



**SENSITIVE**

## OBSERVER REPORT

CALL	
Call:	HORIZON-ER-JU-2022-02
Topic:	HORIZON-ER-JU-2022-FA7-03
Type of action:	IA
Service:	EU- RAIL
Call deadline:	December 14 2022
Submission model:	Single

EVALUATION	
Evaluation model:	Single
Panel(s):	Panel 1 IA
Observer(s):	Kristin Oxley

### TABLE OF CONTENTS

1. BACKGROUND AND SCOPE .....	2
2. OBSERVER ASSESSMENT .....	2
Methodology .....	2
Assessment .....	2
Recommendations .....	7

## 1. BACKGROUND AND SCOPE

Background and scope
<p>This report describes the observer’s assessment of the evaluation of the following call:</p> <p><b>Call for proposals:</b> HORIZON-ER-JU-2022-02  <b>Topic:</b> HORIZON-ER-JU-2022-FA7-03  <b>Deadline:</b> December 14 2022  <b>Budget:</b> EUR 14,7 MILL  <b>Budget for the topic:</b> EUR 2,3 MILL</p> <p>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.</p> <p>The objective is to give independent advice for improving the evaluation processes for EU funding.</p>

## 2. OBSERVER ASSESSMENT

### Methodology

Methodology
<p>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.</p> <p>Prior to the consensus stage, the observer attended a web-based briefing with the call coordinator and followed the progress of the individual assessment phase through SEP.</p> <p>During the consensus stage, which was done remotely, the observer attended the general briefing, the consensus meeting and the review of QC feedback in full.</p> <p>The observer analysed written information pertinent to the call such as HE reference documents, proposal submission and evaluation guide, the call text, Q and A, IERs and CR, etc. The observer furthermore analysed relevant research literature on grant peer review and carried out comparisons with evaluation procedures at national and international levels.</p>

### Assessment

Assessment
<p><b>Scale of complexity of the evaluation task</b></p> <p>This re-evaluation of the single proposal received in response to the call HORIZON-ER-JU-2022-02, topic HORIZON-ER-JU-2022-FA7-03 was not complex, but demanded high attention to rigour and thoroughness, given that the redress review of the first evaluation carried out concluded that a re-evaluation was necessary.</p> <p>The EU- RAIL evaluation team was well prepared to deal with the evaluation task. Prior to undertaking the evaluation, experts received comprehensive briefing material – both written information and an online briefing session - which gave guidance on the relevant policy context, the call, and the evaluation task. The briefing included a very detailed guide to experts on how to evaluate and score the award criteria, including good examples of how to draft appropriate consensus reports. This was very good. The proposal was evaluated by a panel of experts consisting of six members, including one expert specialised in financial aspects.</p> <p>The on-site briefing of experts was comprehensive and clearly communicated the principles of the evaluation process and the relevant policy context. The briefing also conveyed practical and logistical information pertaining to reimbursement of expenses, etc, which was very useful.</p> <p>In their review of good practice grant peer review internationally, the European Science Foundation underlines that "before the tasks of both individual/remote reviewers and panel members begin, it is essential that their assignments are clearly described and communicated" This was clearly the case in this evaluation.</p> <p>Overall, evaluation tasks were handled with professionalism and thoroughness in all stages of the evaluation.</p>

## Transparency of the procedures

A pre-requisite for confidence in the integrity of peer review is that its workings should be transparent. This was clearly the case in this evaluation, which was conducted according to high standards of transparency

The set of evaluation criteria and sub criteria against which the proposal was evaluated were clear and well explained in guidelines. The moderator and recorder actively ensured that experts used the scores appropriately in the course of the consensus meeting. While the literature on peer review suggests that peer review procedures tend to include informal elements that might impact the outcome of the evaluation, no such informal elements were observed during this evaluation. The panel was meticulously reviewing the proposal in line with the stated criteria, scoring it according to the stated guidelines.

The use of an Independent Observer serves to verify that the stated guidelines and procedures are followed in practice and further contributes to the transparency of the evaluation exercise. The observer function also provides a way for identifying and airing potential problems and make suggestions/recommendations during and after the evaluation so that corrective actions can be adopted.

Detailed feedback is given to applicants on the results of the evaluation, enabling them to understand the basis for the funding/rejection decision, including the reasoning behind the scores allocated to their proposal. This helps ensure the transparency of the evaluation vis á vis applicants. The consensus report was of excellent quality, with transparent and clear comments in line with the guidelines and information available to applicants through the call and the Q and A.

In sum, the final result was achieved in a transparent way.

## Throughput time of the evaluation and the efficiency of the procedures

### Throughput time

The proposal took considerable time to evaluate. Discussing and agreeing the first draft of the consensus report took a full day. The following day, the panel spent two hours going through the quality control comments and finalising the panel report. The considerable time spent was warranted considering this was a re-evaluation of a proposal and thus ensuring all elements were thoroughly substantiated was important in order to ensure the confidence of the applicant in the assessments made. Research shows that restricted time available for evaluation tends to reduce the ambitions of panellists to execute rigorous reviews and also increases the risk of groupthink, whereby experts tend to focus primarily on reaching consensus, and are not motivated to detect possible weaknesses in their decisions and to realistically appraise alternative decisions. There were no instances of groupthink observed in this evaluation. The discussion was very thorough and detailed, and experts were open-minded and ready to change their initial views based on the arguments of their fellow experts.

Discussions were aided by the high-quality CR draft developed by the recorder prior to discussions. However, the moderator and recorder had somewhat diverging opinions of the appropriate format of the CR, with the moderator preferring a briefer format than that suggested by the recorder. This meant some time had to be spent in the discussion shortening the text. This kind of editorial amendments are often more efficient to carry out offline. Thus, for the future, a procedure whereby moderators consistently check and give feedback on recorders' draft CRs prior to the consensus meeting could be considered, ensuring the format is aligned with moderators' preferences.

### Highly efficient procedures

#### *Very efficient moderation*

Moderation was very good, helping ensure a thorough and probing consensus discussion. Panel interaction potentially ensures a broader knowledge base for decisions and more thorough processing of the information that constitute the basis for decisions. Panel interaction also helps ensure a more uniform understanding of the organisational requirements, including scoring scales and review criteria. However, group interaction also carries the risk of poorer decision making because shared responsibility creates a situation in which everyone withdraws and no one really puts in the necessary effort, known as social loafing. The moderator in this evaluation helped ensure that such social loafing was avoided, ensuring that everybody participated actively in discussion and encouraging the panel go further into relevant details where needed, giving good guidance on appropriate score setting and understanding of review criteria.

#### *Efficient quality control*

The quality control was ensured by internal staff. This was observed to work very well, with good and pertinent comments offered on the draft CR. However, the group's processing of the comments took considerable time, in total two hours. If the quality controller had participated in the meeting in person for the review of the comments, this might have led to somewhat more efficient discussions, as some confusion was observed with respect to the meaning of the comments. While in a normal evaluation, where several panels run in parallel, the participation of the quality controller in the review of comments can prove difficult to ensure logistically, in this instance, with only one proposal under review it might have been feasible. For the future, this option could also be considered for evaluations where the number of panels running in parallel is restricted. Prior observations of EU-level review processes, featuring both in-person QC review and QC review without the presence of the quality controller clearly showed the in-person QC review to be the most efficient. If it is not possible for the QC to remain with the panel for the entire discussion of the comments, merely participating to introduce and explain the comments at the start of the working session could also be effective in speeding up the processing of QC feedback.

*Recorder task handled efficiently*

The recorder developed a draft consensus report based on the IERs prior to consensus discussion. During the consensus phase the recorder supported the moderator in seeking consensus and drafted the final consensus report. The recorder displayed good drafting skills, good command of the English language and good ability to summarize the discussion.

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

This evaluation was carried out remotely, with a virtual consensus meeting in addition to a remote individual evaluation step. The technical infrastructure worked efficiently. No technical problems were observed with either Teams or SEP which were the two main IT-tools in use.

The use of an independent observer in the evaluation exercise ensures comprehensive feedback on the efficiency, reliability and usability of the procedures. Furthermore, in the process of agreeing the panel report, under section 3 "Additional comments and recommendations", the experts are also asked about their views on the efficiency, reliability and usability of the procedures. It is very good that the experts' feedback is systematically collected in this manner.

In addition, the presence of observers from DG MOVE and ERA further helped ensure the reliability of the exercise. They constituted an extra check that the procedures followed were aligned with the intentions of the call and the guidelines communicated to applicants. Both intervened on occasion to ask clarifying questions to ensure that this was the case.

In sum, the procedures and tools used were reliable and efficient.

**Impartiality, fairness and confidentiality of the evaluation**

The observer is convinced that the evaluation was impartial, fair and confidential, constituting an international best-practice example in this respect.

*Appropriate consideration of conflicts of interest and confidentiality*

Proper guidance was given to evaluators through written information material and through the two briefings to ensure impartiality and confidentiality of the evaluation. Cronyism is a concern for many major funders, and this evaluation had appropriate processes in place to counter the presence or perception of such biases. No problems with conflicts of interest were observed or reported in this evaluation.

The importance of keeping the evaluation confidential was underlined to experts on numerous occasions – in the information material supplied to experts, in the general briefing, in the course of consensus discussions, etc. No breach of confidentiality was observed or brought to the attention of the independent observer during the evaluation process.

*Uniform use of evaluation criteria and scores*

Research on peer review shows that experts often have different opinions on how assessment criteria should be interpreted and emphasized in the overall assessment, and different practices regarding how assessments are translated into scores. This evaluation had a number of effective measures in place to ensure that the evaluation criteria and scores were understood and used by experts in a uniform manner, ensuring a fair evaluation. Proposal score interpretation explanations were clearly communicated to experts prior to the individual assessment; the moderator and recorder actively ensured that experts used the criteria and scores appropriately in the course of consensus meetings; quality control was carried out of the draft consensus report to ensure that criteria and sub-criteria were addressed and scored in an appropriate and unambiguous manner.

*Good mix of experts and appropriate number of experts reviewing the proposal*

A good mix of experts is important because there can often be different opinions on what constitutes good research. The proposal was reviewed by six experts which is well beyond the standard observed in other evaluations. Based on a review of over 30 European research funding organizations, the European Science Foundation concluded that "In general, the aim should be to provide at least three expert assessments before a final decision is made". Furthermore, none of the experts involved in the original assessment of the proposal were involved in its re-evaluation. This was very good and ensured an independent and fair re-assessment of the proposal.

*Thorough, fair and impartial consensus discussions.*

The discussion was overall of a very high quality; it was detailed, systematic and probing. Experts were instructed to agree on comments before discussing scores. This is positive as it helps ensure that consensus discussions are not reduced to simple averaging exercises. Focusing on comments rather than scores generally also ensures a more thorough and in-depth discussion. Consensus discussions observed in national funding agencies where this has not been an established principle has often tended to be discussion about numbers rather than content. A focus on comments rather than scores can also make it easier for experts to focus less on their original and preferred score, making for more flexible discussions where consensus can be reached with greater ease.

The moderator and recorder ensured that comments and assessments of all the individual elements that should be assessed under each of the main evaluation criteria were included in the consensus report, and they were

observed to do so systematically and thoroughly. This is very positive as it helps ensure that experts share information not already known to the other experts; more general discussions tend to focus on the information that all experts are already familiar with - so called sampling bias in favour of shared information. Discussion steered by a set of detailed criteria thus helps increase the value-added of the panel exercise; decisions are made based on broader and more varied information.

Experts were observed to be very dedicated and committed to ensuring that the proposal received a fair and detailed review, even if this on occasion meant that discussions went on for quite long.

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

The evaluation was conducted in full conformity with applicable rules and guidelines published in the context of the call. 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. Based on previous experience, which revealed that experts found it particularly challenging to understand the guidelines regarding lump sum and how these should be followed up in their evaluation, dedicated measures aimed at clarifying and aiding experts in this regard were foreseen. This included a dedicated briefing on the subject and the inclusion of a financial expert in each panel. 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 two years spent more than 130 days observing more than 120 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 base, EU-RAIL's procedures are comparably of excellent quality.

The process was very reliable; experts are given extensive information and guidance regarding how to correctly assess proposals. Furthermore, the evaluation constitutes a best-practice example in terms of its fairness. In line with the literature on peer review, a number of procedures are in place to help guard against bias and ensure impartial and fair reviews, including strict rules on conflicts of interest, uniform use of evaluation criteria and scores, and efficient quality control mechanisms.

In contrast to many national level evaluations observed, the process was organised in such a manner that it ensured that the value added associated with grant panel discussions could be reaped. The value added of panel discussions compared to just averaging the scores of a number of individual assessments is that a panel can draw on a larger pool of knowledge as each member has unique knowledge they can contribute, enabling a more holistic assessment of proposals. In a panel context, members are also able to process the information they possess more thoroughly through the discussion, asking questions, considering alternative options, weighing arguments against one another, etc. Through such information processing, the group is better positioned than individuals to identify errors. While many national evaluations observed fail to reap these benefits due to insufficient resources and time invested in both the individual and consensus part of the evaluation process, in this evaluation staff assisted experts in carrying out a very high quality evaluation, ensuring sufficient discussion time and effective moderation which enabled experts to draw effectively on each other's' expertise and effectively spot errors both in their respective individual assessments and in the proposal itself.

### **Overall quality of the evaluation**

The evaluation constitutes an example of international best practice. It was conducted overall to the highest professional and quality standards. Both EU-RAIL staff and external experts demonstrated high competence and professionalism, and extensive quality assurance procedures were implemented. In sum, this meant the proposal was re-evaluated in an independent, efficient, fair, reliable and transparent manner, with high-quality feedback to applicants as the final result.

### **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.

#### ***Quality of the briefing sessions***

Briefings were very informative and well-structured. Briefings carried out prior to the individual evaluation were tailored to the different needs of the experts and the recorder, and a shorter briefing for all involved was carried out immediately preceding the consensus discussions. In this manner experts were provided with a good overview of

the evaluation procedure, the different roles involved and received guidance on how to carry out the evaluation in a timely and tailored manner.

In many national level evaluations observed, briefings are only carried out prior to the consensus stage. The fact that comprehensive briefings are also organised prior to the individual phase is very positive, as this helps ensure a uniform understanding and application of criteria and scoring scales at an early stage in the evaluation, with less effort having to be invested in this at the consensus stage.

To further capitalise on the advantage of the briefing carried out prior to the individual evaluation stage, there appears to be scope for shortening the briefing held immediately preceding the consensus stage somewhat. As it is now, a lot of the information given in the briefing preceding the individual evaluation stage is repeated in the second briefing. Ideally, they should complement rather than duplicate each other, with for example the first briefing focusing predominantly on the information relating to the appropriate development of IERs, and the second briefing focusing more on the consensus meeting as such and its output.

***The understanding by experts of the call (context, topics), of the evaluation process and their role***

Evaluators overall appeared to have a clear understanding of the call, the evaluation process, and their role. However, the experts commented that the call was not sufficiently clear regarding the required TRL level that the proposal was required to reach. While the Q and A published clarified that the required TRL level did not have to be reached by the project budget alone, but should rather be reached with the help of the 10 times leverage budget required, experts still thought this overly ambitious given the relative immaturity of the technology. The failure of the applicant to meet the requirement of the call in this respect was the main reason that it failed. Clarifying this issue was also the main reason the panel needed so much time to discuss the proposal. To prevent similar situations for the future, they recommended that a research and innovation action call, rather than an innovation action call should be the first step when awarding funding to immature technologies like the hyperloop technology.

In this respect, it was positive that observers from DG MOVE and ERA were present, as following the process of operationalising the call text into a detailed assessment of the proposal in question might offer important insight, useful for the drafting of future calls on the subject.

***The process of the individual evaluations and the actors involved***

The process of the individual evaluation proceeded according to schedule and no major issues were observed or reported. During the individual evaluation phase, each expert read the proposal and wrote an IER. Following the completion of the IERs, the designated recorder developed a draft Consensus Report based on the IERs. The quality of the IERs and draft CR was very good.

***The process of the consensus meetings and the actors involved***

Consensus discussions were observed to be very well structured and thorough. Experts were instructed to agree on comments before discussing scores, which as discussed in the above functioned very well. Once comments were agreed for all sub-criteria within a criterion, the panel of experts agreed the score for the criterion in question, adhering faithfully to the scoring guidelines with respect to how the number of shortcomings/weaknesses related to scores.

The standard of experts were very high, and they all contributed actively to the discussion in a constructive and polite manner, assisting the recorder and moderator in suggesting appropriate text for the CR and in general ensuring a very thorough review of the proposal.

***Working conditions for evaluators***

Compared to national and Nordic evaluations observed, remuneration for evaluators is generous. The time originally foreseen for the discussion of the proposal was two hours, but in the end a full day of discussion proved necessary. Experts had been asked to reserve two full days for the review of the proposal, so this was not observed to cause any major issues. However, one evaluator had an engagement in the afternoon and had to be absent for parts of the discussion regarding the implementation criteria. This was however handled well, as he was able to re-join later in the afternoon and review the assessments agreed and suggest revisions to this before giving his consent.

While the need for extended discussion time was handled well overall, it became a long working day for the experts with very few breaks foreseen and just a short lunch break of 25 minutes. As long hours in front of a screen are more tiring than on-site discussions, breaks could have been taken more frequently. According to a 2021 survey of 3,288 grant reviewers for the US National Institutes of Health, reviewers report shorter attention spans and lower engagement during video grant-review meetings than in those held face-to-face. Compared with in-person grant-review meetings, 46% of respondents said that they paid less attention during the video meetings, and 51% said that their engagement was worse<sup>1</sup>.

<sup>1</sup> Kaplan D, Lacetera N, Kaplan C (2008) Sample Size and Precision in NIH Peer Review. PLoS ONE 3(7): e2761. doi: 10.1371/journal.pone.0002761

**Recommendations**

Recommendations	
<ul style="list-style-type: none"> <li> <p><b>Consider making the briefings carried out prior to the individual evaluation and the consensus discussion more complimentary.</b> Currently the two briefings to some extent duplicate each other and there could be scope for tailoring the two more closely to the two different evaluation phases, with for example the first briefing focusing predominantly on the information relating to the appropriate development of IERs, and the second briefing focusing more on the consensus meeting as such and its output.</p> </li> <li> <p><b>Consider giving feedback on the draft consensus report prior to the consensus meeting.</b> This could lead to more efficient consensus discussions as non-substantial elements such as length of comments, appropriate wording, etc. can be more efficiently handled offline rather than collectively in the panel online.</p> </li> <li> <p><b>Consider including the quality controller in the discussion of the QC comments.</b> This might lead to somewhat more efficient processing of the feedback, as some confusion was observed with respect to the meaning of the comments.</p> </li> <li> <p><b>Consider more frequent breaks in order to tailor working conditions to the online format.</b> Discussing online is more tiring than discussing onsite thus there is a need for more breaks than what would normally be the case when discussion are carried out onsite. Time spent on breaks is likely to be more than compensated for by increased focus and efficiency in discussions.</p> </li> </ul>	

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	15.09.2022	Initial version (new MFF).