"la Caixa" Foundation

Postdoctoral
Junior Leader fellowships

Selection process guidelines

lacaixafoundation.org

Fellowships

Table of Contents

Table of Contents 2

Introduction 3

Stage 1: Eligibility Screening 4

Stage 2: Shortlisting 4

2.1. Structure of the Panels 4

2.3. Evaluation of Candidates 6

2.3. Shortlisted Candidates 11

2.4 Feedback on the Evaluation 12

Stage 3: Interviews 13

3.1 Structure of the Committees 13

3.2 Evaluation of candidates 13

3.3 Interview Protocols 18

3.4 Selected Candidates 19

3.5 Feedback on the Evaluation 20

Annex 1: Mathematical Calculations and Formulas 21

1. Shortlisting Stage Formulas 22

2. Interviews Stage Formulas 28

Annex 1.1: Effects of normalization of scores 33

Introduction

The aim of this guidebook is to describe the evaluation process followed by "la Caixa" Foundation (LCF) fellowship programmes. It is intended for both candidates and evaluators participating in the selection processes providing detailed information about the evaluation steps, the selection criteria, the scoring system and the evaluation procedures. The evaluation process is the cornerstone of LCF programmes and it is driven by the following **key principles:**

Transparency. Candidate selection is based on clearly described rules and procedures that are available on the LCF public website. In addition, candidates receive timely information on the status of their applications at each stage of the selection process.

Equity. All candidates are treated equally, pass through the same evaluation steps and are assessed under identical evaluation criteria and procedures laid out in this document without considering any other factors.

Efficiency. LCF's fellowship programmes are characterised by thorough and rigorous compliance with the established procedures. Meeting deadlines, which are known in advance by applicants and evaluators, is of the utmost importance.

Quality. Independent experts conduct the assessments, selected based on their expertise, research performance, and evaluation experience. LCF takes proactive measures to ensure diversity among evaluators considering gender, geographical distribution, professional sector and academic disciplines. The expert database is continuously updated, ensuring one-third of evaluators rotate with each call.

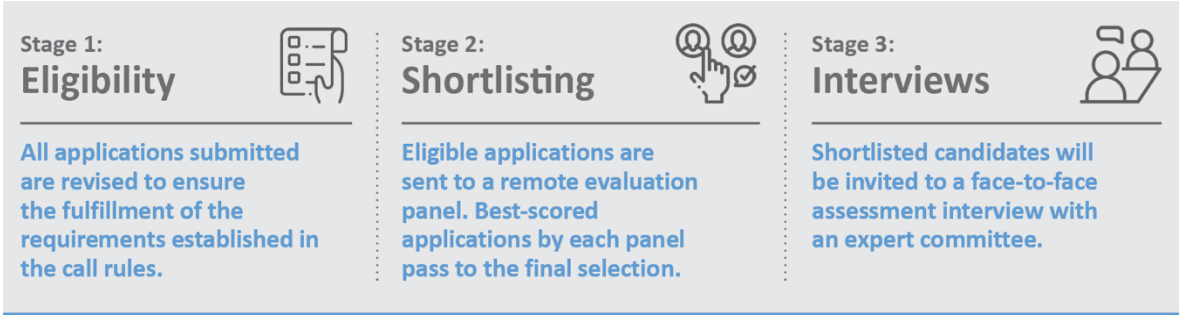
The evaluation and selection of applications takes into account the recommendations of the European Science Foundation published in the *Peer Review Guide*[1].

Likewise, the standards and principles to be followed by all evaluators who take part in the selection processes are ruled by a Code of Conduct that is publicly available on the LCF website.

LCF implements special measures to ensure that the selection process is designed to mitigate potential biases, with a particular focus on **reducing gender bias** and stereotypes. Know more about most common biases while conducting academic assessments here.

[1] *European Peer Review Guide*, European Science Foundation

The assessment process for an application comprises **three stages**:



Stage 1: Eligibility Screening

LCF checks all submitted applications to ensure the fulfilment of the requirements established in the call rules and guarantee that all applications sent for evaluation are eligible.

At this stage, candidates receive timely information about the eligibility of their proposals, and they may be contacted during the process if any information included in the application needs to be added or amended.

Stage 2: Shortlisting

The objective of the shortlisting is to select the best candidates for the final interview stage. With this in mind, the shortlisting stage has been designed with a dual purpose: to promote the best candidates to the final interview phase as well as to ensure the diversity of all disciplines considered in the programme.

2.1. Structure of the Panels

Shortlisting panels are formed based on a research field classification. Each eligible application is sent to **a remote evaluation panel made up of at least two independent experts**, mainly university professors and researchers.

When filling their application in, candidates are self-assigned to the panel that better fits their discipline, and they are evaluated according to their choice. Likewise, evaluators are assigned to the panels according to their discipline.

The composition of the remote panels is **double-blinded** to ensure independence: candidates do not know the identity of the evaluators and evaluators do not know the identity of the other evaluators. When the evaluation processes of all fellowship programmes conclude, a complete list of the participating evaluators is published on LCF website.

The panel structure is designed to ensure representativeness across disciplines. Therefore, to guarantee that candidates from all areas of knowledge are promoted to the final stage, panels with a low number of applications may be merged, grouping closely related fields when necessary. Based on this approach, evaluators will review between 10 and 20 applications.

PANEL DISTRIBUTION

Shortlisting panels are grouped under 2 broad areas of knowledge following the **structure** below:

Life Sciences (LS)

LS1: Medicine, Public Health, Sport Sciences, Nutrition, Clinical Psychology, Health Management

LS2: Animal, Plant, Environmental Biology, Physiology, Ecology and Conservation

LS3: Human Biology, Microbiology, Molecular Biology, Genetics, Cellular Biology, Genomics and Proteomics, Biochemistry

LS4: Agriculture, Veterinary Science, Animal Production, Forestry

LS5: Biotechnology, Bioinformatics, Pharmacy, Food Technology

Physical Sciences, Mathematics and Engineering (PME)

PME1: Theoretical and Applied Mathematics, Computer Sciences

PME2: Physics

PME3: Geology, Earth Sciences, Environmental and Atmosphere Sciences, Mines, Geological Engineering, Oceanography, Hydrology

PME4: Civil and Construction Engineering, Energy, Nuclear Energy and Renewable Energy Engineering

PME5: Chemistry and Chemical Engineering

PME6: Telecommunications, Electronics, Robotics, Biomedical Engineering, Automation Engineering, ICT

PME7: Industrial Engineering, Mechanical Engineering, Metallurgy, Materials Nanotechnology, Aeronautical, Naval and Aerospace Engineering

2.3. Evaluation of Candidates

The candidates' shortlisting is conducted remotely through an online platform specifically designed for this purpose. After logging into the platform, evaluators review and assess all their assigned applications.

The application evaluation mainly consists of **three parts**:

Expertise Level | *Evaluation Criteria and Scoring* | *Justification of the evaluation*

EXPERTISE LEVEL

All evaluators must indicate, for each application assessed, their level of expertise in the discipline of the application.

The definition of the expertise level is:

Level 1: The evaluator's expertise corresponds with the discipline of the application.

Level 2: The evaluator's expertise may not correspond with the discipline of the application. However, their background allows a proper assessment.

Evaluators' expertise level weights the scores accordingly: level 1 experts have a greater impact on the candidate's shortlisting score than level 2 experts.

By default, all evaluators are labelled as level 2. Evaluators with a higher level of expertise must select level 1.

Expertise Level | *Evaluation Criteria and Scoring* | *Justification of the evaluation*

EVALUATION CRITERIA AND SCORING

For each application, three evaluation criteria must be assessed and scored using the following scale:

A final score is obtained by adding the scores assigned to each criterion, considering the weight of each criterion as well as the level of expertise selected by each evaluator.

Each evaluator must score, for the same application, **three criteria**:

50% Excellence of Curriculum Vitae
30% Quality and potential of the research project
20% Impact

1. EXCELLENCE OF CURRICULUM VITAE (50%)

Summary: *The assessment will focus on the quality, relevance, and influence of the candidate's contributions, considering the diversity of outputs and the candidate's specific role. The evaluation will consider the career trajectory in relation to the discipline, career stage, and opportunities available, including non-linear paths. Progression, coherence, mobility, and potential for independence, creativity, and leadership will be assessed.*

Specifically, the main elements assessed will be:

» **Research contributions and outputs.** Quality, originality, and influence of the candidate's contributions will be assessed, including publications and other relevant outputs (e.g. open research outputs, software, datasets, patents, prototypes, and societal or policy contributions), taking into account the candidate's specific role in joint outputs and recognising non-traditional contributions (e.g. large collaborations, interdisciplinary or translational research), without reliance on journal-level metrics, number of publications, citation counts, or author position.

» **Career context and trajectory.** The candidates' track record will be assessed in relation to their career stage, the scientific discipline, and the opportunities available. Candidates at an earlier stage of their career, or those who have experienced career breaks, mobility limitations, caring responsibilities, or other non-linear career paths, will be duly taken into account and will not be penalised for having a shorter track record.

The assessment will also consider the extent to which the candidate demonstrates progression appropriate to their field, as well as any relevant interdisciplinary or intersectoral experience, and evidence of mobility, adaptability, and capacity to benefit from the programme. Any changes in the career path should be appropriately justified and supported, and any professional experience should be clearly explained in terms of its relevance to the proposed project.

» **Researcher's potential and independence.** Evidence of creativity, intellectual maturity, problem-solving capacity, and independent thinking will be considered, together with early signs of leadership (e.g. mentoring, supervision, contributions to team activities, or coordination roles).

2. QUALITY AND POTENTIAL OF THE RESEARCH PROJECT (30%)

Summary: *The assessment will focus on the scientific excellence, originality, and ambition of the project, including its potential to advance knowledge beyond the state of the art. The evaluation will consider the clarity and coherence of the objectives and methodology, as well as the feasibility and robustness of the work plan, including risks. The suitability of the host environment will be assessed in terms of its fit with the project and the candidate's training needs.*

Specifically, the main elements assessed will be:

- **Scientific excellence.** The scientific excellence of the proposal will be assessed in terms of its novelty, originality, and ambition, as well as its potential to advance knowledge beyond the current state of the art. The evaluation will consider the clarity and pertinence of the objectives, hypotheses, and scientific rationale, and the extent to which the proposal is coherent, well-structured, and grounded in the relevant state of the field. Particular attention will be given to the innovative aspects of the proposed research including its potential to generate new knowledge within the scientific discipline, as well the use of new technologies or methodologies where appropriate.

- **Feasibility and methodological robustness.** The coherence between the research objectives and the proposed methodology, and the soundness of the approach, will be assessed. The assessment will consider the coherence, appropriateness, and feasibility of the proposed methods and work plan, including the identification of potential risks and the appropriateness of mitigation measures. Where relevant, the integration of the sex and/or gender dimension will be taken into account.

- **Suitability of the host environment.** The suitability of the host environment will be assessed in terms of the quality of the fit between the project, the host institution, and the expertise of the host group and supervisor. The evaluation will consider the extent to which the chosen environment provides appropriate conditions and support for the implementation of the project and the candidate's training, including, where relevant, planned secondments. The experience and expertise of the supervisor(s) will also be taken into account. Evidence of engagement with new professional or scientific environments, as well as interdisciplinary or intersectoral dimensions, will also be considered where relevant.

3. IMPACT (20%)

Summary: *The assessment will focus on the potential impact of the project, including its contribution to knowledge, methodological innovation, and, where relevant, broader societal or economic benefits. Differences across disciplines will be taken into account, including indirect or long-term impact in basic sciences. The added value of the fellowship for the candidate's development and career progression will also be assessed.*

Specifically, the main elements assessed will be:

» **Added value for the researcher** The extent to which the fellowship contributes to the candidate's development, independence, and competitiveness for future funding and career opportunities will be assessed, taking into account the candidate's career stage and current position.

» **Project potential impact** The potential of the project to generate impact will be assessed in its broadest sense, including the advancement of scientific knowledge, methodological innovation, and open science practices, as well as, where relevant, benefits to society, policy, industry, or specific user communities. The evaluation will recognise that the nature and timescale of impact vary across disciplines. In particular, in basic sciences, where immediate societal impact is not expected and the impact may be indirect, long-term, or enabling.

ADDITIONAL EVALUATION

In addition to the three criteria mentioned above, evaluators must assess **four additional aspects**:

1	2	3	4
The candidate's academic and / or professional potential	*The societal impact of the statement of purpose*	*The impact of the fellowship on the candidate's trajectory*	*Expository clarity demonstrated in the statement of purpose*

Each of these aspects must be scored using the following scale:

Poor	Acceptable	Good	Very Good	Exceptional

The additional evaluation provides complementary information to the final interview evaluators. Moreover, it may be used to break draws in case of equal scores between candidates.

JUSTIFICATION OF THE EVALUATION

Each evaluator must provide a written rationale (qualitative evaluation) for the different criteria for each application, along with a concise brief that includes the reasoning behind their evaluation, as well as the proposal's strengths and weaknesses. The main objective of the rationale is to guarantee that the assessment was thorough and specifically tailored to each proposal, and to provide the information needed to generate a qualitative feedback report for the candidates.

The use of AI tools by evaluators in writing their comments is not permitted, except for translation or grammar correction purposes, provided that such tools do not compromise the confidentiality of the information. In all cases, the content of the report must be original and authored by the expert.

*Those comments and observations **will not be reviewed or filtered** by LCF which is why evaluators should be extremely careful with their wording. In any case, comments should have a strictly professional and constructive tone.*

After completing the three steps in the shortlist stage, experts must submit their evaluations within the established deadline.

Once all evaluations for a candidate have been submitted, **an AI-based tool will compile the individual reports into a consolidated summary**. This compiled report will then be reviewed by a human expert to ensure consistency, accuracy, and alignment with the original evaluations. All comments will be made accessible to candidates and members of the Selection Committee, along with the overall score of the proposal.

DISCREPANCIES

Once the evaluations are submitted by each panel, the system may detect significant discrepancies among experts' scores for the same application. If any, these applications are referred to those experts to review their original scores if deemed appropriate[2] within the established deadline.

[2] *For more information about the detailed calculations of this aspect, see section 1.2 of the Annex 1.*

2.3. Shortlisted Candidates

LIST OF SHORTLISTED CANDIDATES

The shortlist of candidates is not based on consensus or discussion among evaluators. It is an individual assessment. Therefore, the ranking of shortlisted candidates results from the aggregation and weighting of the scores given by the evaluators to each application, sorted by highest to lowest score on each shortlist panel.

The number of **shortlisted candidates who passes to the final stage** depends on the number of fellowships to be awarded and the distribution of applications received by discipline. Once the number of candidates to be shortlisted is settled, the shortlisting involves **two steps**:

STEP 1 / Selection of 70% of candidates to be shortlisted: Best scored candidates by each remote panel are selected following a proportional distribution.

See example below:

Interview Committee X

Total candidates to be shortlisted: **45** > **Step 1** / 70% of the candidates to be shortlisted: **32**

Shortlisting panels

Shortlisting Panel	Eligible applications	% over total eligible applications	Proportional Distribution (70%)
Panel 1	45	24.59%	8 (7.86)
Panel 2	30	16.39%	5 (5.24)
Panel 3	40	21.86%	7 (6.99)
Panel 4	15	8.20%	3 (2.62)
Panel 5	11	6.01%	2 (1.92)
Panel 6	22	12.02%	4 (3.84)
Panel 7	20	10.93%	3 (3.50)
Total	**183**	**100%**	**32**

Step 2 / Selection of 30% of candidates to be shortlisted: The remaining applications that were not shortlisted in the previous step are grouped under a single ranked list per committee[3] (after normalising their scores). The best scored applications on this list are shortlisted regardless of the panel they have been self-assigned to. Non-shortlisted applications remain in a single waiting list per committee.

For more information about the specific calculations of the ranking see *List of shortlisted candidates and single reserve list* in Annex 1.

This methodology has a dual purpose: to guarantee excellence and representation. It guarantees that the best candidates from each discipline are selected (70%) while ensuring that best candidates are likewise selected regardless of their discipline (30%).

[3] *See section 2.1 Structure of the Panels*

DRAWS

In the event of draws involving two or more applications, these will be resolved considering the final score in each individual criterion prioritised according to their weight (C1>C2>C3). Firstly, a comparison of the scores of C1 will be made. If the draw persists, the same process will be followed considering C2 and so forth.

If the draw persists, it will be resolved by introducing the results of the additional evaluation aspects given by each evaluator[4].

2.4 Feedback on the Evaluation

To maximise transparency, the following **information** is released once the shortlist stage is concluded:

- **Feedback to Candidates**

  All candidates receive information on their score, position within the panel and general statistics of the selection process.

  In addition, all candidates obtain information on the quartile in which their application is located, for each criterion evaluated, in comparison to the rest of applications assessed by the same panel.

  Finally, non-preselected candidates obtain a qualitative report according to the justifications of the evaluations carried out by the experts. This report is written by an external figure to the ”la Caixa” Foundation called "Reporter" who will make a report based on the information contained in the evaluation of each expert. The "reporter" does not act in any case as evaluator of the application and is totally impartial in the writing of the report.

- **Feedback to Shortlisting Evaluators**

  Shortlisting evaluators are provided access to anonymized scores and comments from their fellow experts within the same panel.

- **Feedback to Interview Committee Evaluators**

  Evaluators who take part in interview committees will also have access to scores, the information shared with the candidates and anonymized comments from the shortlisting evaluators to support the final assessment.

[4] *For more information about the detailed calculations of this aspect, see section 1.4 of the Annex 1.*

Stage 3: Interviews

Shortlisted candidates are invited to an interview as the final stage in the selection process. The overall purpose of the interview is to select the candidates with the highest potential according to the selection criteria. This process is specifically designed to mitigate biases and ensure objectivity and efficiency.

3.1 Structure of the Committees

The number of committees will be determined based on the number of applicants called to interviews within each disciplinary field. These multidisciplinary committees will be formed by 4 to 8 university professors, researchers or professionals with expertise in the disciplines assessed. Each committee will be chaired by an officer from LCF who will moderate the session and ensure that the interviews are carried out according to the scheduled plan.

Following the same structure as the shortlisting stage, committees are based on **two areas of knowledge**:

Life Sciences

Physical Sciences, Mathematics and Engineering

A maximum of candidates to be interviewed per committee is established. If needed, the committees can be split or merged to adjust their capacity. In case of merging, close-related fields will be considered.

If committees are split, candidates will be distributed among the different subcommittees in a sequential manner according to their shortlisting score. Therefore, the candidate with the highest score will be assigned to subcommittee 1, the next one to subcommittee 2, and so forth until all candidates have been distributed.

3.2 Evaluation of candidates

PREPARATION OF INTERVIEWS

Prior to the interview, evaluators will be provided access to an online platform with all necessary information about the interviewed candidates. This information includes general statistics of the selection process, scores, position, and evaluation comments for each candidate from the shortlisting stage as well as specific guidelines to conduct the evaluation and general information of the call.

Expertise Level | *Evaluation Criteria and Scoring* | *Justification of the evaluation*

EXPERTISE LEVEL

In the same way as in the previous stage, all evaluators must indicate, for each application assessed, their level of expertise in the discipline of the application.

The definition of the expertise level is:

Level 1: The evaluator's expertise corresponds with the discipline of the application.

Level 2: The evaluator's expertise may not correspond with the discipline of the application. However, their background allows a proper assessment.

The evaluators' expertise level weights the scores accordingly: level 1 experts have a greater impact on the candidate's final score than the level 2 experts.

By default, all evaluators are labelled as level 2. Evaluators with a higher level of expertise must select level 1.

Expertise Level | *Evaluation Criteria and Scoring* | *Justification of the evaluation*

EVALUATION CRITERIA AND SCORING

For each application, three evaluation criteria must be assessed and scored using the following scale (including decimals):

Each evaluator must score, for the same application, **three criteria**:

50% Academic and Professional Career
30% Candidate's Potential
20% Motivation and Impact of the Project

1. ACADEMIC AND PROFESSIONAL CAREER (50%)

Summary: The assessment will focus on the depth, coherence, and influence of the candidate's academic and professional trajectory, based on qualitative evidence. The evaluation will consider the quality and significance of research contributions, experience in conducting and managing research activities, and the coherence of the career path in relation to the proposed project. Progression, interdisciplinarity, and non-linear trajectories will be taken into account.

Specifically, the main elements assessed will be:

- **Professional experience and research management.** The candidate's professional and academic experience will be assessed, including prior training to carry out their proposal. The evaluation will consider the extent and quality of experience in managing and conducting research activities, including coordination of tasks, contributions to collaborative projects, supervision of students, and engagement in scientific service (e.g. peer review or community activities). The assessment will also take into account the candidate's ability to manage research resources and planning, as well as the recognition and relevance of their contributions to the field, including, where applicable, involvement in project management, lectures at congresses, awards, competitive funding and other merits.

- **Quality and significance of contributions to the field.** The quality, originality, rigour, and influence of the candidate's research contributions, as described during the interview, will be assessed. Particular attention will be given to significance and depth of the candidate's achievements within their specific career context, including early-stage researchers who will not be penalised for having a relatively short track record.

  The evaluation will consider a broad range of research outputs, including experimental, theoretical, computational, methodological, collaborative, open science, software, datasets, patents, translational contributions. The assessment will not rely on publication counts, journal prestige, citation metrics, or author position.

- **Coherence of the research trajectory.** The consistency and focus of candidates' trajectory will consider the extent to which the career path demonstrates a logical progression and alignment with the proposed project. Non-linear trajectories, career breaks, mobility constraints, or transitions across sectors will be duly considered and will not be penalised, provided they are coherently explained. Interdisciplinary or intersectoral paths will be positively considered where they are relevant and well justified Any professional experience should be clearly explained in terms of its relevance to the proposed project.

2. CANDIDATE'S POTENTIAL (30%)

Summary: The assessment will focus on the candidate's potential to develop into an independent, creative, and responsible researcher. The evaluation will consider reasoning and communication skills, including the ability to articulate complex ideas clearly and respond critically. Evidence of creativity, originality, leadership, and entrepreneurial mindset will be assessed, together with the ability to work effectively in collaborative and interdisciplinary environments.

Specifically, the main elements assessed will be:

- **Reasoning and communication.** Clarity, coherence, and precision in presenting scientific ideas, including the ability to clearly and precisely communicate complex reasoning and very specific matters, so that the ideas introduced can be understood in a clear and structured manner. The evaluation will consider the candidate's capacity to articulate sound scientific reasoning, respond critically to questions, and present ambitious and yet rrealistic ideas aligned with the scope of their project

- Leadership, originality and creativity**.** The candidate's creativity, intellectual curiosity, and capacity to develop original and innovative ideas will be assessed, including the ability to explore new approaches or address unexplored areas beyond the state of the art. The evaluation will consider evidence of emerging leadership, independence and entrepreneurial mindset including the ability to take initiative, make informed decisions, inspire, organise, collaborate, or mentor. Particular attention will be given to the candidate's potential to generate new knowledge and develop novel approaches or technologies.

- **Interpersonal and collaborative competences** Ability to work effectively in collaborative, interdisciplinary, and diverse environments. The evaluation will consider the candidate's capacity for teamwork, openness, and willingness to learn, as well as their ability to contribute constructively to group dynamics.

3. MOTIVATION AND IMPACT OF THE PROJECT (20%)

Summary: The assessment will focus on the clarity and coherence of the candidate's motivation, including the alignment between the project, the host institution, and their career objectives. The evaluation will consider how convincingly the candidate articulates the expected impact of the project, taking into account disciplinary differences. The added value of the fellowship for the candidate's development and future career will also be assessed.

Specifically, the main elements assessed will be:

- **Motivation and clarity of purpose.** The strength, clarity, and coherence of the candidate's motivation will be assessed, including their rationale for selecting the project, the host institution, and the Junior Leader programme, and how these choices align with their longer-term objectives. The evaluation will consider the extent to which the candidate demonstrates a clear understanding of the project's objectives, structure, and underlying rationale, as well as its feasibility within the proposed timeframe.

- **Expected impact (scientific, societal, technological or enabling)** The evaluation will consider how clearly and convincingly the candidate articulates the expected impact of the project, including its potential scientific contribution and, where relevant, broader societal, technological, or policy impact.

  The evaluation will recognise that the nature and timescale of impact vary across disciplines. In particular, in basic sciences, where immediate societal impact is not expected and the impact may be indirect, long-term, or enabling. Consideration will also be given to applications that entail contact with new academic, cultural, or scientific environments as well as interdisciplinary and intersectoral approaches.

- **Added value for the researcher** The extent to which the fellowship will contribute to the candidate's development, independence, and competitiveness for future opportunities will be assessed. The evaluation will also consider the candidate's awareness of the challenges and risks associated with the project, and their capacity to address them, as well as their determination to successfully complete the project within the established duration of the fellowship.

3.3 Interview Protocols

FORMAL ASPECTS

The round of interviews will be conducted in strict accordance with the following formal requirements:

- **Punctuality:** Utmost punctuality is expected. The interviews follow a very precise schedule, and no flexibility is allowed in relation to this time schedule.
- **Duration:** Each interview will last 30 minutes, and it will start with a brief presentation by the candidate (10 minutes) and followed by questions asked by the evaluation committee (20 minutes).
- **Language:** Interviews are conducted entirely in English.
- **No supporting materials:** audio-visual materials, presentations or documents will not be allowed during the interview. Nor will the committee accept any documents that have not been included in the application.

No show: Failing to attend the interview entails that the candidate will not be allowed to apply to future calls, except in case of duly justified force majeure.

OPENING OF THE INTERVIEW

The LCF Officer will welcome the candidate and then, the candidate will start the presentation. To ensure independence, the composition of the committee is **blinded**, which means that candidates do not know the identity of the evaluators. For this reason, committee members will not be introduced to the candidates. After the candidate's presentation, the committee members will ask the questions they deem relevant to properly assess the application.

QUESTIONS

There are no specific guidelines to conduct the selection interview. Experts are entitled to establish their own dynamics and tone, depth and scope of the questions asked to each candidate.

Questions raised in the interview aim at testing candidates' capabilities, broadening the information provided in the application and clarifying any aspects that were insufficiently addressed in the application.

3.4 Selected Candidates

RANKING

At the end of each interview, evaluators should score each candidate, according to the selection criteria established. Once all the interviews are concluded, evaluators will give each candidate three scores, one for each criterion.

The selection of candidates is not based on consensus or discussion among evaluators. It is an individual assessment. Therefore, the ranking results from the aggregation and weighting of the scores given by the evaluators to each application, sorted by highest to lowest score on each committee.

Shortlisting Score: The score obtained in the shortlisting stage will be included in the final score weighted as an additional evaluator with expertise level 1.

DISCREPANCIES

Evaluators will be called to revise discrepancies among the scores of candidates in the cut-off threshold, if any, and adjust them if deemed appropriate[5].

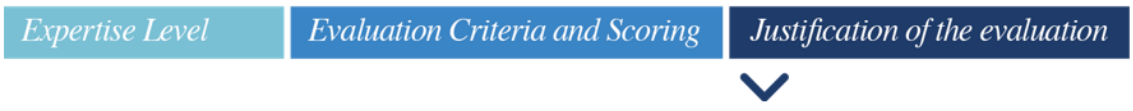
DRAWS

In an event of a draw, it will be resolved by the experts who form the selection committee.

FINAL LIST

Once the process is concluded, all evaluators must ratify the final ranking of the fellowships awarded and the candidates on the waiting list.

The official list of fellows and wait-listed candidates will be published on the LCF website within the deadline established in the rules for participation.



JUSTIFICATION OF THE EVALUATION

At the end of interviews, evaluators must provide a rationale for each application with a short, concise written brief (max. 1.000 characters), which includes the reasoning behind their evaluation and overall impression given by the candidate emphasizing the strengths and weaknesses of each candidate. This rationale must be informed in the online evaluation platform.

[5] *For more information about the detailed calculations of this aspect, see section 2.4 Annex 1.*

The use of AI tools by evaluators in writing their comments is not permitted, except for translation or grammar correction purposes, provided that such tools do not compromise the confidentiality of the information. In all cases, the content of the report must be original and authored by the expert.

Once all evaluations for a candidate have been submitted, an AI-based tool will compile the individual reports into a **consolidated summary**. This compiled report will then be reviewed by a human expert to ensure consistency, accuracy, and alignment with the original evaluations.

3.5 Feedback on the Evaluation

To enhance transparency, the following information is released once the interviews stage is concluded:

» **Feedback to Candidates**

Candidates receive information on their score, position within the panel and general statistics of the selection process.

In addition, candidates obtain information on the quartile in which their application is located, for each criterion evaluated, in comparison to the rest of applications assessed by the same panel.

LCF has no further information on the assessment than the information disclosed to each candidate. Once the evaluation processes of all fellowship programmes have been completed, a full list of the participating evaluators is published on the LCF website.

Annex 1:
Mathematical Calculations and Formulas

This section is aimed to describe the mathematical calculations and formulas behind the different stages of the selection process to maximize transparency and clearness.

The following mathematical procedures rule the selection process:

Scoring	Collecting the candidates' scores given by each of the experts and weighted according to the corresponding evaluation criteria.
Discrepancies	Reviewing the candidates' scores for whom there are significative differences between expert assessments.
Normalization	Normalizing the scores to mitigate the differences in scale and dispersion between different experts.
Expertise	Weighting scores according to the expertise acknowledged by the evaluators.
Draws	Resolving draws between candidates with equal scores.
Reserve list	Establish a reserve list to retrieve candidates in case of a withdrawal.
Feedback to candidates	Presenting the scores and information to make it available to candidates.

1. Shortlisting Stage Formulas

1.1. SCORING

Every application in a panel is reviewed by a certain number n of *evaluators* (usually between 2 and 4), who are independent experts in the discipline specific for that panel or a close - related disciplinary field. The evaluation coming from each expert, for a given candidate, consists of three scores between 1 and 8 (admitting decimals), corresponding to three different *criteria*. We call these the *primary scores,* and we denote them by

$$score1(\mathrm{c},\ \mathrm{e},\ \mathrm{crit})\ \in [1,8],\ \text{with}\ \ crit \in \{1,2,3\}\,,\ e\ \in \{1,2,\dots,\ n\}$$

which designates the primary score of the candidate c, given by the evaluator e, for the criterium $crit$.

Every call may indicate specific weights for the three different criteria, and we denote them by

$$weight(crit),\ \ \text{with}\ \ \ crit \in \{1,2,3\}$$

Then, the added scores of a given candidate for a given evaluator is computed as follows:

$$\mathrm{score1.sum}(c,e) = \sum_{crit=1}^{3} score1(c,e,crit)\cdot \mathrm{weight}(crit)$$

At this stage, we calculate the candidate's score as follows:

$$score(c) = \frac{1}{n}\sum_{e=1}^{n} score(c,e)$$

which is the average of all expert scores given to the candidate. However, this value will not be used until step *1.5. Feedback to Candidate.*

1.2. DISCREPANCIES

The evaluation system identifies significant discrepancies among experts' scores for the same application. When detected, these applications are referred back to the corresponding experts to review and adjust the scores, if deemed appropriate. The detection process involves the following two steps:

1. The primary score $score(c,e)$ of each candidate given by each evaluator, is recalculated by subtracting the evaluator's mean, and dividing it by his standard deviation. That is,

   $$score_{norm}(c,e) = \frac{score(c,e) - mean_e}{stdev_e}$$

   where $mean_e$ denotes the average of all primary scores given by evaluator e and

   $$stdev_e = \sqrt{\frac{\sum_{c=1}^{n}(score(c,e) - mean_e)^2}{n-1}}.$$

In this way the new average of all scores is set to 0 and the standard deviation is 1, so allowing a better comparison among scores.

2. For each candidate we compute the difference between the highest and the lowest normalized score among all those obtained from the different evaluators. That is

$$diff(c) = max_e\big(score_{norm}(c,e)\big) - min_e\big(score_{norm}(c,e)\big)$$

where:

- $score_{norm}(c,e)$ is the normalized score given to candidate by evaluator.
- max_e and min_e represent the maximum and minimum scores given by any evaluator for candidate.

If this difference is equal or larger than 2, the scores of this candidate are considered discrepant.

RANKINGS AND SINGLE RESERVE LIST

1.3 NORMALIZATION OF SCORES

To be able to compare the scores of the candidates assessed by different evaluators, they are normalized according to the following procedure:

1) Calculate the average of the scores of all candidates assessed by each evaluator:

$$mean_e = \frac{1}{n}\sum_{c=1}^{n} score(c,e)$$

2) Compute the standard deviation of the evaluator:

$$stdev_e = \sqrt{\frac{\sum_{c=1}^{n}(score(c,e) - mean_e)^2}{n-1}}$$

3) Adjust the standard deviation to control the dispersion[6] of the scores when *stdev* << 1:

$$stdev_e = \begin{cases} thrs \text{ if } stdev_e < thrs \\ stdev_e \text{ if } stdev_e \geq thrs \end{cases}$$

where:

$$thrs = \frac{stdev_{committee}}{2}$$

4) Finally, standardise the score of each candidate according to the mean and standard deviation of the evaluator. Hence, if the candidate *C* has been evaluated by the evaluator *e*, then

$$score_{norm}(c,e) = \frac{score(c,e) - mean_e}{stdev_e}$$

6. For more details about the implementation of this lower bound see Annex 1.1.

1.4 EXPERTISE

To determine the ranking of shortlisted candidates, the expertise of the evaluators with the discipline of the candidate assessed is considered. Experts indicate their expertise level for each application assigned and their scores are weighted accordingly.

To do so, the final normalised score of a candidate is the result of averaging the normalised scores obtained from the evaluators, weighted by the different expertise levels of each of them**.** More precisely:

» If the expertise levels of all n evaluators coincide, we compute the simple mean

$$score_{norm}(c) = \frac{1}{n}\sum_{e=1}^{n} score_{norm}(c,e)$$

» If the expertise levels do not coincide, we compute a weighted mean, where an additional weight of 0.5 is divided between the experts with Level 1. In other words,

$$score_{norm}(c) = \sum_{e=1}^{n} score_{norm}(c,e) \cdot weight(c,e)$$

where,

$weight(c,e) = \frac{1}{n+0,5}$ if the expertise of this evaluator is Level 2, and

$weight(c,e) = \frac{1+0,5/m}{n+0,5}$ if the expertise of this evaluator is Level 1, and there are m evaluators with Level 1.

Note that if all evaluators declare the same level of expertise (that is if $m = 0$ or $m = n$) then all of them have weight $\frac{1}{n}$ and hence the weighted mean equals the regular mean.

Example

There are 3 experts in the panel ($n = 3$). Evaluators 2 and 3 have indicated Level 2 while evaluator 1 has indicated Level 1. Then

$$weight(c,1) = \frac{1}{3,5} = 0,29, weight(c,2) = weight(c,3) = \frac{1,25}{3,5} = 0,36.$$

At this stage, using the procedure described above, every candidate has a final normalised score denoted as $score_{norm}(c)$. This score reflects all normalised scores obtained from evaluators on the candidate's panel considering their expertise level. Ordering the candidates based on this final score **allows to have a ranking of all candidates of each panel.**

1.5 LIST OF SHORTLISTED CANDIDATES -DRAWS AND RESERVE LIST

The candidates to be shortlisted depends on the number of fellowships to be awarded. Each call establishes a predefined number of candidates per committee who will be promoted to the interview phase. Given a particular committee, we define:

N =Number of candidates to be promoted to the face-to-face interview;

P = Number of panels associated to the given committee;

$can(P)$ = Number of candidates assigned to the panel P;

Total number of applications in the given committee, calculated as:

$$C = \sum_{P} can(P)$$

The N shortlisted applicants are selected in two steps considering that:

N= N1 + N2

N1: Selection of 70% of candidates to be shortlisted

In the first step N_1 candidates will be promoted, where N1 equals 70% of the total of *N* candidates, rounded to the nearest whole number. That is,

$$N_1 = round(0.7N)$$

The first N_1 candidates are shortlisted proportionally to the number of applications $can(P)$ compared to the total C. In this way, the first

$$f \in (P) = round\left(\frac{can(P)}{C} \cdot 0.7N\right)$$

candidates ranked in the panel *P* are shortlisted and pass to the final stage, where this number is also rounded to the nearest whole number. In case the rounding gives 0, at least one candidate will be assigned[7].

The best scored applications of each panel will be the N1 candidates shortlisted.

N2: Single reserve list with the remaining 30% of candidates

In the second step, the remaining 30% of candidates ($N_2 = N - N_1$) will be chosen.

To select the N_2 candidates, a single ranking will be made. This ranking is established by joining the candidates of the panels belonging to the same committee that have not been shortlisted within *N1*. This single ranking is made according to the normalised scores of the candidates, computed in the previous step. With these normalised scores, the best scored applications on this list (N_2) are shortlisted regardless of the panel they have been self-assigned to. Likewise, non-shortlisted applications remain in a single waiting list per committee.

7. If the rounding system leads to a total larger (resp. smaller) than the 70% of N, the candidate in excess (resp. defect) will be removed from (resp. assigned to) the panel with the lowest (resp. highest) value of finP *before rounding.*

In case two or more final scores coincide in the ranking above, the system uses the normalised scores obtained in each of the three criteria separately, to resolve the draw. Specifically, this is done as follows:

» Every candidate's normalised score is divided into three normalised scores, one for each criterion, computed by adding the normalised scores of all evaluators, weighted by their expertise level. In other words,

$$score_{nom}(c, crit) = \sum_{e=1}^{n} score_{norm}(c, e, crit) \cdot weight(e, c)$$

where we recall that $score_{norm}(c, e, crit)$ is the normalised score given by evaluator e, to candidate c , for criterion $crit$, and it is calculated in the same way as $score_{norm}(c, e)$ but for each criterion.

Specifically, given a criterion $crit \in \{1,2,3\}$ and an expert $e \in \{1, \dots, m\}$ the normalization is carried out by first calculating the average and the standard deviation of each evaluator in the set of *n* candidates and for each criterion.

$$mean_{e,crit} = \frac{1}{n}\sum_{c=1}^{n} score(c, e, crit)$$

$$stdev_{e,crit} = \sqrt{\frac{\sum_{c=1}^{n}\left(score(c, e, crit) - mean_{e,crit}\right)^2}{n-1}}$$

where $score(c, e, crit)$denotes the score of the candidate c obtained from the expert e for the criterion $crit$.

In the same way as at *Section 1.3 Rankings and single reserve list*, dispersion of scores when $stedv \ll 1$ must be adjusted. Therefore:

$$stdev_{e,crit} = \begin{cases} thrs \text{ if } stdev_{e,crit} < thrs \\ stdev_{e,crit} \text{ if } stdev_{e,crit} \geq thrs \end{cases}$$

where:

$$thrs = \frac{stdev_{committee,crit}}{2}$$

Then, this score ($score(c, e, crit)$ is normalized as follows:

$$score_{norm}(c, e, crit) = \frac{score(c,e,crit) - mean_{e,crit}}{stdev_{e,crit}}$$

» The criterion with maximum weight is the one which is used to resolve the draw. If the draw persists, the criterion with the second highest weight will be used, and so on until the last criterion is reached.

» If the draw still persists, the additional evaluation is considered: Each evaluator assesses four qualitative aspects for each candidate, that are translated into numerical values between 1 and 5:

Qualification	Poor	Acceptable	Good	Very Good	Exceptional
Numerical Value	1	2	3	4	5

For this purpose, the system computes the total sum of these values given by all evaluators of the panel to each candidate and uses this score to resolve the draw.

1.6 FEEDBACK TO CANDIDATES

To provide comprehensible feedback on their assessment in the shortlisting stage, candidates will receive their normalized total score rescaled and the quartile in which they are ranked compared to the other candidates in their panel. The ranking will not be affected by any of the subsequent steps.

» Total normalized score
Candidates will be provided with their total normalized score rescaled in the interval *[1, 8]*. To rescale the scores $score_{norm}(c)$, we first rescale to obtain a temporary score from 0 to 1:

$$temp.score(c) = \frac{score_{norm}(c) - min_{comite}\big(score_{norm}(c)\big)}{max_{comite}\big(score_{norm}(c)\big) - min_{comite}\big(score_{norm}(c)\big)}$$

That is, to the candidate's final normalised score, we subtract the minimum normalised score among all candidates and divide by the difference between the maximum and the minimum normalised score, again among all candidates. Every score is now between 0 and 1.

Now, a second rescaling calculation is made to obtain the **rescaled final score of each candidate**:

$$score_{resc}(c) = temp.score(c) \times \Big(max_{committee}\big(score(c)\big) - min_{committee}\big(score(c)\big)\Big) + min_{committee}\big(score(c)\big)$$

Therefore, all scores are translated to the interval

$\big[min_{committee}\big(score(c)\big) max_{committee}\big(score(c)\big)\big]$, where *score(c)* is calculated in section *1.1 Scoring.*

» Quartile distribution
The candidate will be informed of the quartile assigned for each of the selection criterion assessed. To determine this position, the rankings of each panel and each criterion are divided into four equal parts or quartiles $Q1$, $Q2$, $Q3$ and $Q4$, where $Q1$ corresponds to the top group of the $n/4$ highest normalised scores, and $Q4$ to the bottom group with the $n/4$ lowest ones.

To provide this position, normalised scores of each criterion are used in the same way than explained in Section *1.4 Draws.*

2. Interviews Stage Formulas

2.1 SCORING

Once the interviews have concluded and experts have scored all candidates, the system considers the weight of each criterion evaluated and calculate an initial score for each candidate from each expert. The final scores generated in the shortlisting stage are included as if they came from one additional expert in the committee with Level 1 of expertise:

$score(c, e)$= score given to candidate c by expert e.

Assuming that there are n candidates and m evaluators (including the shortlisting score), then $c \in \{1, \dots, n\}$ and $e \in \{1, \dots, m\}$.

Consequently, every candidate has m scores: one from each expert, plus the one coming from the shortlisting stage. These scores take values from 1 to 8.

2.2 NORMALIZATION

The normalization or standardization of scores given by one expert in relation to all candidates evaluated is performed according to the following procedure. For each evaluator $e \in \{1, \dots, m\}$:

- The expert's mean score is calculated

$$mean_e = \frac{1}{n}\sum_{c=1}^{n} score(c, e)$$

- The standard deviation of this same set of scores is also obtained from

$$stdev_e = \sqrt{\frac{\sum_{c=1}^{n}(score(c,e) - mean_e)^2}{n-1}}$$

- Finally, the set of scores is of every evaluator (also the ones coming from the shortlist stage) are normalized by

$$score_{norm}(c, e) = \frac{score(c, e) - mean_e}{stdev_e}$$

To normalize the scores coming from the shortlist stage, they are treated as if they were evaluated by the same evaluator.

Observations

With this procedure, the original scores

$$score_{c,e} \in [1,8]$$

are converted in new quantities

$$score_{norm}(c, e) \in (-\infty, \infty)$$

The mean of the new scores of each expert is 0 and its standard deviation is 1. In this way, the possible different tendencies of the evaluators (giving higher or lower scores in general, for example) are eliminated. The new scores will be higher or lower, depending on how far they are from the average of the original scores, and how frequent this distance is. (See *Annex 1.1: Effects of normalization in the scores* for further details about the effects of normalization).

2.3 EXPERTISE

Experts have declared an expertise level with the specific discipline of the candidate assessed.

Supposing we have *m* evaluators (we are including here the shortlisting score), the weights would be distributed in the following way:

- Every expert has an ensured weight of $\frac{1}{m+1}$ and moreover
- there is an additional weight of $\frac{1}{m+1}$ to be uniformly distributed among those experts with Level 1 of expertise (*k evaluators*), among which we always find the shortlisting score.

Hence,

$$weight(c,e) = \frac{1}{m+1} \text{ if the expertise of this evaluator is Level 2, and}$$

$$weight(c,e) = \frac{1}{m+1} + \frac{1}{k(m+1)} \text{ if the expertise of this evaluator is Level 1}$$

Example 1:

A committee is formed by 5 experts: $e \in \{1,2,3,4,5\}$.

Experts 3 and 5 have declared Level 1 of expertise for a certain candidate *c* (hence $k = 3$). As a result, the weights are distributed as follows:

Expert e	Level of Expertise	$\mathbf{weight_{c,e}}$	$\mathbf{weight_{c,e}}$ (num)
1	2	1/7	0,1429
2	2	1/7	0,1429
3	1	1/7 + 1/21	0,19
4	1	1/7 + 1/21	0,19
5	2	1/7	0,1429
Shortlisting score	1	1/7 + 1/21	0,19
	TOTAL	6/7 + 3/ 21	1

Example 2:

A committee is formed by 5 experts: $e \in \{1,2,3,4,5\}$.
No expert has declared Level 1 of expertise for a certain candidate *c* (hence $k = 1$). As a result, the weights are distributed as follows:

Expert e	Level of Expertise	$\boldsymbol{weight_{c,e}}$	$\boldsymbol{weight_{c,e}}$(num)
1	2	1/7	0,1429
2	2	1/7	0,1429

3	2	1/7	0,1429
4	2	1/7	0,1429
5	2	1/7	0,1429
Shortlisting score	1	1/7 + 1/7	0,2857
	TOTAL	6/7 + 1/7	1

2.4 DISCREPANCIES

The evaluation system detects significant discrepancies among experts' standardized scores for the same application. To detect discrepancies, for each candidate we compute the difference between the highest and the lowest score among all those obtained from the different evaluators. That is

$$diff(c) = max_e\big(score_{norm}(c,e)\big) - min_e\big(score_{norm}(c,e)\big)$$

where:

- $score_{norm}(c,e)$ is the normalized score given to candidate c by evaluator e.
- max_e and min_e represent the maximum and minimum scores given by any evaluator for candidate e.

If this difference is equal or larger than 2, the scores of this candidate are considered discrepant.

The scores coming from the shortlisting stage are not considered in this part of the procedure.

Discussion on discrepancies

The system will only highlight significant discrepancies among candidates who are at the bottom of the list with fellowship or at the top of the list without fellowship. The exact number of positions considered will be proportional to the number of fellowships to be awarded by the committee.

Only the assessment of this restricted group of candidates with significant discrepancies will be discussed by the committee. After the discussion, evaluators may either maintain or change their original scores. The revised scores will be considered definitive.

2.5 COMPUTATION OF THE FINAL SCORE

The final normalised score of each candidate is computed adding for the first time the *m* existing scores - one from each expert and one from the shortlisting stage -, all normalized and weighted according to the evaluator's level of expertise. In other words,

$$score_{norm}(c) = \sum score_{norm}(c,e) \times weight(c,e)$$

This final score range **is the one being used to rank the candidates**. This ranking will not be affected by any of the subsequent steps.

Rescale of final score

With the goal of presenting the candidate's scores in a range from 1 to 8, normalised scores are rescaled. The following procedure will be followed: all final norm alised scores are first rescaled to obtain a temporary score from 0 to 1

$$temp.score(c) = \frac{score_{norm}(c) - min_{committee}\left(score_{norm}(c)\right)}{max_{committee}\left(score_{norm}(c)\right) - min_{committee}\left(score_{norm}(c)\right)}$$

That is, to the candidate's final normalised score, we subtract the minimum normalised score among all candidates and divide by the difference between the maximum and the minimum normalised score, again among all candidates. Every score is now between 0 and 1.

Now, a second rescaling calculation is made to obtain the **rescaled final score of each candidate**:

$$score_{resc}(c) = temp.score(c) \times \left(max_{committee}\left(score(c)\right) - min_{committee}\left(score(c)\right)\right) + min_{committee}\left(score(c)\right)$$

Therefore, all scores are translated to the interval $\left[min_{committee}\left(score(c)\right), max_{committee}\left(score(c)\right)\right]$.

The scores from the shortlisting stage are being considered in this part of the procedure.

2.6 SINGLE RESERVE LIST

The reserve list of each committee will be formed by the candidates who did not obtain a fellowship.

If the subcommittees are formed[8], once the candidates to be awarded a fellowship have been determined in each subcommittee, the remaining ones will be joined in a **unique reserve list,** shared across to all subcommittees within the same committee. This list will be ordered based on each candidate's definitive score $score_{resc}(c)$. Then, a **new normalization is applied with the aim of comparing scores from different subcommittees**. Specifically, if *N* subcommittees were created, this second normalization and final reserve list are conducted as follows:

1. For each subcommittee *S*, let us say with *n* candidates in total, we normalize the definitive scores $score_{resc}(c)$by calculating first their average

$$mean_S = \frac{1}{n}\sum_{c=1}^{n} score_{resc}(c)$$

8. For more information of the composition of the committees see Section 3.1 Structure of the Committees of the Selection Process Guidelines.

and then their standard deviation

$$stdev_S = \sqrt{\frac{\sum_{c=1}^{n}(score_{resc}(c) - mean_s)^2}{n-1}}$$

and finally computing the normalized score

$$score_{resc,norm}(c) = \frac{score_{resc}(c) - mean_S}{stdev_S}$$

2. A unique reserve list is created by joining all candidates who were not awarded a fellowship and ordering them by the new normalized score of step 1, $score_{resc,norm}(c)$.

In case of a withdrawal, the fellowship will be awarded to the best ranked candidate in the reserve list. In case of tie between reserve candidates, this will be resolved based on the shortlisting score.

2.7 FEEDBACK TO CANDIDATES

To provide adequate feedback to candidates on their assessment in the interview stage, they will receive their normalized total score, and their position in the quartile in which they are ranked for each criterion compared to the other candidates of their committee or subcommittee.

» Total normalized score

Candidates will be provided with their total normalized score rescaled which has been previously calculated in the section *2.5. Computation of the final score - Rescale of final score.*

» Quartile distribution

The candidate will be informed of the quartile assigned for each of the selection criterion assessed. To determine this position, the rankings of each committee are divided into four equal parts or quartiles Q_1, Q_2, Q_3 and Q_4, where Q_1 corresponds to the top group of the $n/4$ highest normalised scores, and Q_4 to the bottom group with the $n/4$ lowest ones.

In case several subcommittees had been created, the quartiles will be computed separately in each subcommittee.

To calculate the rankings for each committee and for each selection criterion, scores must be normalised in the same way as in section *2.2, Normalization*, but for each selection criterion and without considering the shortlisting score. Once all scores for each evaluator and each criterion are normalised (all the$score_{norm}(c, e, crit)$), the final normalised score for each criterion is:

$$score_{norm}(c, crit) = \sum_{e=1}^{m-1} score_{norm}\,(c, e, crit) \cdot weight(c, e)$$

We recall that, at this point, the number of evaluators is *m-1*, since shortlisting score is not being considered. Therefore, the weights of each candidate-evaluator pair must be adjusted taking into account this condition (there will be one Level 1 expertise less).

Annex 1.1: Effects of normalization of scores

The goal of this annex is to elaborate on detail the process of normalization (or standardization) which will be applied to the scores given by the experts in the selection stages, as well as the effects of this action.

The objective of normalizing each expert's scores before adding them to the others' and comparing between them is to ensure that each evaluator's score carries similar weight in candidate's final score, mitigating the differences in scale and dispersion that might exist among evaluators.

The experts' scores in each of the evaluation criteria can take values in between 1 and 8, and so does the weighted average of these grades computed for every candidate and which we denote by $score_{c,e}$ (where *c* is the candidate and *e* the evaluator).

To normalize the scores of the expert *e*, the average ($mean_e$) and the standard deviation ($stdev_e$) of all their scores are calculated (see section *2.2 Normalization* of the Annex 1 for more details). The original $stdev_e$ is adjusted, setting a lower bound that allows controlling the dispersion of the scores of the evaluators with $stdev_e$<<1, bringing them closer to the mean of the normalised scores (0), and adding justice to the evaluation process.

With these two quantities a new score for each candidate is obtained by

$$score_{c,e}^{norm} = \frac{score_{c,e} - mean_e}{stdev_e}$$

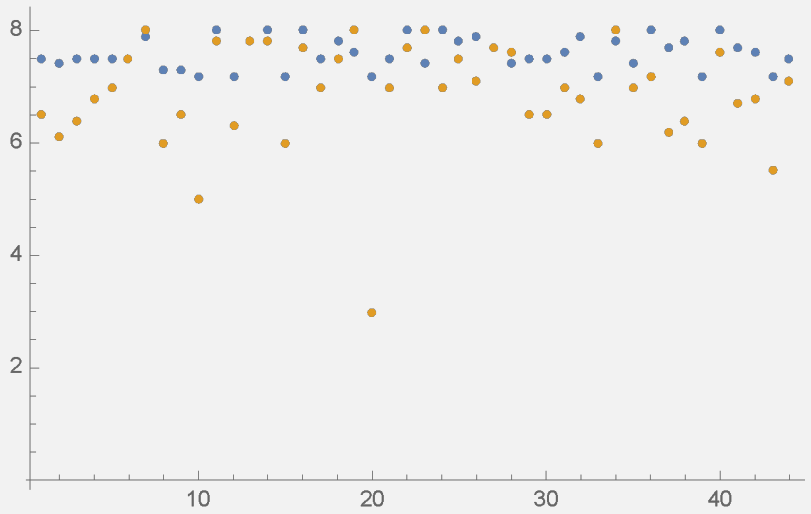
This new score is the one that will be used (after being weighted by the level of expertise of the evaluator for the given candidate) to compute the average score of all the experts' scores for the given candidate.

The performed normalization has the following effects:

- The mean of the scores of each expert is equal to 0, which neutralizes any potential tendencies of the different experts to "grade high" or "grade low".
- The standard deviation of the scores of each expert is equal to 1. This means that, in average, the distance (squared) to the new mean (0) is equal to 1. Approximately 95% of the new scores of each evaluator are between -2 and 2. Scores that were given within a very narrow range ($stdev \ll 1$) will now be more dispersed, while marks given in a large range ($stdev \gg 1$ will now become closer to the mean.
- When the number of applications assigned to an evaluator is very small, there is a risk that, if their standard deviation <<1, the normalization process may produce scores that that deviate significantly from the original scores. However, establishing a bound on the standard deviation helps mitigate this effect.
- Outliers will remain outliers and, in some cases, they may become even more pronounced. If a score was significantly further from the average than the others, the new grade will reflect that difference. Additionally, if the standard deviation of the scores from a specific evaluator is small, this effect may be amplified. However, establishing a lower bound on the standard deviation helps to reduce the impact of outliers.

Illustrative example

Suppose there are 44 candidates and, for the this example, two evaluators. The graph displays the scores from one expert (e=1) in blue, while the scores from the second expert (e=2) are shown in orange. The orange scores are more scattered than the blue scores and include a clear outlier with a grade of 3.

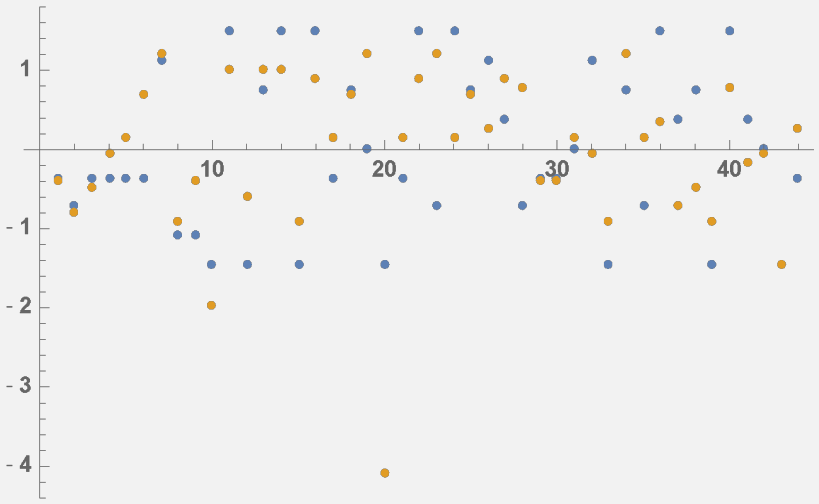


The values computed for this set of scores are:

$$mean_1 = 7{,}6; stdev_1 = 0{,}27;$$

$$mean_2 = 6{,}85; stdev_2 = 0{,}941.$$

The next figure shows the distribution of the new scores after normalization:



The yellow scores display a distribution around the mean that is similar to their original distribution, as their standard deviation was close to 1. The outlier remains present. In contrast, the blue scores are now more scattered than before -even more so than the yellow scores- due to their originally low standard deviation, which caused them to be tightly distributed around their mean.