



LERU

Evaluation Summary Report analysis

- Data-based LERU recommendations on further optimising proposal evaluation under Horizon Europe Pillar 2

Thank you to the European Research Projects Managers network from the League of European Research Universities (LERU) who supported and contributed to the LERU Evaluation Summary Report analysis.

Core group: KU Leuven (Nele Nivelde and Sara Nelissen), Trinity College Dublin (Emma Siddall), University of Copenhagen (Astrid Cermak and Torben Høock Hansen), University of Freiburg (Antonia Huebner-Kruzinna), University of Milan (Anna Sordi, Beatrice Michetti, Deborah Anna De Luca and Luca Corno)

Contributors and collaborators: Ludwig Maximilians Universität München, Lund University, Sorbonne University, University College London, University of Amsterdam, University of Cambridge, University of Helsinki

Content

| | |
|--|-----------|
| Introduction | 3 |
| The value of ESR analysis..... | 3 |
| Transparency and openness of the European Commission..... | 3 |
| About this analysis..... | 4 |
| Summary of main findings with recommendations | 5 |
| Analysis of excellence criterion with recommendations | 8 |
| Objectives..... | 8 |
| State-of-the-art..... | 9 |
| Technological Readiness Level (TRL)..... | 9 |
| Methodology..... | 10 |
| Do No Significant Harm principle (DNSH)..... | 10 |
| Interdisciplinarity including SSH..... | 11 |
| Gender..... | 12 |
| Open Science including management of research outputs..... | 12 |
| Artificial Intelligence (AI)..... | 14 |
| Other comments under Excellence..... | 14 |
| Analysis of impact criterion with recommendations | 15 |
| Pathways to Impact..... | 15 |
| Scale and significance..... | 16 |
| Barriers..... | 18 |
| Dissemination, Exploitation, Communication (DEC)..... | 18 |
| Intellectual Property (IP)..... | 20 |
| Analysis of implementation criterion with recommendations | 21 |
| Work Plan..... | 21 |
| Risks..... | 22 |
| Resources..... | 23 |
| Consortium..... | 24 |
| Annex A: method and descriptive analysis | 25 |
| Coding..... | 25 |
| Statistical analysis..... | 25 |
| Use of shortcomings nomenclature..... | 26 |
| Excellence..... | 26 |
| Impact..... | 28 |
| Implementation..... | 29 |
| Annex B: comparing award criteria with template guidance and standard briefing slides | 30 |
| Table 1: Excellence - Objectives and State-of-the-Art (including TRL)..... | 30 |
| Table 2: Excellence - Methodology, Interdisciplinarity, Gender, DNSH and Open Science..... | 32 |
| Table 3: Excellence - Artificial Intelligence (AI)..... | 36 |
| Table 4: Impact - Pathways To Impact, Barriers, And Scale And Significance..... | 37 |
| Table 5: Impact - DEC Including IP..... | 38 |
| Table 6: Quality and Efficiency of the Implementation - Work Plan, Risks And Resources..... | 40 |
| Table 7: Quality and Efficiency of the Implementation - Consortium..... | 42 |

Introduction

The value of Evaluation Summary Report analysis

As research managers and administrators, we are always eager to read the Evaluation Summary Reports (ESRs) both when a proposal from the Framework programme has been invited for Grant Agreement preparation and when the proposal is not awarded. We believe that ESRs are a source of information which allow us to provide even better advice for the development of future proposals. Large institutions sometimes try to examine their ESRs and make a broader analysis to try to catch trends or identify systematic mistakes in proposals. However, even organisations with top participation in the Framework programme do not usually have enough ESRs from each part of the programme to be able to assess whether what they find is a one-off or a regular feature in ESRs. Also, it has been very difficult to compare ESR investigations across institutions as each organisation has their own approach to ESR analysis.

To overcome these issues a group of LERU research managers and administrators, the same group who have previously focussed on beneficiaries' first experiences with proposal preparation and submission in Horizon Europe,¹ developed a collaborative and systematic approach to ESR analysis under Horizon Europe. We wanted to help each other create a larger data set of ESRs and create a common framework for analysing the ESRs. This work was completed during 2023, through collecting, coding, and comparing a dataset of 129 ESRs. The ESRs were all above the threshold for funding, and all ESRs were from 2022 proposals under Clusters 1, 2, 4, 5 and 6² in the Global Challenges & European Industrial Competitiveness Pillar of Horizon Europe.³ This was a large-scale undertaking, but the scale and design of our analysis led to conclusions that would have been impossible without this unique collaboration. The conclusions both allow us to further optimise our advice to future applicants under the Global Challenges & European Industrial Competitiveness Pillar and to formulate detailed advice towards the European Commission (EC) on how to further align and optimise the instructional documents in the programme. We are now in a position to disseminate the results of our study, sharing it with our LERU colleagues, with research support colleagues more broadly, with the EC, and with the general public.

Transparency and openness of the European Commission

The current study would not have been possible without the thorough commitment of the EC to the principles of transparency and openness in the design and evaluation mechanisms of Horizon Europe. The European Commission communicates all official documents and guidelines related to the Framework Programme under the reference documents on the Funding and Tender Portal.⁴ Hence, the legal framework, strategic plans and work programmes are freely accessible, the latter including the very important General Annexes, which include award criteria, and the description of evaluation procedures in annexes D-F. Furthermore, the proposal templates and guidelines as well as guidelines and Standard briefing slides for HE evaluators are also transparently and openly available for the public to consult.

These principles of transparency and openness in the Horizon Europe design and evaluation are a great value in the Framework Programme and the European Commission is to be applauded for it. The principles greatly contribute to the general attractiveness of the programme. Also to the credit of the Commission are its continuous initiatives to update and align evaluation and guidance documents based on new insights and developments. For example, the Commission undertook to remove duplicate information on deliverables and work package duration

¹ https://www.leru.org/files/Publications/2022.09_LERU-Report_first-experiences-with-proposal-preparation-and-submission-in-Horizon-Europe.pdf

² The challenges covered by these clusters are Health (Cluster 1), Culture, Creativity and Inclusive Society (Cluster 2), Digital, Industry and Space (Cluster 4), Climate, Energy and Mobility (Cluster 5) and Food, Bioeconomy, Natural Resources, Agriculture and Environment (Cluster 6).

³ Cluster 3 "Civil Security for Society" ESRs were not included in this study since the 2022 ESRs were not communicated yet in December 2022 when the dataset was set up.

⁴ Reference documents for Horizon Europe: <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents?selectedProgrammePeriod=2021-2027&selectedProgramme=HORIZON>

in the work package descriptions in the proposal template and they also included an extra slide on score descriptors (“minor shortcoming”, “shortcoming” and “significant weakness”) from the fifth version of the briefing slides for HE evaluators.

That said, some challenges currently remain on the topic of openness. Given the sizeable amount and diversity of openly available documents regarding requirements and evaluation of the programme, many applicants struggle to find the documents they need. Also, given the multitude of documents introducing, explaining, and referring to award criteria (the annexes, the proposal template, the Horizon Europe Programme Guide, briefing slides for HE evaluators, to name just a few) there is the risk of overlap and of variations in wording, framing, and explaining of award criteria. We will return to this issue under the analysis section.

While there is the wide offer of openly available European Commission documents on programme proposals and evaluation, LERU has also become aware of the existence of additional briefing slides. These slides from topic-specific and consensus briefings can be very specific and are currently only shared with evaluators. Considering transparency and consistency across evaluations it would provide applicants with more complete information on how their proposals are evaluated and how ESRs are produced if these briefing slides were published up front as well.

About this analysis

The analysis presented in this report was conducted by a group of LERU research managers and administrators in three stages.

- 1) **Comparing the award criteria, the proposal template, and the briefing slides for HE evaluators.** For each subcriterion⁵ we compared the wording in the three documents to examine whether the three documents explain the (sub)criteria in a uniform way.
- 2) **Describing the ESRs quantitatively.** We looked at elements such as number of comments or words per sub-criterion and the use of score descriptors, e.g. the so-called ‘shortcomings’.
- 3) **Conducting a qualitative analysis per subcriterion.** We looked for patterns in the comments and words used for each subcriterion and per cluster. For subcriteria where we found difference of wording in the award criteria, the proposal template, and the briefing slides for HE evaluators we also paid attention to whether the differences in wording might have an influence on the comments in the ESR.

In this report, the LERU core group presents an overview of our main findings per award criterion, including recommendations for the European Commission where relevant. At the end of the report, we included two annexes for those interested in more detail. Annex A contains an overview of our methods and a brief descriptive analysis of our ESR sample. Annex B contains tables comparing the text from the award criteria with the proposal template instructions and the briefing slides for HE evaluators.

Disclaimer: The report is based on observations from a particular sample that cannot be generalised to all ESRs. The observations are thus presented as indications, examples. It is also important to note that, while we refer to evaluators throughout the document, ESRs are the result of a consensus-based process also involving rapporteurs and EC moderators. Crucial factors affecting the ESR content, like panel-specific guidance or dynamics, or requirements mentioned in the topic text have not been considered in this analysis.

⁵ By the term *subcriterion* we refer to each of the issues to be addressed under the three award criteria (viz. excellence, impact, and quality and efficiency of the implementation). Subcriteria under the excellence criterion would be *objectives, state of the art, methodology, TRLs, Do No Significant Harm, etc.* The EC’s reviewers’ briefing slides (slide 54 in Version 7) also refer to these elements as the *aspects* or *subcriteria* to be addressed in the ESRs under each of the evaluation criteria.

Summary of main findings with recommendations

Before turning to the detailed findings per award criterion, let us first present the main conclusions deduced from this ESR analysis.

Inconsistency of wording across award criteria, proposal template and briefing slides for HE evaluators

For each award criterion mentioned in Annex D,⁶ we compared the instructions in the proposal template⁷ with the guidance given to evaluators in the briefing slides for HE evaluators.⁸ For most subcriteria the wording used is identical in these three documents. Sometimes the wording for the evaluators is slightly shortened as the information needs to suit the slide format. However, for a handful of subcriteria we found inconsistencies between the description in the proposal template and the briefing slides. For the subcriteria Objectives, Pathways to Impact, and Scale & Significance we point out how the use of different words in the proposal template and in the briefing slides may have led evaluators to comment negatively on elements of the proposal which do seem to be in line with the instructions provided in the proposal template.

For the subcriteria Open Science and Dissemination, Exploitation, Communication the guidance in the proposal template is somewhat disorganised with key concepts hidden in sub-bullets or in documents referenced in the template text, whereas the same criteria are presented in the briefing slides in a more clear and straightforward manner.

Length of ESR and number of comments vary across clusters

We noticed remarkable differences in the length of comments between ESRs from the different clusters of the Global Challenges & European Industrial Competitiveness Pillar. Cluster 1 uses the fewest words for feedback, for both the positive and negative comments. For example, many Cluster 1 positive comments on the Work Plan are quite short and of a very general nature. Cluster 4 ESR comments are also short and use fewer words compared to Cluster 2, 5 and 6 ESRs. The clusters using more words tend to have positive comments on e.g. the Work Plan subcriterion providing more details about what evaluators found commendable (e.g., suitable means to monitor progress, work plan underpins level of ambitions etc.). Although the data show no indications of differences in average length of ESRs reflecting in lower or higher scores, this does show that evaluators proceed differently across the clusters.

Generic vs. specific comments

For the studied ESRs, the average number of positive words per subcriterion is 38 compared to the average number of negative words per subcriterion being 15.⁹ This is to be expected as all the ESRs analysed belong to proposals ranked above threshold.

Despite the higher average number of positive words per subcriterion in our dataset, for some subcriteria the positive comments are brief and very similar to each other, using the keywords from the application form and have a tick-the-box feeling, not adding to a more nuanced understanding of the score. For example, the average number of positive words for the subcriteria Gender (16), Interdisciplinarity (30), and IP strategy (16) is lower than for instance the subcriteria Objectives (66), Methodology (64), or Pathways to impact (86). This variation in positive comments may be related to reviewers' preference for or focus on certain subcriteria, lack of reviewers' understanding or background in certain subcriteria, phrasing of evaluation questions (yes/no questions vs. qualitative questions), or idiosyncratic differences across subcriteria. In the ESR analysis we cannot discern the potential impact of consensus discussion dynamics nor the input and influence of rapporteurs and EC moderators on the use of tick-the-box comments in ESRs, but finding clear differences between clusters and evenness within clusters we do think these factors will have had an influence.

⁶ Horizon Europe. Work Programme 2021-2022. 13. General Annexes. European Commission Decision C(2022)2975 of 10 May 2022.

⁷ Horizon Europe Programme Standard Application Form (HE RIA and IA) ver. 2.0, 21 January 2022.

⁸ Standard Briefing Slides for HE Evaluators, ver. 3.0, 18 March 2022.

⁹ We are rounding to the nearest whole number.

Although, “Lack of clarity” and “lack of details” are standard negative comments across subcriteria and clusters LERU is pleased to note that the above subcriteria that often receive tick-the-box positive comments, will generally receive negative comments that are more detailed and specific. This is very useful as specific feedback can help applicants understand how the proposal could have been improved.

Variation of mean scores across clusters

Our quantitative analysis focussed on the mean scores across clusters, both overall and per section of the proposal. Here we identified considerable variation of both the overall scores and the scores for Excellence, Impact, and Quality and Efficiency of the Implementation across clusters. Evaluators for Cluster 4, for example, gave much lower scores on average, both overall and per section, than evaluators for Cluster 1, 2 and 6. Evaluators for Cluster 5, on the other hand, gave higher scores on average, both overall and per section, than evaluators for Cluster 1, 2 and 6.

At present we have no clear indications that these differences in mean scores have any impact on the evaluation of proposals. Especially for the cluster receiving higher scores, it would be interesting to further explore whether a higher percentage of proposals was ranked based on ex aequo parameters, rather than on (only) the evaluation criteria. This would, however, require a more complete dataset. What we do notice, is that idiosyncratic differences in approach exist across clusters, as was also the case with the average number of words per ESR across clusters.

This accumulation of tendencies across clusters strongly suggests that the cluster specific and topic specific guidance, which is not publicly available but which LERU understands is communicated to specific evaluator panels, has a concrete impact on how evaluators proceed and on how ESRs are written across clusters and topics. Although it seems common sense to acknowledge that subcriteria may receive (slightly) different interpretations or even different importance across clusters and topics, it would be advisable for the European Commission to communicate transparently about these cluster specific and topic specific instructions to evaluators. This would further raise transparency of evaluation procedure and expectations regarding award (sub)criteria.

Score descriptors are only used in half of the negative comments

The briefing slides for HE evaluators Version 3.0, which was available at the time of the spring 2022 evaluations when the ESRs in our sample were written, introduced the score descriptors “minor shortcoming”, “shortcoming” and “significant weakness”, but did not offer detailed information on the definition of these terms.¹⁰ When looking at the negative comments in our data, evaluators indicate whether a negative comment is a “minor shortcoming”, “shortcoming” or “major shortcoming” in less than half of the comments. We noticed that the EC terminology “significant weakness” did not occur in the dataset, but “major shortcoming” was used instead. Indicating the level of the shortcoming improves the understanding of what evaluators value negatively, and of the relative weight given to this negative comment.

For the ESRs that *did* make use of the score descriptors, we recorded that proposals with “shortcomings” mentioned in the Excellence section did not score 5 in “Excellence“. We recorded less than ten “major shortcomings” in the sample, which might be due to the fact that all ESRs surveyed were above threshold.

We find it very positive that clear instructions on how to actively use score descriptors have now been included in the expert briefing slides, since Version 5. This will increase standardisation and make the evaluation process more transparent to applicants. We do note that these more recent instructions state that labelling an element in a proposal as a “significant weakness” “will lower the score below threshold” - which was clearly not the case with the proposals with “major shortcomings” in our dataset.

Comments out of scope

In a small number of ESRs, we found comments that seemed to be either misplaced or completely out of scope of the evaluation criteria. We list some examples below:

- The positive comments on Implementation include praise for the strategy on IP management which is a subcriterion that belongs under the Impact section.

¹⁰ A slide on ‘score descriptors’ was included from ver. 5.0 used for the 2023 spring evaluations.

- Commenting on the subcriterion Gender under Methodology, one ESR comments positively on the gender composition of the team, which is not relevant to evaluate for the subcriterion.
- A comment about sub-contracting: “The capacity of the subcontractor to carry out this task is unknown. This is a major shortcoming.” It is not possible to describe the capacity of the subcontractor when the project cannot select the subcontractor prior to signing the Grant Agreement.
- Some ESRs criticise projects for showing differences in the distribution of resources between work packages, which is concerning as resources should be distributed between work packages according to the planned activities. Even distribution of resources is not an award criterion and should therefore not be used as an argument to penalise a project.
- Regarding project management procedures: “The Project Coordinator plans to hire a Project Manager but specificities regarding the required tasks, skills, and overall characteristics of contractual relationship are not clear enough.” Again, a very strange comment as most projects will hire a project manager and their employment is not an award criteria.

Although we cannot measure the exact impact of these misplaced or out-of-scope comments on the project scores, they are more than likely to have some impact on the score, especially if deemed to be a “major shortcoming”.

Overall recommendations to the European Commission

- 1.1. To increase the clarity and consistency of instructions, **further streamlining is needed between the guidance given in the proposal template and in the briefing slides for HE evaluators.** The need is particularly significant in five aspects of the award criteria: Objectives; Open Science; Pathways to impact; Scale and significance; and Dissemination, exploitation, and communication.
- 1.2. Our analysis of 129 Evaluation Summary Reports also reveals the **need for further ensuring the quality of the comments** by placing a greater **emphasis in the training of HE evaluators** and other involved personnel (rapporteurs, EC officers, Quality Controllers). Three identified areas in this respect are:
 - the inclusion of clear justifications not just for weaknesses but also for strengths;
 - the consistent and harmonised use of score descriptors (minor shortcoming, shortcoming, significant weakness), and;
 - the avoidance of misplaced, out-of-scope and tick-the-box type of comments.
- 1.3. Our findings also reveal a **high variability of evaluation practices across clusters**, an issue that could be addressed by implementing **more uniform guidance across the Framework Programme** thus contributing to its robustness and transparency. **Any cluster and topic specific evaluation instructions should also be communicated transparently** upon publication of the work programme or opening of the topic.

Analysis Of Excellence Criterion With Recommendations



Figure 1: Excellence positive (blue) and Excellence negative (black) comments - 25 most frequent words. (made in Pro Word Cloud add-in for PowerPoint).

Objectives

The first excellence subcriterion we analysed, is clearly seen as very important to evaluators as objectives are always commented upon, be it positively, negatively, or both. All ESRs had positive comments and just over half also had at least one negative comment. The subcriteria of Objectives have one of the highest average numbers of positive words used for commenting (67).

When comparing the proposal template wording and the briefing slides for HE evaluators we found that the concept of “clarity” is included in the evaluators’ guidance but not in the proposal template where being “brief” is stressed. This may have led to a discrepancy between what the applicants try to achieve and evaluators expect to see. Applicants focus on being “brief”, as is asked for in the template, and evaluators find fault in the lack of clarity of the objectives. “Lack of clarity” and “insufficient detail” (or similar) was mentioned by the evaluators in half of the negative comments. In fact, most of the negative comments for the Objectives subcriterion criticised that proposals were not “painting the whole picture”, specifying that the proposal lacked a more thorough benchmarking, literature review, or integration of theory into methodology. Around half of the negative comments for this subcriterion were labelled as “minor shortcomings”.

In line with the template guidance, evaluators also want to know how the project reaches its objectives and refer to different concepts of “measurability”. For example, ESRs ask for the use of Key Performance Indicators or state that the objectives are “not clearly measurable” or KPIs are “insufficiently specified to enable precise monitoring”. While we understand why HE evaluators might find this relevant, monitoring and measurements of project progression are part of the implementation section (milestones and deliverables). As applicants are encouraged to be brief under the current subcriterion, they might be hesitant to provide the same information here and under implementation and decide to adhere to the brevity advice in this section.

Recommendations to the European Commission for the Objectives subcriterion

- 2.1. The expert briefing slides and the template guidance should be aligned. The template explanations should include the concept of “clarity”, if this is something evaluators focus on, and it is mentioned in the briefing slides.

State-of-the-art

For the state-of-the-art subcriterion, there is coherence between the award criteria, the guidance in the proposal template and the briefing slides for HE evaluators. The key words “ambitious”, “novel”, and “new” are used in the same way across the three documents.



Beyond state-of-the-art (B-SOTA) is described at length in all ESRs, often giving detailed examples of what elements in the proposal are seen as ambitious and/or B-SOTA. In the ESR comments B-SOTA is very often related to objectives, methodology or Technology Readiness Level (TRL) when evaluating the level of ambition and innovation. This is clearly not a box which is just ticked off by evaluators.

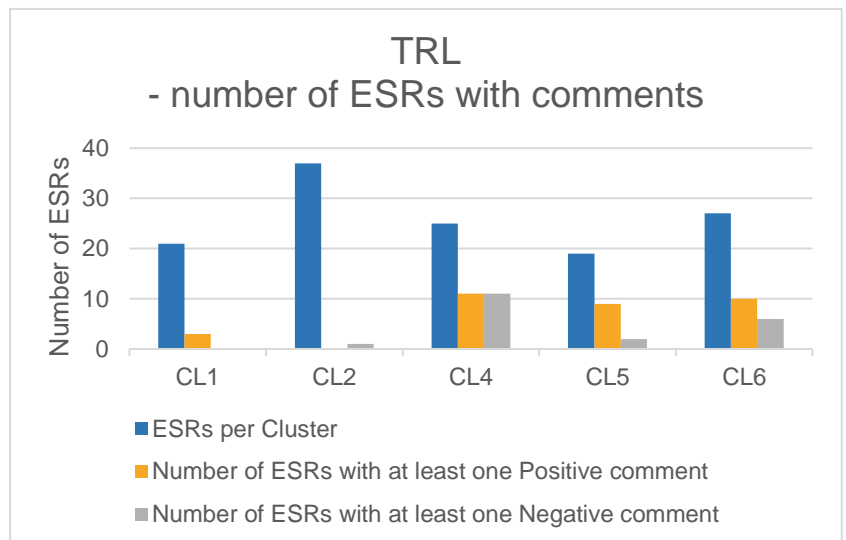
Only 6 ESRs contained negative comments and no positive comments. 4 of those scored 3 or 3.5 in Excellence, and two scored 4. Only one of the 6 proposals was funded, suggesting that getting only negative comments on this subcriterion could make it difficult for a proposal to be funded.

Figure 2: B-SOTA positive comments, adverbs and adjectives, 25 most used words (made in Pro Word Cloud add-in for PowerPoint).

Technology Readiness Level (TRL)

TRL is not mentioned under the award criteria, but template guidance and the expert briefing slides seem to be aligned on the topic, both referring to level of Research and Innovation (R&I) maturity and warning against the differing expectations regarding R&I maturity in e.g. Research and Innovation action (RIA) and Innovation action (IA) types of projects.

It is evident by the number and length of the comments under the TRL subcriterion that this is a concept of greater relevance to some clusters than others. For example, comments on TRL in Cluster 4 are more detailed (using more words) than in Cluster 2 and Cluster 1, where only a few ESRs comment on TRL. However, it seems that TRL is not a pivotal aspect in the evaluation outcome, even in Cluster 4. None of the 11 Cluster 4 ESRs containing comments on TRL were funded, even though 6 of these ESRs had positive TRL comments only, while 5 ESRs had both positive and negative TRL comments.



Methodology

Award criteria, template guidance and briefing slides seem properly aligned on the topic of methodology, all three of them referring to the clarity of scientific methodology, including concepts, models and assumptions underpinning the work.

As with the subcriteria Objectives and Beyond state-of-the-art the subcriterion of Methodology was always commented on in the ESRs examined.

Looking at the positive words used to describe methodology in the ESRs, they are generally correlated to the score attributed to the excellence section. A word cloud analysis of the positive comments shows that the most frequently used adjective or adverb to comment on the methodology subcriterion in ESRs scoring 5 for the Excellence criterion is “excellent”. For ESRs scoring 4 on Excellence the most common word is “well” and for ESRs scoring 3 on Excellence the most frequent word is “sound”. Also, the number of positive words used to comment on Methodology is high (64). This indicates that the evaluators find the methodology assessment important for the Excellence score.

Often evaluators provide targeted and proposal specific feedback in the ESRs, detailing how the methodology is appropriate or not. However, there is still a fair share of very general comments along the lines of “The methodology is sound, detailed and convincing”.

Recommendations to the European Commission for the Methodology subcriterion

2.2. Revise the evaluation question in the briefing slides to elicit a specific comment from the evaluation panel, rather than a “tick-the-box” comment, as currently still occurs in some ESRs. This clarification may look as follows:

Is the scientific methodology (i.e. the concepts, models and assumptions that underpin the work) clear and sound? What elements contribute to (lack of) clarity and soundness?

Do No Significant Harm principle (DNSH)

DNSH is briefly mentioned in the methodology section of the proposal template but the inclusion of DNSH in the proposal text is not mandatory for applicants, so DNSH is **not** an official subcriterion for evaluating the Excellence section. However, the standard evaluation form HE RIA and IA Version 2.0 includes the DNSH principle under “Other questions” as shown at the picture to the right.

This means that evaluators have to indicate whether a proposal complies with the DNSH principle on a scale ranging from “No”, to “Partially”, to “Yes” and the two additional options “Not Applicable” and “Cannot be assessed”. Although the briefing slides state the DNSH principle and refer to the six environmental objectives, they do not provide detailed instructions or criteria on how to assess the DNSH principle. The slides refer to a support video on the DNSH topic on the Funding and Tender Portal, which is not available at the time of writing this report. So exactly how the evaluators decide between “no” and “partially”, for example, is left to their own discretion and needs to be justified by the evaluators in a comment box. This might lead to evaluators choosing “Not applicable” or “yes” in order to avoid further commenting.

In our sample, evaluators seemed to respect the fact that DNSH is not an evaluation subcriterion and only in 5 ESRs did we find comments on DNSH. There was only 1 negative comment related to DNSH, and there were also

Do no significant harm principle

Is this proposal compliant with the 'Do no significant harm' principle?

Not applicable
 Yes.
 Partially
 No
 Cannot be assessed

If Partially/No/Cannot be assessed please explain.

no comments labelled as “shortcomings”. This underpins the instruction given to evaluators in the briefing slides not to score the proposals with respect to DNSH.

Recommendations to the European Commission for the DNSH criterion

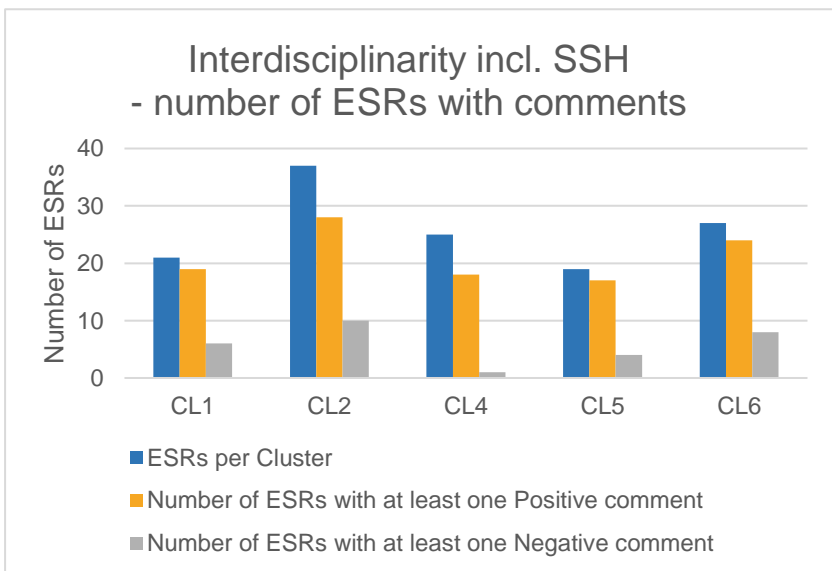
2.3. The expert briefing slides should specify in what cases the DNSH principle “cannot be assessed” and when the DNSH principle is “not applicable”. The slides should also provide guidance for evaluators on how to determine whether a proposal is to be considered under the “no”, “partially” or “yes” categories.

2.4. The DNSH principle could meaningfully be integrated into the ethics table in Part A.

Interdisciplinarity including Social Sciences and Humanities (SSH)

The Interdisciplinary Approach is mentioned in the award criteria, while the SSH integration is not. The template guidance covers both the Interdisciplinary Approach and SSH integration, specifying that reference to SSH integration is mandatory for SSH flagged topics and indicating that this should be included under the interdisciplinary approach subcriterion. The expert briefing slides state that evaluators should consider “how expertise and methods from different disciplines will be brought together and integrated in pursuit of the objectives”. However, the evaluators are not given further information on appropriate interdisciplinary integration methods. Recent research has shown that one of the challenges to achieving successful interdisciplinary research is the education of evaluators on how to score interdisciplinary proposals.¹¹ Therefore, we recommended in a previous report¹² that the European Commission considers providing further guidance to evaluators on this subcriterion.

Interdisciplinarity and/or multidisciplinary are required elements in almost every topic of the work programmes under the Global Challenges & European Industrial Competitiveness Pillar. We therefore expected evaluators to invariably comment on this subcriterion, if only to affirm the projects’ compliance with this requirement. However, in around 14% of the analysed ESRs this subcriterion is commented on neither positively or negatively.



Positive comments mainly focus on the excellent integration and combination of different disciplines. Quite often it is the project itself that is multi- or interdisciplinary in nature. Other positive comments are about how the interdisciplinarity and the SSH contribution are adequately considered and accompanied by clear and convincing explanations. Negative comments mirror the positive ones and very often stress that the integration of social science and humanities is insufficiently explained, or that the proposal does not sufficiently explain how methods from different disciplines will be brought together and integrated in pursuit of the project objectives.

Overall, when there are ESR comments on interdisciplinarity, including SSH integration, they seem to be congruent with the provided guidelines. Indeed, in the ESRs evaluators account for the integration of the different disciplines and the correct use of SSH disciplines in achieving the project objectives. However, as far as SSH integration is concerned, for an optimal analysis of this subcriterion to be possible, we would need to take into account the demands of each individual topic, which is beyond the scope of this report.

¹¹ <https://www.shapeidtoolkit.eu/wp-content/uploads/2021/05/Top-ten-tips-evaluation.pdf>

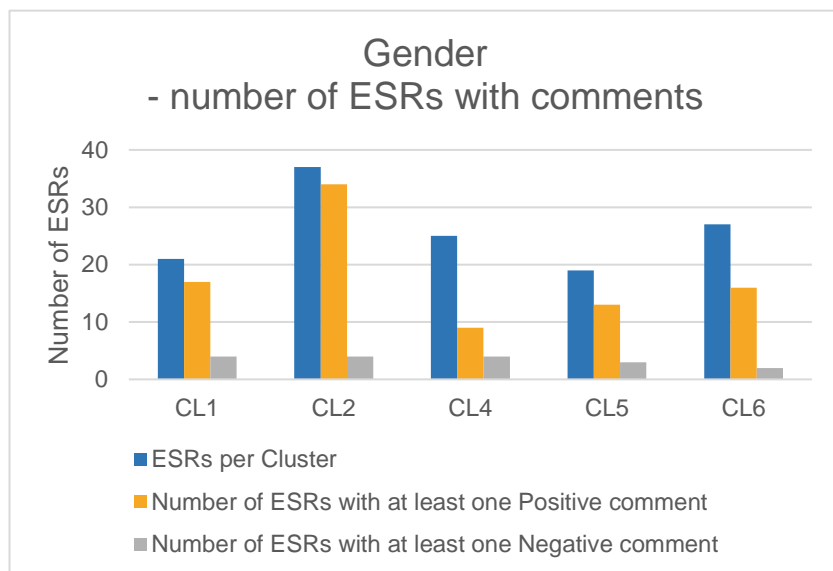
¹² LERU Report on first experiences with proposal preparation and submission in Horizon Europe (2021) Available at www.LERU.org In addition, further clarity is needed on how the cross-cutting priorities are considered during evaluation. P.8

Recommendations to the European Commission for the Interdisciplinarity including SSH subcriterion

2.5. We recommend that the EC develops a slide dedicated to the evaluation of interdisciplinary research to be included in the Standard briefing slides for HE evaluators.

Gender

The gender dimension in the R&I content is referred to in a uniform way in the award criteria and in the template guidance and briefing slides.



Most ESRs contain at least one comment on the gender/sex aspects of the research/research methodology in the proposals. Around half of the comments on gender were rather generic comments simply stating whether gender was considered in the proposal. The other half of the comments were more specific comments highlighting aspects of the project to which gender related (e.g., general strategy, research activities, dissemination, stakeholder engagement, composition of themes, design of the empirical studies, different measures of inclusion and sustainability) and pointing to precise examples referring to the project (e.g., “environmental pollutants such as endocrine disruptors that indeed pose gender-specific hazards”). Overall, the negative comments had a slight tendency to be more specific.

environmental pollutants such as endocrine disruptors that indeed pose gender-specific hazards”). Overall, the negative comments had a slight tendency to be more specific.

Recommendations to the European Commission for the Gender subcriterion

2.6. Revise the evaluation question in the briefing slides in order to elicit more specific reviewer comments, specifying why and for what aspects the gender dimension has (not) been considered. This will help to optimise applicants’ grasp of the subcriterion of Gender. This specification may look as follows:

*Has the gender dimension in research and innovation content been properly taken into account? **Why and for what aspects has it (not) been taken into account?***

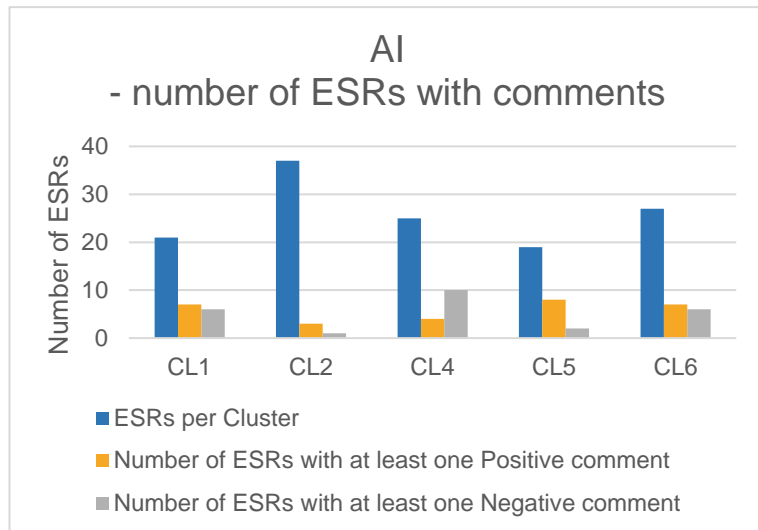
Open Science (OS) including management of research outputs

The Award criteria and the briefing slides for HE evaluators both contain the elements of ‘open science practices’ and ‘management of research outputs’ (although named ‘research data management’ (RDM) in the briefing slides). ‘Engagement of citizens, civil society and end-users’ in the award criteria, however, seems to have disappeared into ‘all open science practices beyond mandatory’ in the briefing slides.

When looking at the template guidance in the proposal it is a very information-dense section, introducing a great number of concepts related to Open Science, such as sharing of research, research output management, reproducibility of research outputs, open access, open peer-review, involving stakeholders, co-creation of R&I, and a similarly sizeable enumeration follows under the data management section. All of these concepts, many of which are still not yet fully familiar to most researchers, are only briefly introduced in the template guidance, with (sometimes) an example or a brief explanation, in a rather heterogeneous list of bullets, enumerations and exclamation mark bullets (⚠️). We feel this might not be the optimal approach to instruct applicants on the Open Science

to implement one or more of the recommended open science practices, e.g., citizen science, co-creation approaches.

Artificial Intelligence (AI)



AI has been part of the briefing slides for HE evaluators since the beginning of the Horizon Europe programme. However, it was not a part of the template guidance version 2.0, which was used by applicants in the spring of 2022, when the proposals whose ESRs we analysed were written.¹⁴ At this time AI was only mentioned in the ethical self-assessment in Part A.

Looking at the evaluator briefing slides there are several aspects of AI which evaluators are expected to consider as additional questions in the evaluation form, including technical robustness, accuracy, and reproducibility of developed or used AI systems, their social robustness, their reliability and function as intended and their ability

to provide a suitable explanation of the decision-making process.

Cluster 4 contains the most ESRs with comments on how to integrate the AI dimension, Cluster 2 has the fewest ESRs which comment on AI. However, both positive and negative AI comments are found in ESRs from all clusters. The Cluster 4 ESRs contain very detailed comments on AI robustness and its components and in Cluster 5 evaluators often focus on testing/validation of the robustness of the AI system. Cluster 6 is the only cluster in which evaluators commented on integration of AI in the methodology.

Recommendations to the European Commission for the AI subcriterion

2.10. Applicants and evaluators will require more thorough guidance on when it is relevant to deal with, consider, and specify AI in detail in future proposals. It is important to have guidelines both for development of AI in projects but also for projects where AI systems are integrated in the methodology. As AI is a moving target and contains both ethical and technical aspects it is important to regularly update the guidelines in both the template and the standard slides for HE evaluators.

Other comments under Excellence

We made a category for Excellence comments that did not fit the award criteria from the evaluation matrix. The eighteen positive comments mention e.g. multi-actor approach, international collaboration and New European Bauhaus as elements covered in the proposal. “Important strength”, “significant for the proposal”, “bring substantial assets to” are some of the statements leading us to think that these elements might have had a positive influence on the score of the evaluation. For the four negative comments, these elements are mentioned as “minor shortcomings” and did not seem to affect the score. Ethics is mentioned once as a negative comment.

¹⁴ The proposal template was updated in November 2022 (ver. 3.2) to include text on artificial intelligence (AI) and how to deal with AI in the proposal. This text is almost identical to the AI instructions provided in the briefing slides for HE evaluators.

Analysis Of Impact Criterion With Recommendations

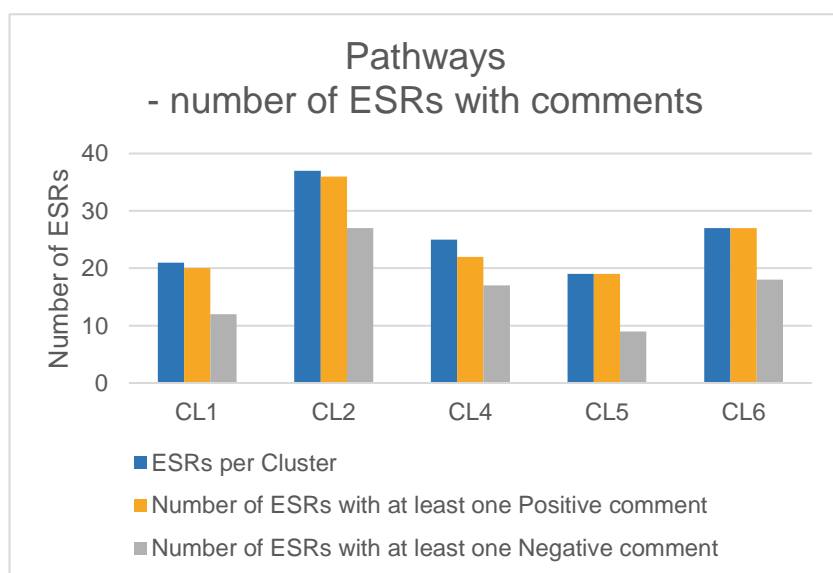


Figure 4: Impact positive (blue) and Impact negative (black) comments - 25 most frequent words. (made in Pro Word Cloud add-in for PowerPoint).

Pathways to Impact

The template guidance and the briefing slides for HE evaluators both focus on the need for the project to “illustrate its **contributions to the expected outcomes** specified in the topic as well as the wider impacts in the respective destinations”; while the award criterion instead requires the analysis of the “credibility of the pathways to achieve the expected outcomes and impacts specified in the work programme” without highlighting that the expected outcomes are specified at topic level and expected impacts are specified at destination level. Also, the template guidance alternates between the terms “wider expected impacts” - cf. “expected impacts” as in the topic texts - and “**wider impacts**”, printed in bold. This is potentially confusing.

Our analysis found that only 30% of the reviewers' comments in the impact section referred to the phrase “pathways to impact”, which may further speak to the discrepancy between the terminology used in the three EC documents.

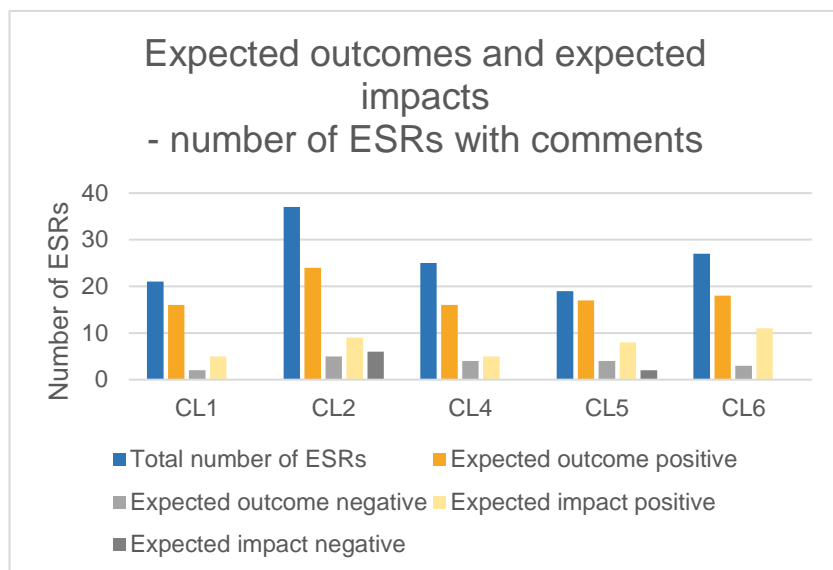


The difference in the terminology used between the award criteria, the expert briefing slides and the template guidance may also create a greater focus on the topic level expected outcomes rather than the destination level expected impacts. This discrepancy is evidenced by the fact that 68% of the ESRs commented positively on the contribution the project would make to the topic level expected outcomes, whereas only 28% included positive comments on the contribution to destination level expected impacts. When the same analysis was conducted for negative comments, 13% of the ESRs included negative comments on the pathways to topic

level expected outcomes but only 6% included negative comments on pathways to destination level expected impacts.

While Pathways to impact is commented upon by evaluators in all ESRs. We noticed that the impact summary table (Section 2.3 of the proposal) is rarely directly mentioned in the ESR comments but most of the points contained in this table are the object of evaluation in the impact section, and specifically the pathways to impact. There is

a significant negative correlation between the number of negative comments and the impact score, which means that a higher number of words in negative comments on Pathways to Impact reflects in a lower impact score.



When we examined this analysis by cluster we noted that the number of positive comments on expected outcomes was fairly even across the clusters. However, the negative comments on expected outcomes showed some variation with 20% of the Cluster 5 ESRs receiving criticism in this area whereas only 9% of Cluster 1 ESRs received negative comments on expected outcomes.

For expected impacts the number of positive comments were highest for Cluster 5 and Cluster 6 at 40% and 42% respectively, whereas only 19% Cluster 4 ESRs received positive comments in this category. The number of negative comments on expected impacts shows the greatest variation with no comments received for the Cluster 1, 4 or 6 ESRs, but 16% of the Cluster 2 ESRs and 10% of the Cluster 5 ESRs had negative comments.

Therefore, our analysis suggests a strong variation across the clusters in the level of importance placed on the contribution of the project to the expected outcomes and impact, with the greatest importance placed in Clusters 2, 5 and 6. Although this variation may be at least partly due to varying ways of impact-thinking across clusters, and different types of evaluators being involved under different clusters, it may also be the result of different emphasis placed by the different EC project officers or Research Executive Agencies during the evaluator briefing sessions. For that reason, we plead for a greater degree of transparency around this process.

Recommendations to the European Commission for the Pathways to Impact subcriterion

- 3.1. There should be better alignment of the wording in the award criterion, the evaluators' briefing slides, and the template guidance. All three official documents should mention the phrase "pathways to impact" and should highlight that these pathways should map expected outcomes at the topic level and expected impacts at the destination level.
- 3.2. There is a need for a greater degree of transparency on the emphasis placed on the alignment of proposals with expected outcomes versus expected impacts in evaluator briefing sessions across the different clusters.
- 3.3. Evaluators should be guided with a more detailed "Pathways to impact" evaluation method, so that this criterion can be properly addressed.
- 3.4. Additional destination specific supporting guidelines should be provided to evaluators to better verify the compliance of impact in respect to work programme destination level expected impacts. These guidelines should be made publicly available upon the opening of the relevant calls for proposals.

Scale and significance

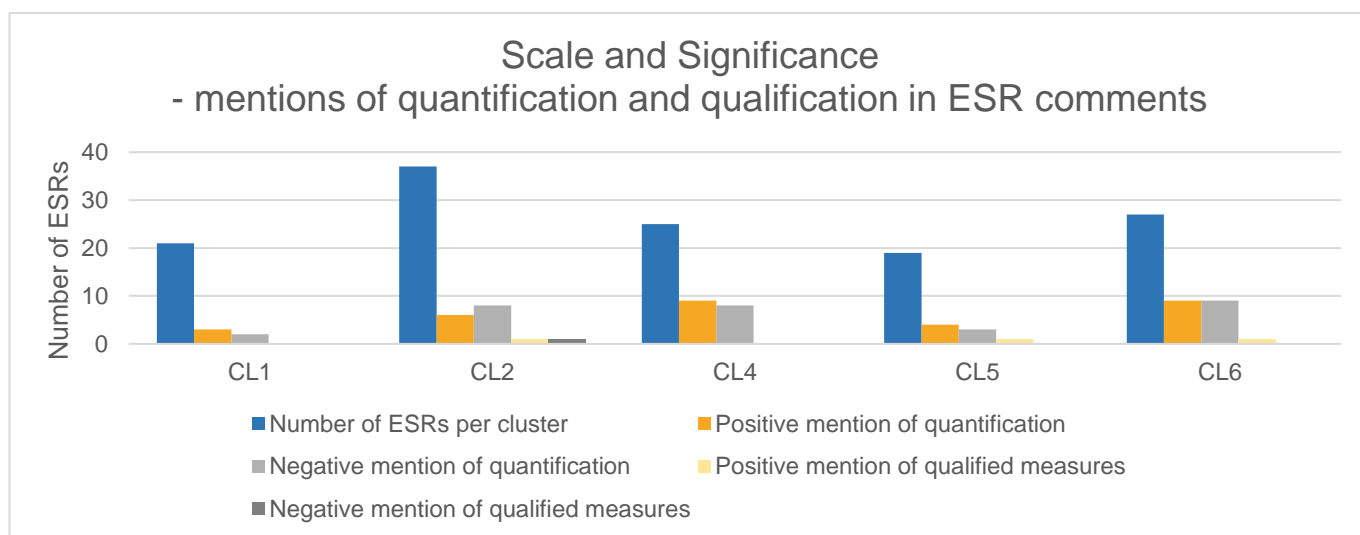
In our analysis we noted a rather significant difference between the wording in the documents guiding proposal writing and evaluation and seek that future documents make this wording consistent, see Annex B, Table 4. The award subcriterion mentions "scale and significance [...] contributions", and the template guidance "indication of the scale and significance [...] contribution [...]". Provide quantified estimates where possible and meaningful". The expert briefing slides, however, pose the question "Are the scale and significance [...] estimated and quantified [...]". This discrepancy between the three documents on the need for quantification of scale and significance

measure appears to have a bearing on the number of negative comments received under this subcriterion, especially for disciplines which tend to be more qualitative in nature, such as those funded under Cluster 2 and 6.

Scale and significance were commented on positively in 89 ESRs and negatively in 57 ESRs. When the clusters were compared, we found that scale and significance was commented on positively more often in Cluster 1, 2 and 6 when compared to the other clusters. Negative mentions of scale and significance were higher in cluster 2 and 6 when compared to the other clusters, but no significant difference was found across clusters. However, we found a statistically significant negative correlation between the number of negative comments and the impact score, which means that a higher number of negative comments on scale and significance reflects in a lower impact score.

Cluster 2 was also the only cluster where there were more negative than positive comments on the lack of quantification of scale and significance. The quantification of scale and significance seems least important in Cluster 1 and most important in Clusters 2, 4 and 6.

31 ESRs included positive comments on the use of quantification of Scale and Significance measures. However, only 3 comments positively mention the use of qualitative measures for Scale and Significance. 31 ESRs included negative comments on the lack of quantification of Scale and Significance measures, and one comment explicitly criticised the use of qualitative rather than quantitative measures of scale and significance.



The findings above illustrate (yet again) the danger of differences in wording and message in award criterion, template guidance and briefing slides. On top of that, it illustrates the risk of qualitative things being measured by quantitative indicators - a danger which also the COARA Agreement on Reforming Research Assessment¹⁵ warns against. We feel that it is particularly important to address the issues highlighted above as there is a significant negative correlation between the number of negative comments and the impact score.

Recommendations to the European Commission for the Scale and Significance subcriterion

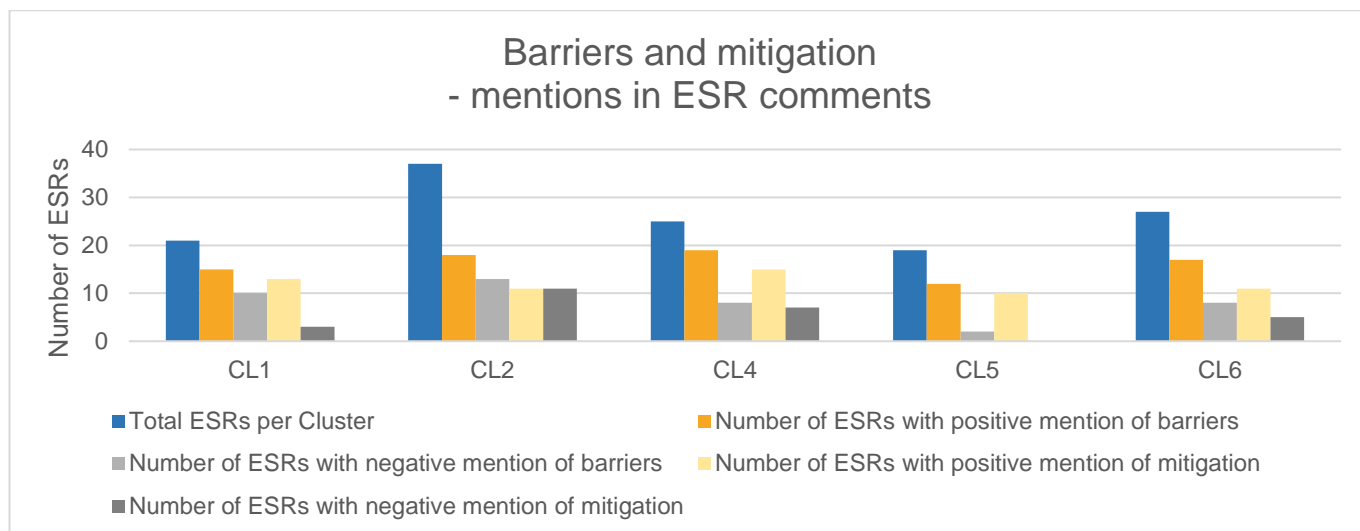
3.5. While we recognise that the European Commission wants to ensure that proposals demonstrate the level and value of contribution a potential project will make to the expected outcomes and impacts for call topics, we recommend that further consideration is given to how these contributions can be assessed, using both qualified and quantified measures. This is particularly important as we seek to ensure the further integration of SSH disciplines (or even Arts, Humanities, and Social Sciences (AHSS)) into proposals, many of whom engage in qualitative research.

3.6. We recommend that the European Commission reconsiders the wording used in the award criteria, template guidance and expert briefing slides, making it clear that measurements must be included in this section but specifying also that it is accepted that quantitative measures are not always appropriate and that for certain types of projects qualitative measures can be accepted as well.

¹⁵ https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

Barriers

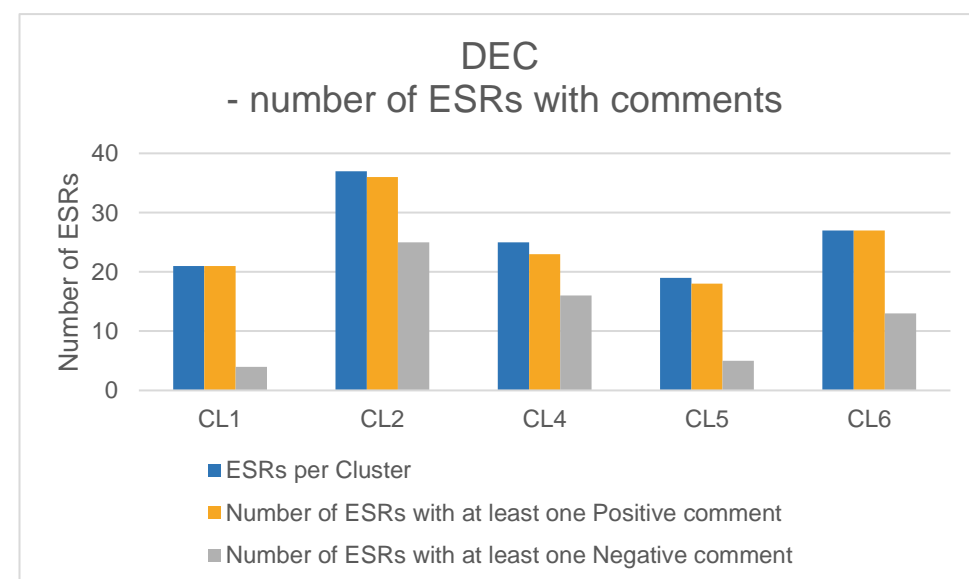
Barriers are not mentioned as a subcriterion in the award criteria but we noted from our ESR analysis that in order to score well a proposal needs to address both potential barriers to project impact as well as appropriate mitigation strategies. Cluster 2 received the highest number of negative ESR comments for this subcriterion, with both the correct identification of barriers and realistic mitigation measures being criticised. In Cluster 1 most of the negative comments criticised the barriers identified rather than mitigation. Only 2 ESRs in Cluster 5 included a negative comment on barriers and no comments on mitigation.



Dissemination, Exploitation, Communication (DEC)

The subcriterion Measures to maximise impact - Dissemination, Exploitation and Communication describes the necessity to explain the suitability and quality of the measures to maximise impact, as set out in the plan for DEC. The template guidance reiterates the necessity to come up with a first plan for DEC and states the need to describe the target groups addressed. Next, in a heterogeneous list of sub-bullets on various aspects of the plan for DEC, a third sub-bullet conveys key instructions to applicants on what to also include under this section, viz. the need for the measures to be proportionate, concrete and to extend during and after the project’s lifetime. Measures also need to be best suited to reach the target group addressed. The need for the measures to be of good quality mentioned in the briefing slides as well as the award criteria is not resumed or explained in the template guidelines

- apart from the fact that measures need to be concrete. All aspects referred to jointly in the award criterion and the template guidance are described in the briefing slides as elements to be considered for the evaluation of this subcriterion.



The DEC-subcriterion is almost always commented on in the ESRs: only four of the ESRs we analysed did not receive a single positive comment on DEC. Over half of the ESRs received at least one negative com-

ment. Proposals under all clusters received more positive than negative DEC comments, and the proportion of negative DEC comments is the highest under clusters 2 and 4. There is a significant negative correlation between the number of negative comments and the impact score, which means that a higher number of negative comments on DEC reflects in a lower impact score.

Within the DEC comments we noticed a number of interesting trends, so we recorded the number of references to frequently recurring themes under this section in the ESRs, as displayed in the X-axis in the graph below. The graph shows the relative occurrence of each of these themes under the DEC section. We recorded frequent positive comments on clarity and measurability (KPIs) - both aspects of the *quality* of measures -, as well as on suitability of the measures and on the described target groups.

Positive comments on Dissemination, Exploitation and Communication were counted separately, but they jointly make up about 25% of all positive comments. In the negative DEC comments there is a strikingly high number of comments on exploitation measures.

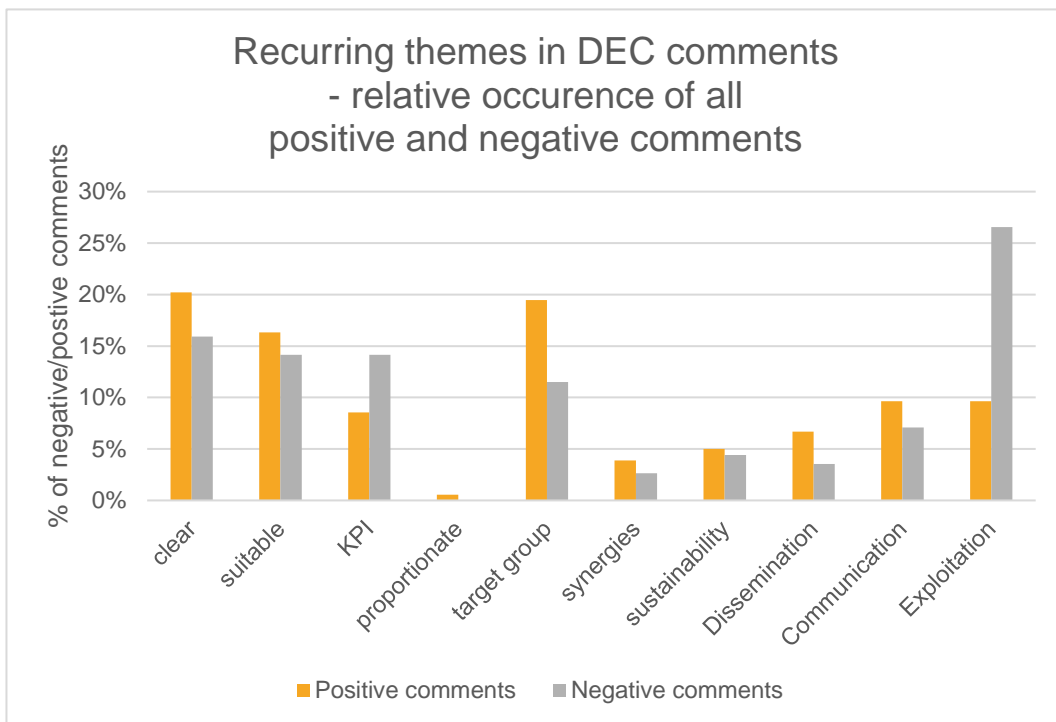
Under the measurability theme (KPIs to monitor success of the measures) we also found more negative than positive comments, stating that KPIs were either missing or dysfunctional.

Other negative comments focus regularly - as can be expected - on clarity and suitability of the measures or on description of target groups. Proportionate measures, measures during and beyond the project lifetime and synergies (i.e., references to collaborations and synergies with previous or other related projects or initiatives, an aspect which is not included under the award criterion or template guidelines but sometimes mentioned in the topic text) make out a small amount (5% or less) of both positive and negative DEC comments. Notice that none of the ESRs included any negative comments on the point of measures being proportionate to the scale of the project.

We conclude that the themes referred to only in the sub-bullets in the template guidance (proportionate, during and beyond the project lifetime) are referred to less in the ESRs than the themes introduced in the award criterion itself and in the first general section of the guidance (clarity (including measurability), suitability, target groups, and DEC-measures).

The data furthermore revealed a different approach between evaluators across clusters: Cluster 1 and 2 ESRs write positive DEC comments primarily in a joint approach, whereas Cluster 4, 5 and 6 ESRs write positive DEC comments mostly separately for Dissemination, Exploitation and Communication. We cannot at this point conclude whether this different approach has an effect on the impact score, but it reflects a lack of uniformity in evaluation approach across clusters and might be due to cluster specific informal or written instructions made available to evaluators.

Further comparison across clusters shows us that the ESRs commented most positively on clarity, suitability, and target groups in Cluster 1, 2, 4, 5 and 6. Cluster 4 positive comments focus more often on the use of KPIs than is the case in Cluster 1, 2, 5 and 6.

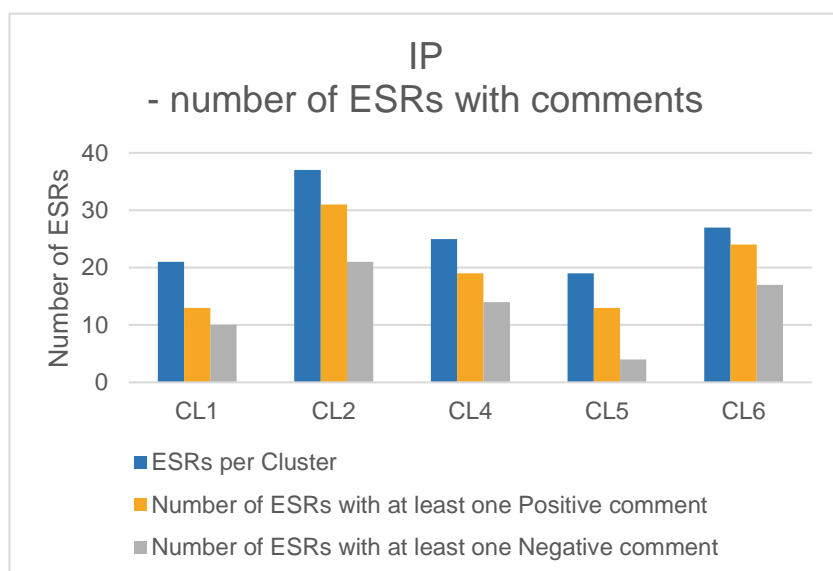


Recommendations to the European Commission for the DEC subcriterion

- 3.7.** For reasons of user friendliness and transparency, it seems advisable that all key elements that are to be described under the DEC part of the template, are included in the main text of the template guidance, and not distributed over the main text and a bullet list containing other types of information on the DEC sub-criterion.
- 3.8.** There is a need for more transparency within the guidance on evaluating proposed strategies for dissemination, exploitation and communication in evaluator briefings across the different clusters.

Intellectual Property (IP)

The award criterion for this section does not explicitly reference the IP strategy, but it can be inferred that it falls under the suitability assessment for DEC measures. The template guidance asks applicants to outline their IP management strategy and gives examples of IP protection such as patents, and copyright. Applicants are also asked to highlight how this IP strategy will support the exploitation of results. The expert briefing slides asks if the IP strategy is properly outlined and suitable to support the exploitation of the project results.



Unsurprisingly most of the Cluster 4 ESRs received comments under this subcriterion, however it is surprising that only 64% of the Cluster 1 and 70% of the Cluster 5 ESRs received comments under this section. 30 ESRs, close to 25%, did not contain any comments on IPR.

Positive comments on IP praise the plan for “being suitable and well considered”. A third of the 77 positive ESR comments praised the inclusion of an IP strategy that supports the exploitation of results, which matches the wording of the guidance documents. In Cluster 1, it was two-thirds of the positive comments that praised the inclusion of an IP strategy that supports the

exploitation of results. In Cluster 5, however, only one of the 11 positive comments mentions that the IP strategy supports exploitation of results. The rest of the comments are of a more general nature, for example “IPR has been appropriately addressed.”

The alignment of the IP strategy with open science principles was the third most common positive comment under this subcriterion, however the wording of the guidance in the template and in the expert briefing slides do not explicitly make the connection between open science and IP.

The negative comments under this section focused on insufficiently addressed IP strategies, which were often criticised for being generic, and we often noted criticisms that the IP strategy was not suitable to support the exploitation of the project results.

Recommendations to the European Commission for the IP subcriterion

- 3.9.** We recommend that the wording of the template guidance and expert slides are amended to include mention of open access approaches such as CC BY licenses to support the exploitation of project results if appropriate.
- 3.10.** We also suggest that the template guidance should be revised to direct applicants to consider the connection between the Open Science described in Section 1.2 and the IP strategy devised in Section 2.2.

Analysis of Implementation Criterion with Recommendations

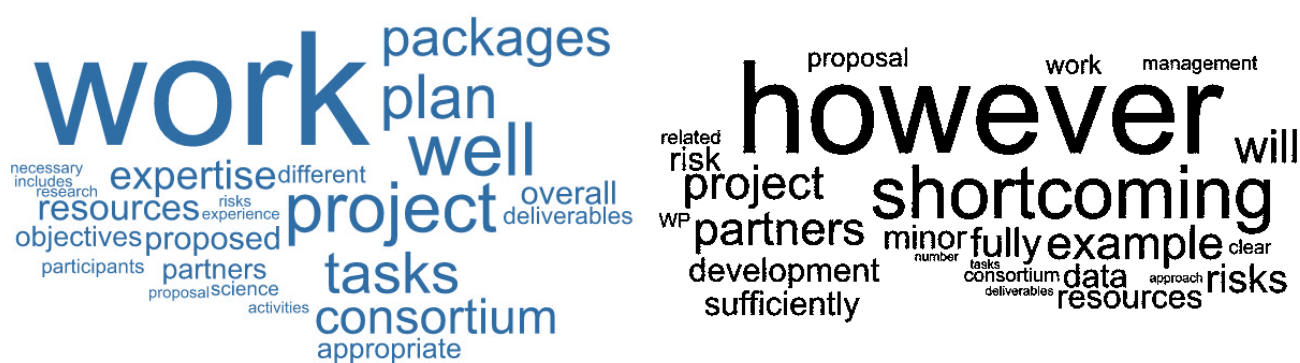


Figure 5: Implementation positive (blue) and implementation negative (black) comments - 25 most frequent words. (made in Pro Word Cloud add-in for PowerPoint)

Work Plan

For the work plan subcriterion, the award criteria, template guidance, and briefing slides seem to be properly aligned.

Only three of the ESRs had no positive comments about the proposal Work Plan, while 54 did not have any negative comments. This might indicate that this proposal element is well known by both applicants and evaluators.

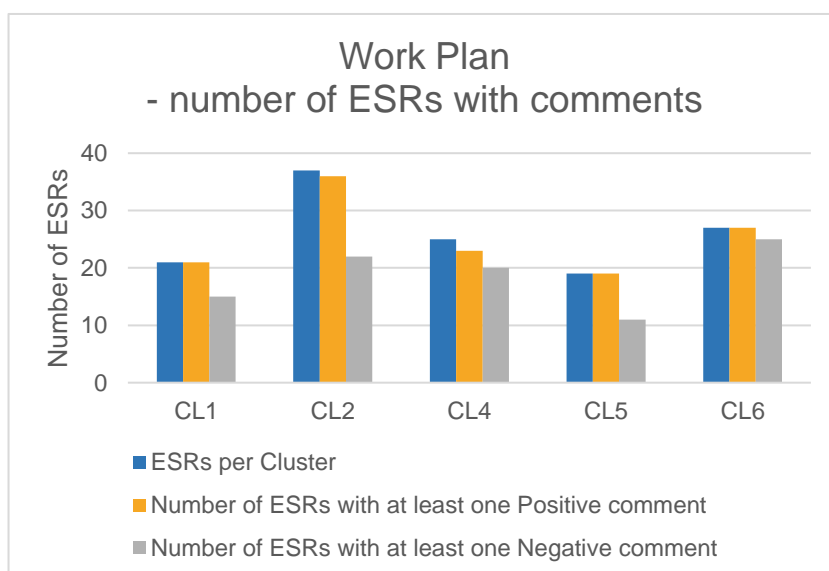
In many ESRs the work plan is simply given an overall generic comment such as “the overall work plan is well-structured /logical/well-developed/well-described.”

Milestones are mentioned in 94 of the ESRs, no less than 103 times. The majority of comments are positive, and it is very of-

ten the timing of the milestones and/or their relevance that draws praise. Conversely, if deemed badly timed, insufficient in numbers, or if lacking some sort of verification measure or being the wrong kind of milestone according to the evaluators’ preferences, milestones can be singled out for critical remarks as was the case in 23 of the ESRs. For good reasons, the scrutiny of milestones is often linked to the evaluation of monitoring measures.

Given that deliverables are important elements in all projects, it is noteworthy that their inherent qualities are very rarely commented upon. Most positive comments about deliverables are of a quite general nature using words like “relevant”, “feasible”, or as in one case “listed clearly”. Negative comments about deliverables are rare, and again mostly of a very general or standardised nature (e.g., one proposal was criticised for having too many deliverables).

A number of comments, both positive and negative, focuses on the numbers or timing of (specific) deliverables, milestones etc. Some negative comments criticised the limited number or wide temporal spread of milestones, without any qualification of why this is insufficient. This can give the impression that proposals must meet some unspoken standards.



Recommendations to the European Commission for the Work Plan subcriterion

- 4.1. When commenting on the inadequacy of the number and/or timing of deliverables or milestones, evaluators should always qualify why this needs to be different as there are no rules for number and timings in the template guidelines.
- 4.2. We recommend that during briefings evaluators are explicitly told that the numbers of milestones and deliverables are not fixed to a given number or temporal spacing but must be judged against the specific proposal.

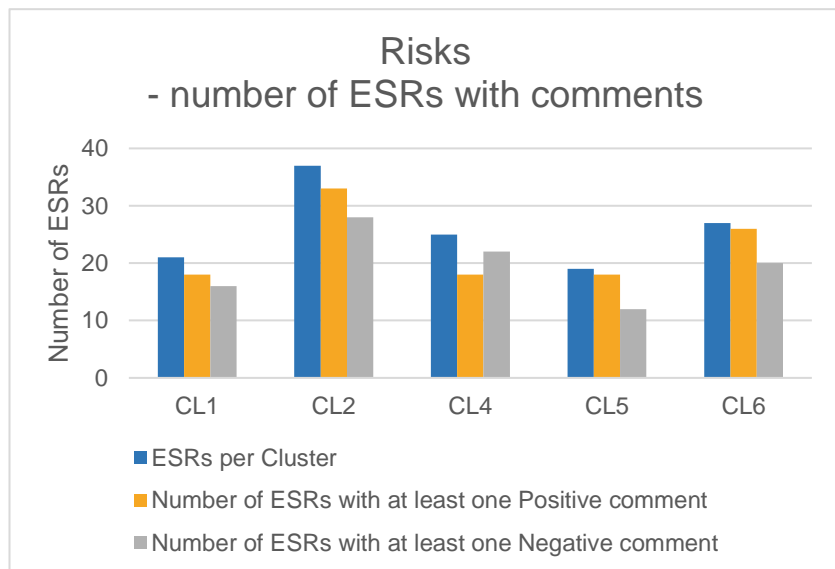
Risks

In the award criteria “Critical risks for implementation” is mentioned simply with the words “...assessment of risks...”, while in the template guidance it is present in the form of Table 3.1e and the template guides the applicants by stating what is needed:

“A list of critical risks, relating to project implementation, that the stated project's objectives may not be achieved. Detail any risk mitigation measures. You will be able to update the list of critical risks and mitigation measures as the project progresses (Table 3.1e).”

The briefing slides pose the question whether critical risks, relating to project implementation, are identified and whether proper risk mitigation measures are proposed.

In many of the ESRs with only positive comments (a total of 55 ESRs) related to the Risks subcriterion, the comments are of a very general nature, simply stating that risks and mitigation plans are appropriate. While general comments are also seen in the ESRs with only negative comments about Risks (30 ESRs), more of the negative comments offer specific explanations.



Out of the 129 ESRs, 17 mention ‘minor shortcomings’ and 19 mention ‘shortcomings’ under the Risk subcriterion with none seen as ‘major shortcomings’ out of a total of 80 negative comments. Many of the negative comments sound quite severe, but the lack of indication of the degree of shortcoming by evaluators makes it impossible to appreciate the critique’s influence on the score for Implementation.

In ESRs with both positive and negative comments on risks it is often the case that identification of risks is deemed good while the mitigation measures are criticised for being either generic and/or -critique of risk mitigation measures as they

are used in more than half of the negative comments.

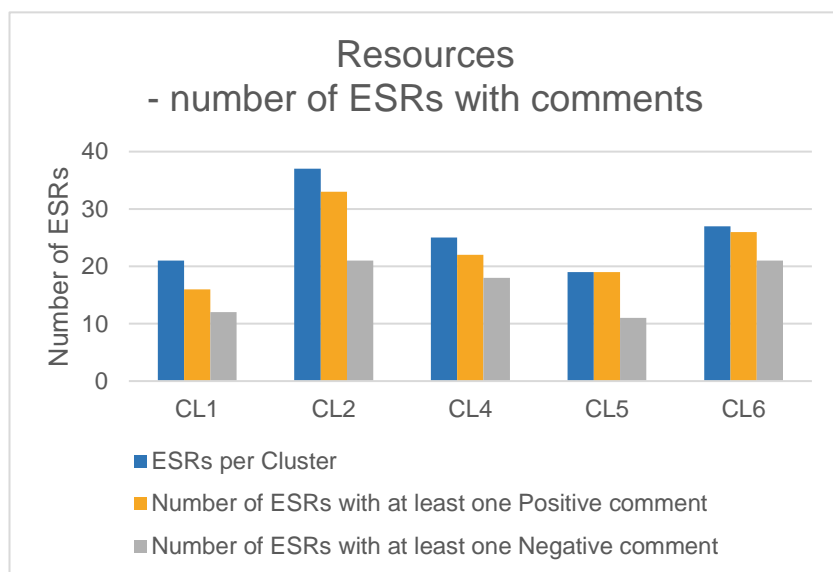
Some negative comments leave the reader with a sense of them being easy criticisms, for instance, one negative comment states that the application should have addressed “...the risks that some concepts/methods might require substantial revision. This is a shortcoming.” In another ESR it is simply stated that “Economic risks” have not been included. One can argue that such types of risks can be found in most research and innovation proposals. It is therefore important to further qualify and explain such negative remarks, in order for them to make sense and to be of benefit to the reader of the ESR.

Recommendations to the European Commission for the Risks subcriterion

- 4.3. It is recommended that the expert briefing slides emphasise that comments should be qualified in case of negative remarks.
- 4.4. More qualitative negative comments might be prompted by the following revised question in the briefing slides: *Are critical risks, relating to project implementation, identified and proper risk mitigation measures proposed? If not, what critical risks seem to be missing or why are (some of) the proposed mitigation measures not appropriate?*

Resources

For the Resources subcriterion, the award criteria, template guidance, and briefing slides seem to be properly aligned.



A lot of comments on resources are rather general and seem standardised. Almost all ESRs contain positive comments under this subcriterion. Positive comments are often limited to a single sentence, and very often include a statement about resources being “adequate”, “appropriate”. Or “justified” while negative comments can be paraphrased into “allocation of resources is not sufficiently clear”. In some cases, the evaluators also note that a specific task or work package looks under-financed.

Based on the negative comments on resources, it seems that very specific issues can lead to either a “Shortcoming” or “Minor shortcoming”.

E.g., in one negative ESR comment the evaluators find an explanation of the use of internally invoiced goods and services for a partner is unclear, which is named as a ‘shortcoming’.

One point that must be taken seriously is the instances where evaluators claim that uneven distribution of resources is deemed a problem. In all projects there will be differences due to a multitude of reasons; personnel costs vary enormously across the member states, as do costs of consumables, some beneficiaries have larger roles, and some have smaller roles in the project.

We also found a large number of very general negative remarks on the distribution of resources, such as “However the budget for some of the proposed high-cost activities is not fully realistic. This is a major shortcoming” and “...some of the costs claimed by some of the partners are not sufficiently justified”.

Recommendations to the European Commission for the Resources subcriterion

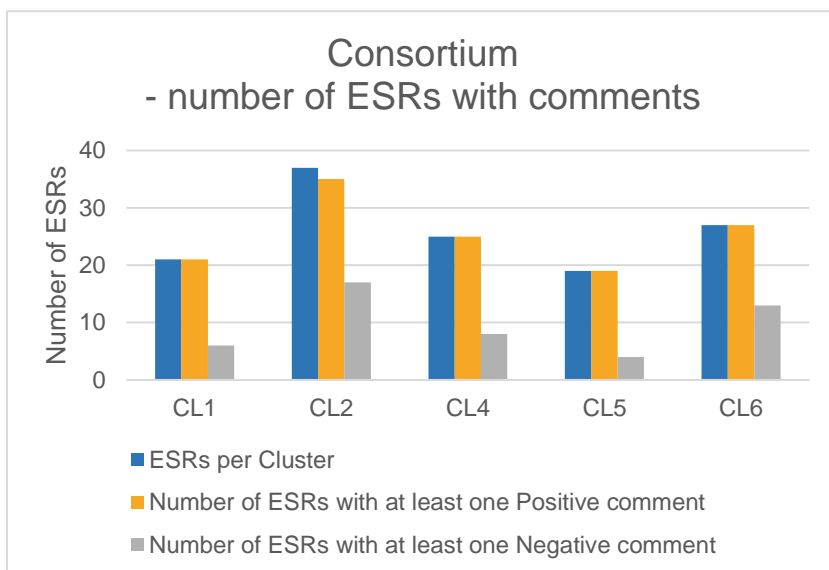
- 4.5. The briefing slides should include a (second) footnote instructing that applicants are not expected to present an even distribution of resources across partners and work packages. Distribution of resources should only be evaluated in light of their alignment with tasks, objectives, and deliverables of the specific work packages.
- 4.6. Negative comments on distribution of resources should be specified. This could be prompted by using the following revised question in the briefing slides: *Are the resources allocated to the work packages in line with their objectives and deliverables? If not, for which tasks/WPs are the resources inadequate and how would that reflect upon the objectives of the work package?*

Consortium

This is a subcriterion where the guidance in the proposal template and the briefing slides for HE evaluators are both very comprehensive. The ESRs comments on this subcriterion are quite long, especially for the positive comments. That the consortium, its composition, and coverage of relevant competences is seen as important to the evaluators is indicated by the 31 ESRs where the “shortcoming-scale” is used.

While comments do at times have a general tone to them, evaluators do in many cases make an effort to exemplify why they praise something. This is very commendable and such practices should be encouraged.

Not all evaluation panels include reflections about all the bullet points mentioned in the briefing slides for HE evaluators. It is, for instance, not often that evaluators make comments about competencies on open science (only mentioned in 17 of the ESRs), gender aspects (7 of the ESRs), or SSH (only mentioned in 17 of the ESRs, all from non-Cluster 2 proposals). It is interesting that when evaluators do give negative comments on competencies on gender, open science, and/or SSH, they are judged as minor shortcomings, while e.g., missing industry/commercial involvement results in more harsh critique.



Recommendations to the European Commission for the Consortium subcriterion

4.7. It is recommended that HE evaluators are fully familiar with gendered research and innovation as well as open science, so they can optimally assess the required competencies relevant for the consortium. The briefing slides can play an instrumental role to this end.

4.8. SSH HE evaluators of high quality and with exemplary competencies should be allocated at least to all proposals where SSH is mentioned in the topic text.

Annex A: Method and descriptive analysis

Coding

11 LERU universities each provided up to 20 of their ESRs from 2022 to include in the sample. Each ESR was included after an initial check of proposal numbers to avoid duplicates.

The core group developed a template in Excel. The template included all of the scores and award criteria divided into subcriteria present in the ESRs for Excellence, Impact and Implementation. Each ESR was given an anonymous ID code and entered into the Excel template by the partner providing the ESR. When all of the information was collected, all ESRs were copied into one large sheet.

Following that, the core group developed a codebook for a quantitative content analysis based on the available information in the ESRs and based on discussions and meetings with the full group. For example, for each sub-criterion, the codebook asked for the number of positive comments [0-...], the number of negative comments [0-...], positive adjectives used, negative adjectives used, the number of minor shortcomings mentioned [0-...], the number of shortcomings mentioned [0-...], the number of major shortcomings mentioned [0-...], the number of positive words [0-...], the number of negative words [0-...].

Following this, the core group coded the Excel template according to the codebook in Excel.

The core group had weekly meetings to discuss difficulties and ambiguities whilst coding to ensure that all ESRs were coded consistently.

Statistical analysis

Frequencies, means, standard deviations, crosstabs, and Pearson correlation coefficients were calculated in the statistical software program SPSS (version 27).

ESR sample

$N = 129$

70% Partner, 30% coordinator

47% funded, 47% not funded, 6% on reserve list

88% RIA, 10% IA, 2% CSA

| | No of ESRs | LERU as co-ordinator | LERU as partner | Funded | Not funded | Re-serve list | RIA | IA | CSA |
|-----------|------------|----------------------|-----------------|--------|------------|---------------|-----|----|-----|
| Cluster 1 | 21 | 8 | 13 | 9 | 10 | 2 | 21 | 0 | 0 |
| Cluster 2 | 37 | 17 | 20 | 20 | 16 | 1 | 37 | 0 | 0 |
| Cluster 4 | 25 | 7 | 18 | 5 | 20 | 0 | 21 | 3 | 1 |
| Cluster 5 | 19 | 2 | 17 | 14 | 4 | 1 | 16 | 3 | 0 |
| Cluster 6 | 27 | 4 | 23 | 12 | 11 | 4 | 18 | 7 | 2 |
| Total | 129 | 38 | 91 | 60 | 61 | 8 | 113 | 13 | 3 |

Scores

Overall scores [10-15]: $M = 12.6$, $SD = 1.5$

Excellence score [3-5]: $M = 4.1$, $SD = 0.6$

Impact score [3-5]: $M = 4.2$, $SD = 0.6$

Implementation score [3-5]: $M = 4.2$, $SD = 0.5$

Mean scores per Cluster (SD in brackets):

| | Overall | Excellence | Impact | Implementation |
|-----------|------------|------------|-----------|----------------|
| Cluster 1 | 12.6 (1.4) | 4.1 (0.6) | 4.2 (0.6) | 4.3 (0.5) |
| Cluster 2 | 12.6 (1.5) | 4.2 (0.6) | 4.1 (0.5) | 4.2 (0.6) |
| Cluster 4 | 11.6 (1.4) | 3.8 (0.6) | 3.9 (0.6) | 3.9 (0.5) |
| Cluster 5 | 13.6 (1.3) | 4.4 (0.6) | 4.6 (0.6) | 4.6 (0.4) |
| Cluster 6 | 12.7 (1.4) | 4.2 (0.6) | 4.3 (0.5) | 4.3 (0.5) |

Use of shortcomings nomenclature

There were large differences between the clusters regarding the use of the term “shortcoming” (either “minor shortcoming”, “shortcoming” or “major shortcoming”). All of the ESRs in Cluster 6 of our sample included at least one comment specifying a shortcoming, whereas in Cluster 4 the use of the term was less prevalent. Here 60% of the ESRs included at least one comment specifying a shortcoming.

The mean scores of the remaining ESRs in each cluster that had no shortcoming ranged between 11.5 and 14.5, as can be seen in the table below. This means that according to their score, the proposals had some weaknesses, but evaluators chose to comment negatively without using the term “shortcoming”.

| | % of ESRs where evaluators used some form of the term "shortcoming" | Mean score of the remaining ESRs with no shortcomings |
|-----------|---|---|
| Cluster 1 | 76% | 12.4 |
| Cluster 2 | 70% | 12.5 |
| Cluster 4 | 60% | 11.5 |
| Cluster 5 | 89% | 14.5 |
| Cluster 6 | 100% | NA |

The distribution of the usage of “shortcomings” across clusters and award criteria is shown in the table below. In the Excellence section, there were approximately 1.84 shortcomings (including “minor shortcomings”, “shortcomings” and “major shortcomings”) per ESR across all clusters. In comparison, the mean number of shortcomings per ESR was of 1.31 in the Impact section and 1.33 in the Implementation section.

Only in very few cases, evaluators repeated the use of “major shortcomings” when pointing out significant weaknesses in the proposals.

In Cluster 6, evaluators used the term “shortcomings” more frequently than in the other clusters.

Table: No. of minor shortcomings, shortcomings and major shortcomings per ESR and per cluster

| | No. of ESRs | Excellence | | | Impact | | | Implementation | | |
|--------------|-------------|-------------------|-------------|-------------------|-------------------|-------------|-------------------|-------------------|-------------|-------------------|
| | | Minor shortcoming | Shortcoming | Major shortcoming | Minor shortcoming | Shortcoming | Major shortcoming | Minor shortcoming | Shortcoming | Major shortcoming |
| Cluster 1 | 21 | 0.43 | 1.00 | 0.14 | 0.43 | 0.48 | 0.00 | 0.19 | 0.33 | 0.05 |
| Cluster 2 | 37 | 0.22 | 1.00 | 0.00 | 0.35 | 0.78 | 0.00 | 0.27 | 0.59 | 0.03 |
| Cluster 4 | 25 | 0.64 | 0.80 | 0.00 | 0.60 | 0.20 | 0.00 | 0.48 | 0.80 | 0.00 |
| Cluster 5 | 19 | 0.84 | 1.37 | 0.00 | 0.21 | 0.95 | 0.00 | 0.53 | 1.05 | 0.00 |
| Cluster 6 | 27 | 1.11 | 1.81 | 0.07 | 0.96 | 1.41 | 0.07 | 1.00 | 1.11 | 0.30 |
| Mean | 129 | 0.61 | 1.19 | 0.04 | 0.52 | 0.78 | 0.02 | 0.49 | 0.77 | 0.08 |
| Per criteria | | 1.84 | | | 1.31 | | | 1.33 | | |

Excellence

The table below presents Pearson correlation coefficients between the excellence score and subcriteria scores.

Annex A: Method and descriptive analysis - LERU ESR analysis

| | |
|---------------------------|-------------------------|
| State of the art positive | $r = 0.229, p < 0.01$ |
| State of the art negative | $r = -0.44, p < 0.001$ |
| Objectives positive | Not significant |
| Objectives negative | $r = -0.47, p < 0.001$ |
| Method positive | $r = 0.315, p < 0.001$ |
| Method negative | $r = -0.587, p < 0.01$ |
| AI positive | Not significant |
| AI negative | $r = -0.24, p < 0.05$ |
| DNSH positive | Not significant |
| DNSH negative | Not significant |
| Gender positive | $r = 0.264, p < 0.01$ |
| Gender negative | $r = -0.265, p < 0.01$ |
| ID/SSH positive | $r = 0.267, p < 0.01$ |
| ID/SSH negative | $r = -0.220, p < 0.05$ |
| OS/RDM positive | Not significant |
| OS/RDM negative | $r = -0.178, p < 0.05$ |
| TRL positive | Not significant |
| TRL negative | $r = -0.290, p < 0.001$ |

There is a significant positive correlation between the excellence score and impact score (Pearson coefficient of 0.664, $p < 0.01$) indicating the higher the excellence score, the higher the impact score and vice versa .

There is also a significant positive correlation between excellence score and implementation score (Pearson coefficient of 0.554, $p < 0.001$) indicating the higher the excellence score, the higher the implementation score and vice versa.

There is also a significant positive correlation between impact score and implementation score (Pearson coefficient of 0.566, $p < 0.001$) indicating a higher score on impact goes together with a higher score on implementation .

In excellence, the most words were used in the following subcriteria: positive comments objectives ($M = 66.5, SD = 47.8$), positive comments methodology ($M = 63, SD = 50$), state of the art positive comments ($M = 49, SD = 39$). The fewest words were used in DNSH negative comments ($M = 0.3, SD = 3$), DNSH positive comments ($M = 0.6, SD = 3$), gender negative comments ($M = 3, SD = 9$).

Overall, there were a lot more positive comments than negative in this sample of ESRs (which is not surprising as these were all projects that received scores above the threshold). We found most of the positive comments in: 1) the methodology section (on average 3 comments per ESR), 2) Open Science/RDM (on average 2 comments per ESR), 3) objectives (on average 2 comments per ESR). In terms of negative comments, we found most in: 1) methodology (on average 1.5 comments per ESR), 2) objectives (on average less than 1 comment per ESR), 3) state of the art (on average less than 1 comment per ESR).

When looking at the TRL comments, we found the fewest positive and negative comments in Cluster 2 (0 positive comments in 38 cases, 0 negative comments in 37 cases). In Clusters 4, 5 and 6, we found most positive comments (1 or 2 comments) on TRL. In Cluster 4 and 6 we found most negative comments (1 or 2 comments).

Most words ($M = 20$) were used in the positive description of TRL in Cluster 5 versus Cluster 2 where least words ($M = 0$) were used for positive descriptions of TRL. Most words were used in negative description TRL in Cluster 4 ($M = 14$) versus least words ($M = 0$) were used in Cluster 1.

Mentioning of shortcomings: “major shortcomings” were not often mentioned (except twice in methodology and once in gender). Most shortcomings were mentioned in methodology (18 times “minor shortcoming”, 35 times the term “shortcoming”). Also, in the objectives section, several shortcomings were mentioned (14 times “minor shortcoming”, 22 times “shortcoming”). The least mention of the word “shortcoming” was in DNSHP (0 times),

and in gender (3 times use of term “minor shortcoming”, 2 times “shortcoming”, 1 time use of “major shortcoming”).

For the general subcriteria (such as state of the art, objectives, methodology) Cluster 1 uses least words in the feedback (both positive and negative). Also, Cluster 4 uses a lower number of words than the other clusters. Clusters 2, 5 and 6 use most words especially in the positive descriptions.

In 67% of these ESRs, the gender dimension is commented positively, in 13.5% negatively. Gender is most commented positively in Cluster 2 (in 34 ESRs) and least in Cluster 4 (in 9 ESRs). Gender is most negatively commented in Cluster 2 (5 ESRs) and least in Cluster 6 (2 ESRs).

The ESRs in this sample could have up to 3 negative comments on objectives and still get funded (although most funded projects got 0 negative comments, 15 projects got 1 negative comment and were funded, 3 projects 2 negative comments and were funded and 1 project 3 negative comments and was funded).

Impact

The table below presents Pearson correlation coefficients between the impact score and the subcriteria scores.

| | |
|-------------------|-------------------------|
| Pathways positive | $r = 0.304, p < 0.001$ |
| Pathways negative | $r = -0.423, p < 0.001$ |
| S&S positive | $r = 0.255, p < 0.01$ |
| S&S negative | $r = -0.489, p < 0.01$ |
| Barriers positive | $r = 0.260, p < 0.01$ |
| Barriers negative | $r = -0.322, p < 0.01$ |
| DEC positive | $r = 0.299, p < 0.001$ |
| DEC negative | $r = -0.462, p < 0.001$ |
| IP positive | Not significant |
| IP negative | $r = -0.234, p < 0.01$ |

In the impact section in general, we also found more positive comments than negative comments. Most positive comments were found in 1) DEC (on average 3.5 comments per ESR), 2) pathways to impact (on average 3 comments per ESR), 3) Scale and significance (on average 1 comment per ESR). Most negative comments were found in 1) Pathways to impact (on average 1 comment per ESR), 2) DEC (on average 1 comment per ESR), 3) scale and significance (on average less than 1 comment per ESR).

Scale and significance was mentioned positively at least once in 89 ESRs, and mentioned negatively at least once in 57 ESRs. A positive mention of scale and significance was higher in Clusters 1, 2 and 6. A negative mention of scale and significance was higher in Clusters 2 and 6.

In 40 ESRs, there was a positive mention of pathways to impact while in only 8 ESRs there was a negative mention. In 90 ESRs, there was a positive mention of the expected outcomes while in 18 ESRs there was a negative mention. In 37 ESRs, there was a positive mention of expected impact while in only 8 ESRs there was a negative mention of expected impact. In 40 ESRs, there was positive mention of wider impact on societal impact while in 11 ESRs there was a negative mention of this. In 16 ESRs, there was a positive mention of wider impact on economic impact while in 11 ESRs there was a negative mention of this. In 21 ESRs there was a positive mention of wider impact on scientific impact, in only 1 ESR there was a negative mention of this. The impact canvas was only mentioned once (!) in the pathways to impact evaluation section.

If we look at the different Clusters for the positive mentions of pathways to impact, expected outcomes and expected impact and impact on societal impact, we see some differences. A positive mention of pathways to impact was highest in Cluster 1 (12 times) and Cluster 2 (8 times). A positive mention of expected outcomes was highest in Cluster 2 (24 times). A positive mention of expected impact was highest in Cluster 6 (11 times). A positive mention of wider impact on societal impact was highest in Cluster 2 (15 times).

19 ESRs received 0 positive comments for barriers to impact but got a 5/5 score on impact. 5 ESRs received at least 1 negative comment for barriers to impact but still got a 5/5 score on impact.

The mean number of words IP is highest for Cluster 6 (positive: $M= 24.3$, $SD= 25.2$ / negative: $M= 13.3$, $SD= 20.9$).

Mentioning of “shortcomings” in the impact section: in pathways to impact, we found 9 ESRs with “minor shortcomings”, 27 ESRs with “shortcomings”, and 1 ESR with “major shortcoming”. In scale and significance, we found 10 ESRs with “minor shortcomings” and 21 ESRs with “shortcomings”. In barriers, we found 14 ESRs with “minor shortcomings” and 12 shortcomings. For DEC, we found 14 ESRs with “minor shortcomings”, and 16 ESRs with shortcomings for DEC. For IP, we found 12 ESRs with “minor shortcomings” and 8 ESRs with “shortcomings”.

In impact, the categories with most used words were positive comments on pathways ($M= 85.9$, $SD= 69.2$), DEC positive comments ($M= 81.5$, $SD= 59.4$), S&S positive comments ($M= 31.5$, $SD= 43.9$). Least words were used in the categories negative comments for barriers ($M= 12.9$, $SD= 23.1$), positive comments on barriers ($M= 15.3$, $SD= 15.6$), and negative words for S&S ($M= 15.2$, $SD= 22.2$).

Implementation

The table below presents Pearson correlation coefficients between the impact score and the subcriteria scores.

| | |
|---------------------|--------------------------|
| Consortium positive | $r= 0.237$, $p< 0.01$ |
| Consortium negative | $r= -0.291$, $p< 0.001$ |
| WP positive | $r= 0.288$, $p< 0.001$ |
| WP negative | $r= -0.551$, $p< 0.001$ |
| Resources positive | $r= 0.289$, $p< 0.001$ |
| Resources negative | $r= -0.327$, $p< 0.001$ |
| Risks positive | $r= 0.256$, $p< 0.01$ |
| Risks negative | Not significant |

We found a moderate significant, negative correlation between negative comments on WPs and the implementation score: the more negative comments on WPs, the lower the implementation score, and vice versa. Strangely, negative comments on risks are not significantly correlated with implementation score.

We found more positive comments than negative ones. Most positive comments were found in: 1) consortium (on average 4 per ESR), 2) WPs (on average 3.5 per ESR), and 3) resources (on average 1.5 comments per ESR). Negative comments: 1) WPs (on average 1 negative per ESR), 2) risks (on average 1 negative comments per ESR), and 3) resources (on average 0.5 negative comments per ESR).

There are 4 ESRs that got at least one negative comment on risks but still received a 5/5 score on implementation.

In implementation, most words were used for consortium positive comments ($M= 74.8$, $SD= 45.2$), WP positive comments ($M= 56.3$, $SD= 36.2$), WP negative comments ($M= 34.1$, $SD= 41.2$). Least words were used for consortium negative comments ($M= 11.5$, $SD= 19.6$), resources negative comment ($M= 13.2$, $SD= 20.1$).

In terms of consortium, 13 ESRs mention “minor shortcomings”, 15 mention “shortcomings”, while 3 ESRs mention “major shortcomings”. For the WPs, 15 ESRs mention “minor shortcomings”, 29 ESRs mention “shortcomings”, and 1 mentions “major shortcomings”. In terms of resources, 12 ESRs mention “minor shortcomings”, 12 mention “shortcomings” and 2 mention “major shortcomings”. In terms of risks, 17 ESRs mention “minor shortcomings” while 19 ESRs mention “shortcomings”.

Annex B: Comparing Award criteria with Template guidance and Standard briefing slides

| EXCELLENCE | IMPACT | QUALITY AND EFFICIENCY OF THE IMPLEMENTATION |
|---|--|--|
| <ul style="list-style-type: none"> ✓ Clarity and pertinence of the project's objectives, and the extent to which the proposed work is ambitious, and goes beyond the state-of-the-art. ✓ Soundness of the proposed methodology, including the underlying concepts, models, assumptions, interdisciplinary approaches, appropriate consideration of the gender dimension in research and innovation content, and the quality of open science practices including sharing and management of research outputs and engagement of citizens, civil society and end users where appropriate. | <ul style="list-style-type: none"> ✓ Credibility of the pathways to achieve the expected outcomes and impacts specified in the work programme, and the likely scale and significance of the contributions due to the project. ✓ Suitability and quality of the measures to maximize expected outcomes and impacts, as set out in the dissemination and exploitation plan, including communication activities. | <ul style="list-style-type: none"> ✓ Quality and effectiveness of the work plan, assessment of risks, and appropriateness of the effort assigned to work packages, and the resources overall. ✓ Capacity and role of each participant, and extent to which the consortium as a whole brings together the necessary expertise. |

Table above: Excellence, impact, implementation award criteria with bold as shown in the Standard briefing slides for HE evaluators ver. 3.0.¹⁶

Table 1: Excellence - Objectives and State-of-the-Art (including TRL)

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|-------------------------------------|--|--|
| | Version 1.2 25 May 2021 and the next version, version 2.0 - 21 January 2022 have identical explanations apart from two paragraphs have changed place in the impact section. There are also no relevant changes in text compared to the current version 3.2 - 15.11.2022 where lump sum and information on artificial intelligence have been added. | We have used ver. 3.0 which is identical to the current ver. 7.0 (27 October 2023) for all the criteria we are commenting on in this report. |

¹⁶ Standard briefing slides for HE evaluators https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/experts/standard-briefing-slides-for-experts_he_en.pdf

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|---|--|--|
| (The following aspects will be taken into account, to the extent that the proposed work corresponds to the description in the work programme) | <ul style="list-style-type: none"> The following aspects will be taken into account only to the extent that the proposed work is within the scope of the work programme topic. | Proposals aspects are assessed to the extent that the proposed work is within the scope of the work programme topic |
| Clarity and pertinence of the project's objectives | <ul style="list-style-type: none"> Briefly describe the objectives of your proposed work. Why are they pertinent to the work programme topic? Are they measurable and verifiable? Are they realistically achievable? | Assess the project's objectives: <ul style="list-style-type: none"> Are they clear and pertinent to the topic? Are they measurable and verifiable? Are they realistically achievable? |
| and the extent to which the proposed work is ambitious and goes beyond the state of the art | <ul style="list-style-type: none"> Describe how your project goes beyond the state-of-the-art, and the extent the proposed work is ambitious. Indicate any exceptional ground-breaking R&I, novel concepts and approaches, new products, services or business and organisational models. Where relevant, illustrate the advance by referring to products and services already available on the market. Refer to any patent or publication search carried out. | <ul style="list-style-type: none"> Is the proposed work ambitious and goes beyond the state-of-the-art? Does the proposal include ground-breaking R&I, novel concepts and approaches, new products, services or business and organisational models? |
| TRL, not mentioned in Award Criteria | [bullet in the section about Objectives and ambition] <ul style="list-style-type: none"> Describe where the proposed work is positioned in terms of R&I maturity (i.e. where it is situated in the spectrum from 'idea to application', or from 'lab to market'). Where applicable, provide an indication of the Technology Readiness Level, if possible distinguishing the start and by the end of the project ⚠ Please bear in mind that advances beyond the state of the art must be interpreted in the light of the positioning of the project. Expectations will not be the same for RIAs at lower TRL, compared with Innovation Actions at high TRLs. | <ul style="list-style-type: none"> Is the R&I maturity of the proposed work in line with the topic description? <p>Please bear in mind that advances beyond the state of the art must be interpreted in the light of the positioning of the project. For example, expectations will not be the same for RIAs at lower TRL, compared with Innovation Actions at high TRLs.</p> |

Table 2: Excellence - Methodology, Interdisciplinarity including SSH, Gender dimension, DNSH and Open Science including management of research outputs

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|--|--|--|
| Soundness of the proposed methodology , including the underlying concepts, models, assumptions, | <ul style="list-style-type: none"> Describe and explain the overall methodology, including the concepts, models and assumptions that underpin your work. Explain how this will enable you to deliver your project’s objectives. Refer to any important challenges you may have identified in the chosen methodology and how you intend to overcome them. | <p>Assess the scientific methodology:</p> <ul style="list-style-type: none"> Is the scientific methodology (i.e. the concepts, models and assumptions that underpin the work) clear and sound? |
| DNSH, not mentioned in Award Criteria | <p>[bullet in start of Methodology]</p> <p>⚠ Where relevant, include how the project methodology complies with the ‘do no significant harm’ principle as per Article 17 of Regulation (EU) No 2020/852 on the establishment of a framework to facilitate sustainable investment (i.e. the so-called ‘EU Taxonomy Regulation’). This means that the methodology is designed in a way it is not significantly harming any of the six environmental objectives of the EU Taxonomy Regulation.</p> | <p>Extra slide under additional questions</p> <p>In line with the European Green Deal objectives, economic activities should not make a significant harm to any of the six environmental objectives (EU Taxonomy Regulation)</p> <ul style="list-style-type: none"> Applicants can refer to the DNSH principle when presenting their research methodology and the expected impacts of the project, to show that their project will not carry out activities that make a significant harm to any of the six environmental objectives of the EU Taxonomy Regulation. <p>However, evaluators will not score applications in relation to their compliance with the DNSH principle unless explicitly stated in the work programme (currently, this is the case only for actions in the European Innovation Council Work Programme 2021).</p> |
| inter-disciplinary approaches, | <ul style="list-style-type: none"> Explain how expertise and methods from different disciplines will be brought together and integrated in pursuit of your objectives. If you consider that an inter-disciplinary approach is unnecessary in the context of the proposed work, please provide a justification. | <ul style="list-style-type: none"> Is it clear how expertise and methods from different disciplines will be brought together and integrated in pursuit of the objectives? if applicants justify that an inter-disciplinary approach is unnecessary, is it credible? |

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|---|---|--|
| SSH not mentioned in the award criterion | <p>[bullet in start of Methodology]</p> <ul style="list-style-type: none"> For topics where the work programme indicates the need for the integration of social sciences and humanities, show the role of these disciplines in the project or provide a justification if you consider that these disciplines are not relevant to your proposed project. | <ul style="list-style-type: none"> For topics indicating the need for the integration of social sciences and humanities, is the role of these disciplines properly addressed? <p>+ extra slide on integration of SSH:</p> <p>Assessing the effective contribution of social science and humanities disciplines and expertise as part of the scientific methodology of the project</p> <p>When the integration of SSH is required, applicants have to show the roles of these disciplines or provide a justification if they consider that it is not relevant for their project. A proposal without a sufficient contribution/integration of SSH research and competences will receive a lower evaluation score.</p> |
| appropriate consideration of the gender dimension in research and innovation content, | <ul style="list-style-type: none"> Describe how the gender dimension (i.e. sex and/or gender analysis) is taken into account in the project's research and innovation content. If you do not consider such a gender dimension to be relevant in your project, please provide a justification. Note: This section is mandatory except for topics which have been identified in the work programme as not requiring the integration of the gender dimension into R&I content. Remember that that this question relates to the content of the planned research and innovation activities, and not to gender balance in the teams in charge of carrying out the project. Sex and gender analysis refers to biological characteristics and social/cultural factors respectively. For guidance on methods of sex / gender analysis and the issues to be taken into account, please refer to https://ec.europa.eu/info/news/gendered-innovations-2-2020-nov-24_en | <ul style="list-style-type: none"> Has the gender dimension in research and innovation content been properly taken into account? <p>+ extra slide on what is meant by gender dimension:</p> <p>Addressing the gender dimension in research and innovation entails taking into account sex and gender in the whole research & innovation process.</p> |

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|--|---|---|
| <p>and the quality of <i>open science practices</i>, including sharing and <i>management of research outputs</i> and <i>engagement of citizens, civil society and end-users</i> where appropriate.</p> | <ul style="list-style-type: none"> Describe how appropriate <i>open science practices are implemented as an integral part of the proposed methodology</i>. Show how the choice of practices and their implementation are adapted to the nature of your work, in a way that will increase the chances of the project delivering on its objectives. If you believe that none of these practices are appropriate for your project, please provide a justification here. Open science is an approach based on open cooperative work and systematic sharing of knowledge and tools as early and widely as possible in the process. Open science practices include early and open sharing of research (for example through preregistration, registered reports, pre-prints, or crowd-sourcing); research output management; measures to ensure reproducibility of research outputs; providing open access to research outputs (such as publications, data, software, models, algorithms, and workflows); participation in open peer-review; and involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science). Please note that this question does not refer to outreach actions that may be planned as part of communication, dissemination and exploitation activities. These aspects should instead be described below under 'Impact'. <i>Research data management and management of other research outputs</i>: Applicants generating/collecting data and/or other research outputs (except for publications) during the project must provide maximum 1 page on <i>how the data/ research outputs will be managed in line with the FAIR principles</i> (Findable, Accessible, Interoperable, Reusable), addressing the following (the description should be specific to your project): | <ul style="list-style-type: none"> Are open science practices implemented as an integral part of the proposed methodology? Is the research data management properly addressed? + extra slide on Open Science and mandatory and recommended practices: <p>Open science is an approach based on open cooperative work and systematic sharing of knowledge and tools as early and widely as possible in the process, including active engagement of society.</p> <p>Mandatory OS practices</p> <ul style="list-style-type: none"> Mandatory in all calls: Open access to publications; <i>RDM in line with the FAIR principles including data management plans</i>; open access to research data unless exceptions apply ('as open as possible as closed as necessary'); access and/or information to research outputs and tools/instruments for validating conclusions of scientific publications and validating/re-using data. Additional obligations specific to certain work programme topics. <p>Reflect both in lower score when not sufficiently addressed</p> <p>Recommended OS practices</p> <ul style="list-style-type: none"> All open science practices beyond mandatory Evaluate positively when sufficiently addressed |

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|-------------------------------------|---|---|
| | <p>Types of data/research outputs (e.g. experimental, observational, images, text, numerical) and their estimated size; if applicable, combination with, and provenance of, existing data.</p> <p>Findability of data/research outputs: Types of persistent and unique identifiers (e.g. digital object identifiers) and trusted repositories that will be used.</p> <p>Accessibility of data/research outputs: IPR considerations and timeline for open access (if open access not provided, explain why); provisions for access to restricted data for verification purposes.</p> <p>Interoperability of data/research outputs: Standards, formats and vocabularies for data and metadata.</p> <p>Reusability of data/research outputs: Licenses for data sharing and re-use (e.g. Creative Commons, Open Data Commons); availability of tools/software/models for data generation and validation/ interpretation /re-use.</p> <p>Curation and storage/preservation costs; person/team responsible for data management and quality assurance.</p> <ul style="list-style-type: none"> Proposals selected for funding under Horizon Europe will need to develop a detailed data management plan (DMP) for making their data/research outputs findable, accessible, interoperable and reusable (FAIR) as a deliverable by month 6 and revised towards the end of a project’s life-time. For guidance on open science practices and research data management, please refer to the relevant section of the HE Programme Guide on the Funding & Tenders Portal. | |

Table 3: Excellence - Artificial Intelligence (AI) (added from version 3.2)

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|--|--|---|
| <p>AI, not mentioned in Award Criteria</p> | <p>[In the template ver. 2.0 this was part of the Additional questions section but a Methodology bullet was inserted right after DNSH from template version 3.2 - 15. November 2022]</p> <p>If you plan to use, develop and/or deploy artificial intelligence (AI) based systems and/or techniques you must demonstrate their technical robustness. AI-based systems or techniques should be, or be developed to become:</p> <ul style="list-style-type: none"> • technically robust, accurate and reproducible, and able to deal with and inform about possible failures, inaccuracies and errors, proportionate to the assessed risk they pose • socially robust, in that they duly consider the context and environment in which they operate • reliable and function as intended, minimizing unintentional and unexpected harm, preventing unacceptable harm and safeguarding the physical and mental integrity of humans • able to provide a suitable explanation of their decision-making processes, whenever they can have a significant impact on people’s lives. | <p>Extra slide under additional questions</p> <ul style="list-style-type: none"> • Experts must answer an additional question as part of their individual evaluations on whether the activities proposed involve the use and/or development of AI-based systems and/or techniques. • If you answer ‘yes’ to this question, you must assess the technical robustness of the proposed AI-system as part of the excellence criterion (if applicable). • In addition, your answer to this question will help us to with the proper follow-up of any aspects related to Artificial Intelligence in projects funded under Horizon Europe. <p>(*) Technical robustness refers to technical aspects of AI systems and development, including resilience to attack and security, fallback plan and general safety, accuracy, reliability and reproducibility.</p> <p>AI-based systems or techniques should be, or be developed to become:</p> <ul style="list-style-type: none"> • Technically robust, accurate and reproducible, and able to deal with and inform about possible failures, inaccuracies and errors, proportionate to the assessed risk posed by the AI-based system or technique. • Socially robust, in that they duly consider the context and environment in which they operate. • Reliable and function as intended, minimizing unintentional and unexpected harm, preventing unacceptable harm and safeguarding the physical and mental integrity of humans. • Able to provide a suitable explanation of its decision-making process, whenever an AI-based system can have a significant impact on people’s lives. |

Table 4: Impact - Pathways To Impact, Barriers, And Scale And Significance

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|---|--|--|
| | <p>The results of your project should make a contribution to the expected outcomes set out for the work programme topic over the medium term, and to the wider expected impacts set out in the ‘destination’ over the longer term.</p> <p>In this section you should show how your project could contribute to the outcomes and impacts described in the work programme, the likely scale and significance of this contribution, and the measures to maximise these impacts.</p> | |
| <p>Credibility of the pathways to achieve the expected outcomes and impacts specified in the work programme</p> | <p>Provide a narrative explaining how the project’s results are expected to make a difference in terms of impact, beyond the immediate scope and duration of the project. The narrative should include the components below, tailored to your project.</p> <p>(a) Describe the unique contribution your project results would make towards (1) the outcomes specified in this topic, and (2) the wider impacts, in the longer term, specified in the respective destinations in the work programme.</p> | <p>Assess the proposed pathways towards impact:</p> <ul style="list-style-type: none"> Is the contribution of the project towards the (1) expected outcomes of the topic and (2) the wider impacts, in the longer term, as specified in the respective destinations of the WP, credible? <p>+ slide explaining pathways</p> |
| <p>and the likely scale and significance of the contributions due to the project.</p> | <p>(b) Give an indication of the scale and significance of the project’s contribution to the expected outcomes and impacts, should the project be successful. Provide quantified estimates where possible and meaningful.</p> <ul style="list-style-type: none"> ‘Scale’ refers to how widespread the outcomes and impacts are likely to be. For example, in terms of the size of the target group, or the proportion of that group, that should benefit over time; ‘Significance’ refers to the importance, or value, of those benefits. For example, number of additional healthy life years; efficiency savings in energy supply. <p>Explain your baselines, benchmarks and assumptions used for those estimates. Wherever possible, quantify your estimation</p> | <ul style="list-style-type: none"> Are the scale and significance of the project’s contribution to the expected outcomes and impacts estimated and quantified (including baselines, benchmarks and assumptions used for those estimates)? <ul style="list-style-type: none"> ‘Scale’ refers to how widespread the outcomes and impacts are likely to be. For example, in terms of the size of the target group, or the proportion of that group, that should benefit over time; ‘Significance’ refers to the importance, or value, of those benefits. For example, number of additional healthy life years; efficiency savings in energy supply. |

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|--|--|---|
| | <p>of the effects that you expect from your project. Explain assumptions that you make, referring for example to any relevant studies or statistics. Where appropriate, try to use only one methodology for calculating your estimates: not different methodologies for each partner, region or country (the extrapolation should preferably be prepared by one partner).</p> | |
| <p>Barriers, not mentioned in award criteria</p> | <p>(c) Describe any requirements and potential barriers - arising from factors beyond the scope and duration of the project - that may determine whether the desired outcomes and impacts are achieved. These may include, for example, other R&I work within and beyond Horizon Europe; regulatory environment; targeted markets; user behaviour. Indicate if these factors might evolve over time. Describe any mitigating measures you propose, within or beyond your project, that could be needed should your assumptions prove to be wrong, or to address identified barriers.</p> | <ul style="list-style-type: none"> Are potential barriers to the expected outcomes and impacts identified (i.e. other R&I work within and beyond Horizon Europe; regulatory environment; targeted markets; user behaviour), and mitigation measures proposed? Is any potential negative environmental outcome or impact (including when expected results are brought at scale, such as at commercial level) identified? Is the management of the potential negative impacts properly described? |

Table 5: Impact - DEC Including IP

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|---|---|--|
| <p>Suitability and quality of the measures to maximise expected outcomes and impacts, as set out in the dissemination and exploitation plan, including communication activities.</p> | <ul style="list-style-type: none"> Describe the planned measures to maximise the impact of your project by providing a first version of your ‘plan for the dissemination and exploitation including communication activities’. Describe the dissemination, exploitation and communication measures that are planned, and the target group(s) addressed (e.g. scientific community, end users, financial actors, public at large). <ul style="list-style-type: none"> Please remember that this plan is an admissibility condition, unless the work programme topic explicitly states otherwise. In case your proposal is selected for funding, a more detailed ‘plan for dissemination and exploitation including communication activities’ will need to be provided as a mandatory project deliverable within 6 months after signature date. This plan shall be | <p>Assess the measures to maximise impact - Dissemination, exploitation and communication:</p> <ul style="list-style-type: none"> Are the proposed dissemination, exploitation and communication measures suitable for the project and of good quality? All measures should be proportionate to the scale of the project, and should contain concrete actions to be implemented both during and after the end of the project. Are the target groups (e.g. scientific community, end users, financial actors, public at large) for these measures identified? |

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|-------------------------------------|---|--|
| | <p>periodically updated in alignment with the project’s progress.</p> <ul style="list-style-type: none"> • Communication¹⁷ measures should promote the project throughout the full lifespan of the project. The aim is to inform and reach out to society and show the activities performed, and the use and the benefits the project will have for citizens. Activities must be strategically planned, with clear objectives, start at the outset and continue through the lifetime of the project. The description of the communication activities needs to state the main messages as well as the tools and channels that will be used to reach out to each of the chosen target groups. • All measures should be proportionate to the scale of the project, and should contain concrete actions to be implemented both during and after the end of the project, e.g. standardisation activities. Your plan should give due consideration to the possible follow-up of your project, once it is finished. In the justification, explain why each measure chosen is best suited to reach the target group addressed. Where relevant, and for innovation actions, in particular, describe the measures for a plausible path to commercialise the innovations. • If exploitation is expected primarily in non-associated third countries, justify by explaining how that exploitation is still in the Union’s interest. • Describe possible feedback to policy measures generated by the project that will contribute to designing, monitoring, reviewing and rectifying (if necessary) existing policy and programmatic measures or shaping and supporting the implementation of new policy initiatives and decisions. | |
| IP, not mentioned in award criteria | <ul style="list-style-type: none"> • Outline your strategy for the management of intellectual property, foreseen protection measures, such as | <ul style="list-style-type: none"> • Is the strategy for the management of intellectual property properly outlined and suitable to support exploitation of results? |

¹⁷ For further guidance on communicating EU research and innovation for project participants, please refer to the [Online Manual](#) on the Funding & Tenders Portal.

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|-------------------------------------|---|--|
| | <p>patents, design rights, copyright, trade secrets, etc., and how these would be used to support exploitation.</p> | <ul style="list-style-type: none"> ○ If exploitation is expected primarily in non-associated third countries, is it properly justified how that exploitation is still in the Union’s interest? <p>+ extra slide on IP</p> <p>Each Horizon Europe beneficiary shall use its best efforts to exploit the results it owns, or to have them exploited by another legal entity, in particular through the transfer and licensing of results. In this respect beneficiaries are required to adequately protect their results -if possible and justified -taking account of possible prospects for commercial exploitation and any other legitimate interest.</p> |

Table 6: Quality And Efficiency Of The Implementation - Work Plan, Risks And Resources

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|--|---|--|
| <p>Quality and effectiveness of the work plan,</p> | <p>Please provide the following:</p> <ul style="list-style-type: none"> • brief presentation of the overall structure of the work plan; • timing of the different work packages and their components (Gantt chart or similar); • graphical presentation of the components showing how they inter-relate (Pert chart or similar). • detailed work description, i.e.: <ul style="list-style-type: none"> ○ a list of work packages (table 3.1a); ○ a description of each work package (table 3.1b); ○ a list of deliverables (table 3.1c) ○ a list of milestones (table 3.1d); | <p>Assess the proposed work plan, and the effort and resources:</p> <ul style="list-style-type: none"> • Is the work plan of good quality and effective? • Does it include quantified information so that progress can be monitored? |

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|---|---|---|
| assessment of risks, | <ul style="list-style-type: none"> a list of critical risks, relating to project implementation, that the stated project's objectives may not be achieved. Detail any risk mitigation measures. You will be able to update the list of critical risks and mitigation measures as the project progresses (table 3.1e); | <ul style="list-style-type: none"> Are critical risks, relating to project implementation, identified and proper risk mitigation measures proposed? |
| and appropriateness of the effort assigned to work packages, and the resources overall . | <ul style="list-style-type: none"> Give full details. Base your account on the logical structure of the project and the stages in which it is to be carried out. The number of work packages should be proportionate to the scale and complexity of the project You should give enough detail in each work package to justify the proposed resources to be allocated and also quantified information so that progress can be monitored, including by the Commission Resources assigned to work packages should be in line with their objectives and deliverables. You are advised to include a distinct work package on 'project management', and to give due visibility in the work plan to 'data management' 'dissemination and exploitation' and 'communication activities', either with distinct tasks or distinct work packages. Please make sure the information in this section matches the costs as stated in the budget table in section 3 of the application forms, and the number of person months, shown in the detailed work package descriptions. a table showing number of person months required (table 3.1f); a table showing description and justification of subcontracting costs for each participant (table 3.1g); a table showing justifications for 'purchase costs' (table 3.1h) for participants where those costs exceed 15% of the personnel costs (according to the budget table in proposal part A) if applicable, a table showing justifications for 'other costs categories' (table 3.1i) if applicable, a table showing in-kind contributions from third parties (table 3.1j) | <ul style="list-style-type: none"> Does it follow a logic structure (for example regarding the timing of work packages)? Are the resources allocated to the work packages in line with their objectives and deliverables? <p>(Footnote) Do not penalise applicants that did not provide detailed breakdown of costs as they are not required.</p> |

Table 7: Quality And Efficiency Of The Implementation - Consortium

| Award criteria (General Annex D) | Template guidance Version 2.0 - 21 January 2022 | Standard briefing slides for HE evaluators Version 3.0 - 18 March 2022 |
|---|--|---|
| <p>Capacity and role of each participant, and the extent to which the consortium as a whole brings together the necessary expertise</p> | <ul style="list-style-type: none"> Describe the consortium. How does it match the project’s objectives, and bring together the necessary disciplinary and inter-disciplinary knowledge. Show how this includes expertise in social sciences and humanities, open science practices, and gender aspects of R&I, as appropriate. Include in the description affiliated entities and associated partners, if any. Show how the partners will have access to critical infrastructure needed to carry out the project activities. Describe how the members complement one another (and cover the value chain, where appropriate) In what way does each of them contribute to the project? Show that each has a valid role, and adequate resources in the project to fulfil that role. If applicable, describe the industrial/commercial involvement in the project to ensure exploitation of the results and explain why this is consistent with and will help to achieve the specific measures which are proposed for exploitation of the results of the project (see section 2.2). | <ul style="list-style-type: none"> Assess the quality of participants and the consortium as a whole: Does the consortium match the project’s objectives and bring together the necessary disciplinary and inter-disciplinary knowledge. Does the consortium include expertise in open science practices, and gender aspects of R&I, as appropriate? For topics flagged as SSH relevant, does the consortium include expertise in social sciences and humanities? Do the partners have access to critical infrastructure needed to carry out the project activities? Are the participants complementing one another (and cover the value chain, where appropriate) In what way does each of them contribute to the project? Does each of them have a valid role, and adequate resources in the project to fulfil that role (so they have sufficient operational capacity)? Is there industrial/commercial involvement in the project to ensure exploitation of the results? <p>(Footnote) Participants’ previous publications, in particular journal articles, are expected to be open access and existing datasets FAIR and ‘as open as possible, as closed as necessary’. Evaluate positively if this is sufficiently addressed.</p> |