER-JU-2023-ExplR-02
- HORIZON-ER-JU-2023-ExplR-03
- HORIZON-ER-JU-2023-ExplR-04
- HORIZON-ER-JU-2023-ExplR-05
- HORIZON-ER-JU-2023-ExplR-06

This call covers the following types of action: RIA, IA, CSA

The report analyses the efficiency of the procedures, usability of the instruments (including IT tools), conduct and fairness of the evaluation sessions, and compliance with the applicable rules.

The objective is to give independent advice for improving the evaluation processes for EU funding.

2. OBSERVER ASSESSMENT

Methodology

Methodology

The observer assessed the quality of the evaluation process with respect to its fairness, efficiency, transparency, consistency and the application of rules, guidelines and best practices.

Prior to the consensus stage, the observer attended a web-based briefing and followed the progress of the individual assessment phase through SEP, including the development of IERs and draft CRs.

During the consensus stage, which was done remotely, the observer attended the general briefing, consensus meetings and monitored the development of CRs in SEP. The observer analysed written information pertinent to the call such as HE reference documents, proposal submission and evaluation guide, the call text, IERs and CRs, etc. The observer furthermore analysed relevant research literature on grant peer review and carried out comparisons with evaluation procedures at national and international levels.

Assessment

Assessment

Scale of complexity of the evaluation task

This evaluation of proposals received in response to the call HORIZON-JU-ER-2023-01 was a moderately complex evaluation. Proposals were evaluated by four panels, consisting of 4-6 experts and one dedicated recorder. Each panel evaluated 2-11 proposals. For one topic only one proposal was received.

The EU- RAIL evaluation team was well prepared to deal with the evaluation task. It has tried and tested routines for the evaluation exercise and experienced and professional staff ensuring these routines are appropriately implemented. Experts received comprehensive briefings on these routines, and active guidance and help throughout the process, ensuring a high-quality process. Unforeseen events like sick leaves and issues associated with conflicts of interest were handled professionally without disrupting the foreseen scheduling of the process.

Overall, evaluation tasks were handled with impressive professionalism and thoroughness in all stages of the evaluation. The excellent planning and preparations carried out by the evaluation team facilitated a smooth-running process.

Transparency of the procedures

The procedures were highly transparent. The review criteria and the evaluation process were comprehensively described in the call and the supporting information was available to all applicants. The evaluation processes observed were clearly aligned with the descriptions given, and the evaluation criteria described were diligently applied.

Evaluators were observed to adhere closely to the directions given in the call text, aided by active moderators and thorough quality controllers. This ensured applications were assessed solely according to the criteria and evaluative dimensions communicated to applicants. While in previous calls observed in national and Nordic contexts, idiosyncratic preferences among evaluators have been observed to play an important role in the faith of applications, the thorough and highly objective review process ensured by the EU- RAIL evaluation team ensured this was not the case in this call.

Consensus reports were of high quality, with judgements that were easily understandable and transparent, and which closely followed the guidelines and information available to applicants.

In sum, the process was exceptionally transparent.

Throughput time of the evaluation and the efficiency of the procedures

The throughput time of the evaluation was excellent. Consensus discussions were scheduled to take 2 hours, with some extra time foreseen for quality control. These schedules were generally adhered to, although some delays were in evidence, with panels on some occasions having to work quite long hours to ensure that they finished their proposals within the overall time period foreseen. The standard working hours foreseen were in general quite long; from 9-18.

While overall, throughput times were excellent, measures aimed at further facilitating efficient discussions could be explored for the future. In other EU-level discussion observed, online consensus meetings have been preceded by a preliminary discussion in SEP, carried out using the comment function in SEP. Such systematic use of SEP discussions has been observed to be a very useful component of the evaluation process which has served to improve the throughput time of the evaluation. It allows consensus discussions to focus primarily on the main issues of contention as many minor issues can be resolved through this kind of written communication. By potentially reducing the time that must be spent in online meetings, such preliminary written discussions can contribute to combating zoom fatigue and in general reduce the detrimental effects of time pressure in online consensus meetings. The social psychology literature on group decision making suggests that time pressure can lead groups to predominantly focus on the information possessed by all members (shared information) over the information than only one or very few possess (unshared information), reducing the richness of knowledge that groups can draw on. Furthermore, time pressure has the effect of increasing groups' desire for uniformity of opinion, reducing their motivation to explore deviant opinions. While moderators and recorders were generally vigilant in ensuring that these dynamics were not a problem in the discussions observed, the use of preliminary SEP discussions might serve to make their task an easier one.

Procedures were highly efficient. The procedures for conducting panel discussion naturally varied somewhat across panels as different moderators had different style; some had a very active approach, while others had a more pared back style. However, they all helped ensured that experts participated actively in the discussion, and that all criteria were appropriately addressed and scored in accordance with the guidance given.

Discussions were overall carried out in a very efficient manner. However, one potential point of improvement that can be considered for the future, is to adopt a simplified discussion of proposals that are clearly below threshold. Some panels were observed to conduct lengthy discussions of proposals uniformly judged below threshold by experts, spending time on detailed wording of minor shortcomings. While it is important to ensure that the feedback that participants receive is not flawed, it is also important to balance an attention to detail with appropriate

progression in discussions. For the future, it might be more efficient to instruct experts to primarily focus discussions for such proposals on shortcomings and weaknesses, avoiding substantial time spent on minor shortcomings. Analogously, calls which has only received one proposal and where the experts overall agree that it is above threshold could benefit from an expedited approach to CR development, as the applicants will surely not complain about details in the CR if they receive funding.

Quality control procedures were overall excellent and very efficient. Compared to the approach observed in other EU evaluations, where quality control has been ensured by external experts, the approach taken in this evaluation whereby internal staff ensured quality control, was much more efficient. It was a smooth process that ensured important control of the quality and consistency of CRs, without demanding excessive resources in terms of time invested. The QC comments did not go into unnecessary detail and focused predominantly on corrections required to establish a correct alignment between comments and scores. In some isolated instances observed, processing the QC comments proved quite time consuming as the process served to re-open the discussion of numerous points, with processing of comments taking up to 1,5 hours. Moderators should thus be vigilant in ensuring that discussion of QC comments is minimal, avoiding the reopening of discussions of the substance of proposals as such.

Practically, the quality controllers gave their comments through the comments function in word, which is preferable to using SEP in this respect as the comments function in SEP functions sub-optimally. While for the future this should be rectified, it is commendable that alternative procedures for collaborative working are ensured in the meantime.

Recorder function. The recorders developed a draft consensus report based on the IERs prior to consensus discussion. During the consensus phase they supported the moderator in seeking consensus and drafted the final consensus report. While some of the recorders were clearly very experienced and efficient, others appeared more unfamiliar with the task. While naturally it is important to train new experts for the role of rapporteur, to minimize potential challenges to the efficiency of discussion that such training entails, new rapporteurs could foremost be employed in the panels that have a low number of proposals to assess.

While overall the standard of recorders was very high in this evaluation, some were truly excellent and were key to facilitating an efficient discussion. In particular one recorder approach was observed to be particularly efficient and could be suggested as an example of best practice for future recorders. Rather than just presenting the cases of diverging opinions in the CR draft, this rapporteur would also suggest appropriate CR text accommodating the different views, greatly facilitating discussions as experts could then just select one alternative and slightly modify this if needed rather than drafting consensus text from scratch.

Efficiency, reliability and usability of the procedures, including the IT-tools

This evaluation was carried out remotely using virtual consensus meetings, in addition to a remote individual evaluation step. Procedures were overall excellently adapted to ensuring a high-quality fully digital evaluation process.

No major issues with the IT-tools were observed. The only minor problem observed of a technical nature was that some experts on occasion had problems joining the Teams meetings, leading to some modest delays, but this was generally solved by rebooting the PC. Furthermore, as discussed in the section on throughput time, the comments function in SEP is not optimal, meaning an exchange of word documents back and forth between QC and panels was the preferred working mode. For the future it would be good if SEP functionality could be enhanced in this respect to avoid the challenges to confidentiality that email exchanges of word documents entails.

The procedures in place were very effective in capitalising on the positive aspects of online evaluation, while minimising negative aspects. While concerns regarding technical difficulties, the excessive strain associated with online meetings, etc were strong in early phases of online discussions observed in the past, these issues were minimal in this evaluation as review procedures and infrastructure were overall very well adapted to the new online reality. However, lack of informal interaction is consistently pointed to as problematic by panellists and few measures addressing this issue has been observed, including in this evaluation. While meetings would start off by an introductory round whereby panellists presented their respective expertise, naturally such presentations tended to be quite brief and offer limited scope for dialogue between experts. Even if panels have a very strict timeline they need to adhere to, it could be worthwhile considering if some more time and effort could be set aside for the introductory phase, as time invested in this respect is likely to pay off in terms of smoother and more efficient consensus discussions. While the informal chat that shared meals and coffee breaks lend themselves to in an onsite evaluation setting are not easily recreated online, means of facilitating more informal interaction at the start of a panel sessions should be explored. A simple google search will provide numerous suggestions for virtual icebreakers, and while it might seem a bit awkward to introduce such get-to-know each other exercises in the context of consensus meetings, it seems warranted to give it a try.

Furthermore, while procedures in this call was overall very efficient in reducing the strain on participants by ensuring meetings were of moderate length and breaks were taken systematically, there were a few instances observed where discussions went on for quite long and requests for breaks from the participants were declined by the moderator due to efficiency concerns. While the intuitive response when one is pressed for time is to cut back on breaks, this tends to be counterproductive as participants' ability to concentrate and contribute actively will deteriorate. Moderators should thus be strongly encouraged to ensure breaks are taken frequently irrespective of the time pressure they face.

Impartiality, fairness and confidentiality of the evaluation

The evaluation was impartial, fair and confidential, constituting an international best-practice example in this respect.

Of utmost importance to ensure the impartiality of the exercise is the requirement for evaluators to be free of any potential conflicts of interest. Requirements in EU evaluations are strict in this regard, and in this evaluation these requirements were clearly communicated to experts in several instances – during the briefings, at the start of panel meetings, etc. Along with strict requirements regarding conflicts of interest, the guiding principles for the evaluation process of independence, impartiality, objectivity, accuracy and consistency were also clearly communicated to experts in the briefing material. Equally, measures intended to ensure the confidentiality of the exercise were consistently communicated and implemented.

The evaluation exercise constitutes a state-of-the-art example of how to ensure fair evaluation. Proposals were evaluated with parity, and criteria and scoring scales were interpreted in a uniform fashion and applied coherently. Effective quality control further helped ensured the consistency of evaluation approach across proposals. A minor issue observed with respect to consistency, was that on occasion panels confused the principle of not comparing proposals in terms of their content as an instruction not to check back with previous assessments to ensure they were applying a consistent evaluation approach. To avoid similar confusion in the future, the EU-RAIL evaluation team should explicitly address the difference between comparing proposals with respect to their content and with respect to the overall evaluation approach, clearly condoning the latter approach as this constitutes an important element in ensuring consistent and fair evaluations across proposals.

Conformity of the evaluation with the applicable rules (including guidance documents)

The evaluation was observed to closely confirm with the applicable guidance documents and the call text. These were communicated to experts in the various instructions and briefings given and reinforced during the evaluation through the active moderation of discussions, and further checked and verified through the QC process. This was excellent.

Quality of the evaluation process in comparison with similar national/international evaluation procedures

In the context of carrying out a PhD on the subject of panel peer review of research grant applications, the observer has over the last three years spent more than 150 days observing more than 140 panels at national, Nordic and EU-level. The observer also knows the literature on grant panel peer review well, including the large empirical literature detailing peer review procedures in use internationally. Based on this knowledge, it is clear that this evaluation process was of comparably excellent quality, constituting the state of the art in terms of ensuring a thorough, fair and transparent evaluation.

The strengths of this evaluation process in comparison to many national and international evaluations observed is that individual experts all follow a standardised and detailed evaluation set up in their individual assessments, providing comments for all criteria and sub-criteria specified in the call. This ensures all applications are assessed with the same rigour and according to the same standards. Furthermore, there is sufficient time available for consensus discussions (2-3 hours) and discussions are clearly structured according to the evaluation criteria with clear instructions for how to translate assessments into scores and calculate overall scores, further contributing to thorough and systematic assessments. Finally, evaluation outputs are thoroughly quality checked, ensuring consistent assessments and high-quality feedback to applicants.

In contrast, evaluation exercises observed at national level tend to structure individual and panel evaluations merely according to the overall criteria of excellence, impact and implementation, leaving experts significant discretion in how to operationalise these in more detail. Furthermore, panel discussions tend to be short, often limited to around 15 minutes, limiting the scope for integrating the substance of individual experts' assessments, often resulting in a tendency to merely average individual scores in order to reach a consensus. The variability inherent in individual assessments thus tend not to be discussed and addressed, and there are no other calibration measures such as quality control foreseen. Evaluation output thus tends to be subject to considerably more variability that that observed in this evaluation.

Overall quality of the evaluation

With reference to the comments in the above section, the evaluation altogether was excellent, constituting a state-of-the-art evaluation process. While the reliability of many national evaluation processes observed might be questioned due to the limited resources invested both on the part of the experts and the organising authority, the thoroughness and consistency of the procedures in this call were very impressive, and a reliable evaluation outcome of excellent quality was achieved.

Other remarks

Quality of the documentation provided to experts beforehand
The quality of the documentation received by evaluators prior to starting evaluations was excellent. Comprehensive written briefing material was made available, including detailed guidelines on the development of appropriate IERs/CRs. However, in other evaluations observed, experts have also been provided with a concrete example of a best practice IER/CR and have reported finding this very helpful. To further guide and assist experts, EU-RAIL could also consider providing experts with such a concrete and comprehensive example.

Quality of the briefing sessions
Briefings were very informative and well-structured. Through the general briefing carried out prior to the individual assessment phase, experts were provided with a good overview of the policy context and received guidance on how to carry out the evaluation. A shorter briefing was carried out immediately preceding the consensus discussion, with contributions from DG MOVE regarding the policy context, the acting Executive Director regarding the evaluation process, a dedicated adviser regarding the lump sum approach and the call coordinator regarding the practicalities associated with the evaluation, including reimbursements. This overall functioned very well.

The understanding by experts of the call (context, topics), of the evaluation process and their role
Evaluators reported to have a clear understanding of the call, the evaluation process, and their role. Moderators were active in ensuring that the evaluation approach taken was aligned with the requirements of the call and relevant guiding documents. Recorders and quality controllers further contributed in this respect.

Criteria and scoring scheme: appropriateness, completeness, relevance, clarity, consistency in application
The criteria and scoring scheme were very thoroughly presented and explained, and evaluators received extensive support from the moderators in applying the evaluation criteria and associated sub-criteria in a consistent fashion. The evaluation set-up, including quality control ensured consistent interpretation of criteria and consistent scoring across proposals.

However, some criteria were observed to cause confusion. Experts were observed to have trouble differentiating between the sub-criterion under excellence with respect to interdisciplinary approaches, and the sub-criterion under implementation regarding extent to which the consortium as a whole brings together the necessary expertise. The moderators were observed to be very helpful in clarifying the difference, but this difference might be explicitly explained in future briefings and guidelines for IER/CR development to avoid confusion.

Similarly, in one panel observed, experts struggled with keeping apart the sub-criterion under excellence associated with the soundness of the methodology and the sub-criterion under impact associated with the credibility of the pathways to achieve the expected outcomes and impact. This could also usefully be explicitly addressed in future guidelines for IER/CR development.

Scores are as a main rule set based on the number of shortcomings identified. Compared to previous evaluations observed, it is highly commendable that this evaluation had made an attempt to rectify this imbalance by highlighting to experts in the briefings that they in addition to identifying proposals' shortcomings should also identify their strengths. However, a systematic approach to ensuring that such strengths affected scores was lacking. This was reflected in the instructions to experts from moderators and the feedback from quality controllers, which focused exclusively on shortcomings when determining scores and when judging the appropriateness of the scores given.

While this is an easy-to-use approach that is effective in ensuring consistency of scoring across proposals, the downside is that such an approach tends to punish highly ambitious proposals excessively for the risks associated with taking large strides forward, while not sufficiently rewarding proposals for the potential associated with such an approach. The risk is thus that only proposals attempting incremental advances will make it through the review process as these will be less risky and thus less vulnerable to fault-finding. The 2022 study of the proposal evaluation system for the EU R&I framework programme found EU proposal evaluation processes to be anti-innovation[1], suggesting this is a system-level problem.

Equating top scoring proposals with fault-free proposals is likely to have a conservative effect, but since this is a system-level problem it is likely not easily solved by EU-RAIL alone, as adjustments in the general scoring instructions would be needed. However, efforts to alleviate the conservative tendencies of the current system could be envisaged. As experts are already instructed to systematically mention strengths in their assessments, a systematic approach to ensure these affect scores could be implemented without changing scoring instructions. Currently a shortcoming amounts to approximately half a point deduction in score. For the future, experts could be instructed to reflect a strength as approximately half a point increase in score. As the current scoring instructions at general level does not define "a small number of shortcomings", there is leeway for interpretation, and thus for allowing strengths identified to "cancel out" shortcomings identified, without being in direct contradiction with the general guidelines. However, the score of 5 would still need to be reserved for proposals with only minor shortcomings due to the manner that this is operationalised in the scoring guidance.

[1] Rodriguez-Rincon, D., Feijao, C., Stevenson, C., Evans, H., Sinclair, A., Thomson, S., & Guthrie, S. (2022). Study on the proposal evaluation system for the EU R&I framework programme. Final Report.

The process of the individual evaluations and the actors involved

The process of the individual evaluations proceeded according to schedule and no major issues were observed or reported. The quality of the IERs appeared to be very good and it is very good that detailed guidelines for their development are provided.

However, experts could be encouraged to briefly re-assess their IERs once they have finished all their designated proposals and make necessary adjustments to ensure a consistent approach is adopted. This could serve to minimise the impact of the sequence effect. In several instances experts were observed to give comments in panel meetings along the lines that "this was the first one I assessed so I did not want to be too strict, I got stricter as I moved along", etc. This kind of sequence effect is normal. Research in psychology, behavioural finance and behavioural economics has highlighted that the sequence in which individuals make decisions affects the decision outcomes through learning[2] as inexperienced decision makers need to form appropriate expectations about expected outcomes. That is, when experts begin evaluating projects, they believe a certain fraction of projects to be good and build a screening function that corresponds to their beliefs about the fraction of good projects in the pool. Over time, they learn and update their screening function depending on the quality of projects they have seen. However, during this learning period, they may (dis)favour a project based on where it appears in the sequence. Alerting experts to this sequence effect and encouraging them to take appropriate measures to ensure a consistent approach to counteract its adverse effects in the individual evaluation phase could thus be considered.

The quality of the draft CRs was overall also very good. However, there were substantial differences between recorders with respect to the style adopted. While some adopted a quite detailed approach, others were very minimalist. While both approaches worked well, some more guidance to recorders with respect to the length and detail expected could be considered. Currently, the moderators do not give feedback to the draft CRs. This kind of feedback has been observed to be important to ensure a high-quality consistent approach to CR development in other EU-level evaluations and could be considered for the future.

The process of the consensus meetings and the actors involved

Consensus meetings were of a very high quality, with dedicated experts and professional staff. The advantage of bringing experts together in a panel meeting is that it gives them the opportunity to jointly re-evaluate their arguments, weigh the arguments of the various experts against each other and distinguish between important and less important arguments. Others' arguments can stimulate new arguments and insights and individual errors and misunderstandings can be uncovered by the group and corrected in the discussion. Panels in this call were observed to consistently capitalize on these benefits associated with panel discussion.

In the panel meetings, experts were observed to actively draw on each other's expertise and aided the recorder in developing high-quality CRs, referring back to the proposal, to relevant annexes, to the call text, etc. in order to ensure that they were making correct and consistent assessments. However, to further capitalise on the draft CRs for the future, experts could be allowed to access these prior to the meeting in order to prepare for discussions. Experts in other EU-evaluations observed where this has been customary reported that this was very useful and contributed to more smooth-running discussions as experts were more prepared.

Moderators consistently ensured that all relevant aspects of each criterion were discussed and that the main points were recorded in the consensus report. While experts at times expressed frustration that each and every sub-criterion had to be addressed, this is a good approach that ensures consistency of evaluation across proposals. While generally moderators were vigilant in ensuring that all experts participated in discussions, explicitly asking for the opinion of those experts less active, there were some instances observed where some experts took on a very pared back role in discussions without being prompted by the moderator to contribute more actively. To capitalize on the richness of expertise represented in the panel it is important that moderators consistently ensure balanced participation in discussions.

Quality controllers contributed effectively to ensuring that panels adopted a consistent assessment approach and had substantial impact in this respect, with experts frequently deciding to change their scores as a result of QC comments. Panels were conscious of processing the QC feedback as quickly as possible upon receipt, which is very positive. This helps ensure early learning about the approach expected, progressively building a better understanding among panelists for the appropriate evaluation approach to take and thus progressively ensuring smoother discussions.

Quality of evaluation summary reports

The quality of ESRs was excellent. While there were considerable differences between panels in terms of the approach taken with respect to ESR length, all proposers received high-quality and helpful feedback, closely aligned with the evaluation criteria, ensuring overall consistency in terms of the substance of reports. This is very important, as it ensures that even applicants who do not receive funding obtain value from the process, gaining insight into the strong and weak parts of their proposal, ensuring that they are better positioned for success in future applications.

Working conditions for evaluators

During consensus discussions, some panels were observed to successfully keep discussions rather brief and take frequent breaks, while others – more pressed for time, took breaks very infrequently. For the future, systematic breaks should be taken even in situations where panels are pressed for time. This is not just important for experts'

[2] The Sequence Effect in Panel Decisions: Evidence from the Evaluation of Research and Development Projects. Paola Criscuolo, Linus Dahlander, Thorsten Grohsjean, and Ammon Salter. Organization Science 2021 32:4, 987-1008

wellbeing, but also for the efficiency of the evaluation exercise. Well rested experts will be able to work more efficiently.

It is very positive that experts can give their comments on the evaluation process both directly to the observer and through a dedicated survey. To further enrich the feedback collected from experts in this respect for the future, this could also be included as a point of discussion in the panel review meeting. Currently, experts' assessment of the call and the quality of proposals as well as their feedback on how future calls might be improved I solicited during the panel review. This overall appeared to work very well and rich feedback was generated. This request for feedback could usefully be complemented with also asking experts explicitly about how the evaluation process as such might be improved, while the exercise is still very fresh in their mind. This could potentially provide richer feedback than what can be obtained in survey format.

Overall conduct of staff
EU-RAIL staff was highly professional and very helpful to both experts and the observer. All moderators had the necessary skills and contributed to creating a positive and productive atmosphere among experts..

Recommendations

Recommendations

- **Consider supplying experts with a concrete example of a best practice IER/CR**
  While the quality of the documentation received by evaluators prior to starting evaluations was excellent, EU-RAIL could also consider providing experts with such a concrete and comprehensive example to further facilitate a consistent approach to IER/CR development.
- **Consider giving feedback on the first CRs developed**
  This kind of feedback has been observed to be important to ensure a high-quality consistent approach to CR development in other EU-level evaluations and could be considered as a complement to the already excellent guidance given.
- **Consider encouraging experts to take measures to counteract the sequence effect**
  Evaluating projects is a learning exercise, but during the learning period, experts may (dis)favour a project based on where it appears in the sequence. Alerting experts to this sequence effect and encouraging them to take appropriate measures to counteract it could thus be considered.
- **Consider introducing a pre-discussion exchange of comments in SEP**
  Such pre-discussions have been observed to be a useful component of the evaluation process in previous EU-level evaluations observed, which has served to improve the throughput time of the evaluation as it allows consensus discussions to focus primarily on the main issues of contention.
- **Consider giving experts access to draft CRs prior to discussions**
  Experts in other EU-evaluations observed where this has been customary reported that this was very useful and contributed to more smooth-running discussions as experts were more prepared.
- **Consider investing more time in the introductory phase of CR discussions**
  While brief introductions of experts are carried out, more time invested in this respect could pay off in terms of smoother and more efficient consensus discussions. Experts consistently point out that a lack of informal interaction is the major drawback to online evaluations and alleviating this somewhat for example by organising some simple icebreaker exercises at the start of meetings appear warranted.
- **Consider a simplified discussion of proposals where a weakness is identified**
  While the provision of high-quality feedback is important for the legitimacy of the evaluation exercise, the subtle nuances of how a minor shortcoming should be worded is less important if a weakness has already been identified, rendering the proposal non-eligible for funding. In such instances, moderators could usefully instruct experts to primarily focus on reaching agreement on shortcomings and weaknesses.
- **Consider improvements to the IT tools**
  SEP should be further improved and adapted to online working, with better functionality for providing comments to CRs, thus avoiding the avoiding the challenges to confidentiality that email exchanges of word documents between QC and panels entails.
- **Consider introducing measures that will award proposal strengths, not just punish shortcomings**
  It is highly commendable that this evaluation required experts to identify proposals' strengths, not just shortcomings, but an approach for translating such strengths into scores was lacking. For the future, experts could be instructed to consistently reflect a strength as approximately half a point increase in score.
- **Consider clarifying the principle of not comparing proposals**
  On occasion panels confused the principle of not comparing proposals in terms of their content as an instruction not to check back with previous assessments to ensure they were applying a consistent evaluation approach. While the former should be discouraged, the latter should be explicitly encouraged to help ensure consistency of assessment across proposals.
- **Consider clarifying certain evaluation criteria**
  Experts were observed to have trouble differentiating between the sub-criterion under excellence with respect

to interdisciplinary approaches, and the sub-criterion under implementation regarding extent to which the consortium as a whole brings together the necessary expertise. Furthermore, some struggled with keeping apart the sub-criterion under excellence associated with the soundness of the methodology and the sub-criterion under impact associated with the credibility of the pathways to achieve the expected outcomes and impact. The differences between them could usefully be explicitly addressed in future guidelines for IER/CR development.

- **Consider more frequent discussion breaks**
  Discussing online is more tiring than discussing onsite, and breaks should be taken frequently. This is important both for the efficiency of the process as well as for the wellbeing of experts and moderators.

- **Consider asking experts for their feedback on the evaluation process in the context of the panel review**
  This will provide an opportunity for collecting rich feedback while the evaluation exercise is still fresh in experts' mind.