

Selection Process- Step 2

Interviews by multidisciplinary committees

PROCEDURES

Introduction

This document explains the procedure for evaluating and selecting candidates in the face-to-face stage by assessment committees.

The objective of this system is to ensure maximum efficiency and objectivity in the assessment of candidates, eliminating, as much as possible, the effect of bias in scorings and minimizing the possibility that an expert's assessment is determined by not strictly technical factors.

In this regard, the procedure is made up of the following steps:

1. *Grading*. Collecting the candidates' scores given by each of the experts and weighted according to the evaluation criteria explained in the document "2. *Evaluation Criteria*"
2. *Normalization or standardization*. Transforming the scores to mitigate the differences in scale and dispersion between different experts.
3. *Detection of discrepancies*. Reviewing the candidates' scores for whom there are significant differences between expert assessments. In case of significant discrepancies, the affected experts have the option of reconsidering their grades.
4. *Weighing*. Adding the (normalized) scores from all evaluators, together with the grade obtained from the preselection procedure, each of them weighted according to the expert's level of expertise with the discipline of the candidate. This level will be declared by the experts previously to the face-to-face interviews.
5. *Presentation*. Rescaling of the scores to present them, after the procedure above, following a scale from 1 to 8.
6. *Reserve list*. Creating a unique reserve list for each area.
7. *Feedback for the candidates*. Separating the candidates, for each evaluation criterion, into four quartiles according to their scores. Inform the candidates of the quartile to which belong.

1. Grading

Experts shall rate the various aspects of a candidate's application during the face-to-face assessment sessions.

The representative of "la Caixa" Foundation will be in charge of entering all experts' scores for each candidate into a database. The final marks generated in the remote assessment (preselection) process will be included. From this moment on, **these marks will be considered as if they came from one additional expert in the committee, who is assumed to possess level of expertise 1** (see step 4). The system will consider the various levels of weighting of each aspect evaluated and arrive at an initial score for each candidate from each expert

$mark_{c,e}$ = mark given to candidate c by expert e .

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

Consequently, every candidate has m marks: one from each expert, plus the one coming from the preselection phase or remote assessment. These marks take values from 1 to 8.

For later use, at this stage we calculate the quantity

$$mark_c = \frac{1}{m-1} \sum_{e=1}^{m-1} mark_{c,e},$$

which is the average of all experts' marks obtained by candidate c , excluding the preselection score. This number will not be used until step 5 (Presentation).

2. Standardization

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

- The expert's mean score is calculated

$$mean_e = \frac{1}{n} \sum_{c=1}^n mark_{c,e}$$

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

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

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

$$\text{mark}_{c,e}^{\text{norm}} = \frac{\text{mark}_{c,e} - \text{mean}_e}{\text{stdev}_e}$$

Observations

With this procedure, the original marks

$$\text{mark}_{c,e} \in [1,8]$$

are converted in new quantities

$$\text{mark}_{c,e}^{\text{norm}} \in (-\infty, \infty).$$

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 marks will be higher or lower, depending on how far they are from the average of the original grades, and how frequent this distance is. (See the annex for further details about the effects of normalization).

3. Detection of discrepancies and an alarms

In the event that the standardized scores given by the different experts to the same candidate show a significant divergence (see below), the system displays an alarm.

To be considered as a significant divergence, the difference between the maximum and the minimum score among all those of the same candidate must be larger than 2. More precisely,

$$\max_e(\text{mark}_{c,e}^{\text{norm}}) - \min_e(\text{mark}_{c,e}^{\text{norm}}) > 2.$$

The scores coming from the preselection process are not taken in this part of the procedure.

The system will only indicate the alarms of those candidates in the last positions with a scholarship or the first positions without fellowship. The exact number of positions to be considered will be proportional to the number of fellowships to be granted by the committee.

The candidates of this restricted group with significant divergences will be discussed by the committee at its final meeting and evaluated again by each expert, who can then maintain or change their original score. The new scores will be entered into the system again and they will be final.

4. Weighting according to expertise with the discipline assessed

The experts of each committee may indicate, via the online application that gives them access to applications, their level of expertise with the discipline of the candidacy evaluated

Each expert can choose between two possible levels of expertise:

LEVEL 1: Their knowledge corresponds, generically, with the field of the discipline evaluated and their evaluation can therefore be considered that of an expert. The remote assessment mark is always considered that of an expert.

LEVEL 2: Their knowledge does not correspond, generically, with the field of the discipline evaluated; therefore, their evaluation cannot be considered strictly that of an expert, but it is sufficient taking into account the characteristics of the call.

The levels of expertise of the different evaluators for a given candidate will result in different weights of the expert's score at the time of adding all the scores of this candidate. Supposing we have m evaluators (we are including here the score of the remote assessment), 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, among which we always find the remote assessment score.

Hence, if k evaluators (including the preselection score) declare level 1 of expertise with the candidate c , the expert e for the candidate c will have a weight of

$$weight_{c,e} = \frac{1}{m+1}, \text{ if the expert } e \text{ has declared level 2, and}$$

$$weight_{c,e} = \frac{1}{m+1} + \frac{1}{k(m+1)}, \text{ if the expert } e \text{ has declared level 1.}$$

Example 1:

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

The experts 3 and 5 have declared a 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	$weight_{c,e}$	$weight_{c,e}$ (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
preselection	1	$1/7 + 1/21$	0,19
TOTAL		$6/7 + 3/21$	1

Example 2:

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

No expert has declared a 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	$weight_{c,e}$	$weight_{c,e}(\text{num})$
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
preselection	1	$1/7 + 1/7$	0,2857
	TOTAL	$6/7 + 1/7$	1

Computation of the final grade

Once the weights of each expert for every candidate have been calculated, the final score of each candidate will be computed adding for the first time the m existing scores (one from each expert and one from the remote assessment), all normalized and weighted according to the expert's level of expertise. In other words,

$$\text{final mark}_c = \sum_{e=1}^m \text{mark}_{c,e}^{\text{norm}} \times \text{weight}_{c,e}.$$

This final score can take any value in $(-\infty, \infty)$, and it is the one being used to rank the candidates. This ranking will not be affected by any of the later steps, which serve only aesthetic purposes.

5. Presentation of the scores

With the goal of presenting the candidate's marks in a range from 1 to 8, the following procedure will be followed:

1. All final marks are first rescaled to obtain a temporary score from 0 to

$$\text{temp mark}_c = \frac{\text{final mark}_c - \min_c(\text{final mark}_c)}{\max_c(\text{final mark}_c) - \min_c(\text{final mark}_c)}.$$

That is, to the candidate's final score, we subtract the minimum mark among all candidates and divide by the difference between the maximum and the minimum score,

again among all candidates. Every mark is now between 0 and 1 but the ordering remains the same it was.

2. All scores are translated to the interval

$$\left[\min_c(\text{mark}_c), \max_c(\text{mark}_c) \right]$$

where we recall that mark_c is the direct average of the expert's scores for the candidate c (see step 1), before normalizing or applying the weights according to expertise, and without taking into account the preselection results.

After this rescaling we obtain the definitive final score of each candidate:

$$\text{mark}_c^{\text{def}} = \text{temp mark}_c \times \left(\max_c(\text{mark}_c) - \min_c(\text{mark}_c) \right) + \min_c(\text{mark}_c)$$

6. Reserve lists

The reserve list of each committee will be formed by the candidates who did not obtain a fellowship, in the order established in step 4 (or 5, since it is the same one).

In the event that the number of candidates to be interviewed is too high to be handled by only one committee, two or more subcommittees will be created in the same area. Candidates will be distributed between among the different subcommittees in a sequential manner according to their score. Therefore, the candidate with the highest score will be assigned to subcommittee 1, the next one to subcommittee 2, and so on until starting over, and until exhausting the candidate list.

In this situation, every subcommittee will be preassigned a number of fellowships to award and this will be done accordingly to the procedure described in the sections above.

Once the candidates to be awarded have been determined in each subcommittee, the remaining ones will be joined in a **unique reserve list**, common to all subcommittees (of the same committee). This list will be ordered according to the definitive score of the candidates $\text{mark}_c^{\text{def}}$, after a **new normalization has been applied**. More precisely, if N subcommittees have been created, the second normalization and final reserve list are done as follows:

1. For every subcommittee S , let us say with n candidates (in total), we normalize the definitive scores $\text{mark}_c^{\text{def}}$ by calculating first their average

$$\text{mean}_S = \frac{1}{n} \sum_{c=1}^n \text{mark}_c^{\text{def}},$$

and then their standard deviation

$$stdev_s = \sqrt{\frac{\sum_{c=1}^n (mark_c^{def} - mean_s)^2}{n-1}}$$

and finally computing the normalized grade

$$mark_c^{def,norm} = \frac{mark_c^{def} - mean_s}{stdev_s}$$

Ordering by this new mark, does not change the previous ranking of the subcommittee. Hence fellowships will be awarded to the same candidates, but a comparison with the other subcommittee's lists will be meaningful.

2. A unique reserve list is created by joining all candidates with no fellowship and ordering by the new normalized score of step 1, $mark_c^{def,norm}$.

In case a resignation occurs, 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 preselection score.

7. Feedback for the candidates

In order to provide adequate feedback to the candidates on the assessment of their candidacy during the face-to-face assessment process, their qualifications for each criterion will be analyzed for comparison with the qualifications of the other candidates of their panel.

This feedback will not take into account the remote assessment mark, since interviewed candidates will have already obtained the corresponding feedback from the remote assessment process.

Candidates will not be informed about the exact marked obtained in each of the criteria, but only about the quartile to which their grade belongs, once again normalized.

Normalization (for each criteria)

A new normalization will take place although in this occasion it will be done separately for each of the criteria used for the candidate's evaluation (for example *potential, background, ...*). Indeed, given a criterium $k \in \{1,2,3\}$, and an expert $e \in \{1, \dots, m\}$, the normalization will be carried out by first calculating the average and the standard deviation in the set of n candidates

$$mean_{e,k} = \frac{1}{n} \sum_{c=1}^n mark_{c,e,k}$$

$$stdev_{e,k} = \sqrt{\frac{\sum_{c=1}^n (mark_{c,e,k} - mean_{e,k})^2}{n-1}}$$

where $mark_{c,e,k}$ denotes the score of the candidate c obtained from the expert e for the criterium k .

With this two quantities, we normalize each of the score lists by calculating

$$mark_{c,e,k}^{norm} = \frac{mark_{c,e,k} - mean_{e,k}}{stdev_{e,k}}$$

Finally, the normalized score of each candidate c with respect to the criterium k is the average of all the normalized scores obtained by each of the evaluators, that is

$$mark_{c,k}^{norm} = \frac{1}{m} \sum_{e=1}^m mark_{c,e,k}^{norm}.$$

Every candidate obtains in this way three normalized scores, one for each of the criteria considered.

Quartile division

For every criterium k the list of candidates will be ordered according to the normalized scores $mark_{c,k}^{norm}$, and divided into four equal parts or quartiles Q_1, Q_2, Q_3 and Q_4 , where Q_1 corresponds to the top grup of the $n/4$ highest scores, and Q_4 to the bottom group with the $n/4$ lowest ones.

The candidate will be informed of the quartile assigned for each of the criteria.

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

ANNEX

About normalization of scores and its effects

The goal of this annex is to comment on the process of normalization (or standardization) which will be applied to the scores given by the experts in the face-to-face assessment phase, as well as the effects of this action.

The objective of normalizing each expert's scores before adding them to the others' or comparing between them is to ensure that the grade of every evaluator has a similar weight in the final candidate's mark, mitigating the differences in scale and dispersion that might exist between them.

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 $mark_{c,e}$ (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 of his or her scores are calculated (see Section 2 of the document for more details). With these two quantities a new score for each candidate is obtained by

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

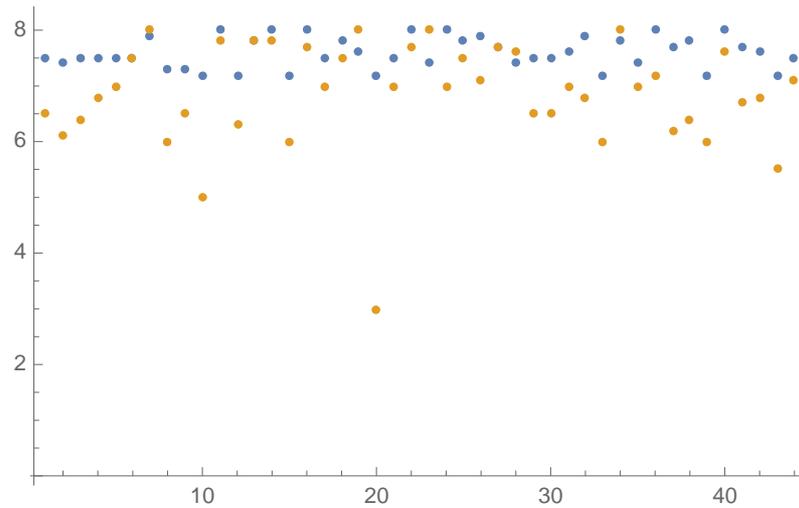
This new score takes values in $(-\infty, \infty)$, and is the one that will be used (after being weighted by the level of expertise of the evaluator for the given candidate) to compute the average grade of all the experts' scores for the given candidate.

The performed normalization has the following effects:

- The mean of the scores of each expert is equal to 0. This cancels the possible (natural) tendencies to "grade high" or "grade low" of the different experts.
- 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 in 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.
- Outliers will still be outliers (and sometimes even more than before). If a grade was much further from the average than the others, the new grade will have the same property. If the deviation of the list was small, this can even be enhanced.

Example

Suppose there are 44 candidates and, for the purpose of this example, two evaluators. The graph below shows in blue the scores of one of the experts ($e=1$) and in orange the scores of the second one ($e=2$), quite more scattered than the blue ones and with a clear outlier, with a grade of 3.

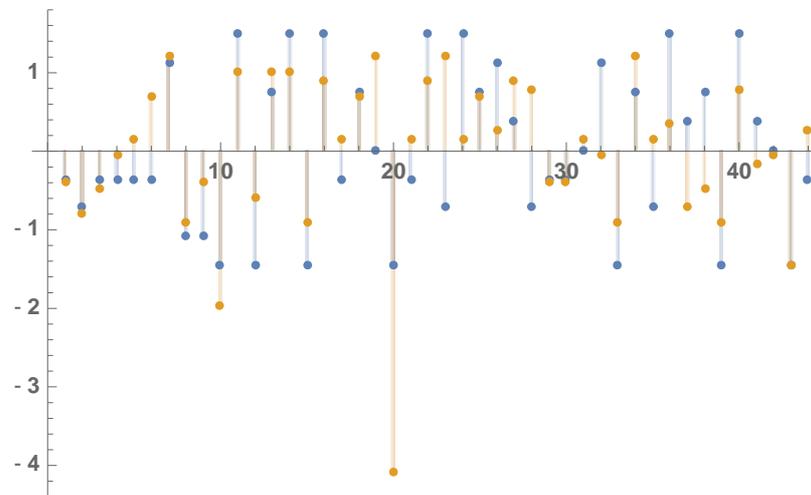


The values computed for this set of scores are:

$$mean_1 = 7,6; \quad stdev_1 = 0,27;$$

$$mean_2 = 6,85; \quad stdev_2 = 0,941.$$

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



The yellow grades show a similar distribution around the mean to the one they had before (their standard deviation was close to 1), and the outlier is still there. Instead, the blue scores are now more scattered than before (even more than the yellow ones!), because their standard deviation was originally very small (they distributed tightly around their mean).

Postdoctoral Junior Leader Fellowships

Rating criteria

GUIDELINES AND CRITERIA FOR THE EVALUATION OF APPLICATIONS DURING THE INTERVIEW SELECTION PROCESSES FOR THE POSTDOCTORAL JUNIOR LEADER FELLOWSHIPS PROGRAMME

General considerations

Several committees will be formed according to the number of applicants called in for interviews. Each of these committees will be made up of university professors or experts in the evaluated disciplines.

To rate candidates during the interview selection process, evaluators will use a rating scale with three aspects for evaluation, each of which will have a certain weight.

These fellowships are restricted to areas of Health Sciences, Technologies, Engineering, Mathematics and Physics. All applications have been previously reviewed to verify that they meet the formal requirements of the call but their scientific content has not been reviewed. We kindly ask you that in case you detect any application that, in your opinion, does not meet this requirement, please inform us as soon as possible to determine its eligibility.

Score

Each evaluated aspect must be rated using one of the values in the scale below:

Rating	Score
Exceptional	8
Excellent	7
Very good	6
Good	5
Average	4
Mediocre	3
Poor	2

Very poor

1

Aspects evaluated

1. CANDIDATE POTENTIAL

Summary: *the candidate's potential, paying particular attention to the candidate's interpersonal skills, such as clarity and consistency of discourse, expression of ideas, ability to present complex reasoning, teamwork, capacity for independent reasoning, originality, entrepreneurship and leadership will be evaluated.*

In this section, the following will be mainly assessed:

- Originality: capacity to go off the beaten path, coming up with creative proposals or delving into relatively unexplored fields. Originality of both the proposal and formula to achieve the objectives pursued.
- Innovation: capacity to create new knowledge or, in case the proposal includes possible outcomes transferable to industry, to break new ground or find new formulas to create wealth. Use of new technologies or innovative use of already existing technologies. Use of new theoretical approaches to existing phenomena or problems or innovative use of already existing theoretical approaches.
- Feasibility: the candidate puts forth ambitious and realistic ideas and their capacities are in keeping with the scope of their proposal.
- Clarity of exposition: capacity to present complex reasoning and very specific matters clearly and precisely. Appropriate wording including technical terms but avoiding use of overly specialised vocabulary.
- Independence and Leadership: aspects of their track record which make it possible to gauge the candidate's capacity to head a research project will be assessed. In this regard, the candidate's scientific output (publications as primary author, relevant contributions, among others) will be considered. The direction of research work, students' supervision and direction, recognitions (awards, invitations to international lectures) and capacity to obtain resources for their own research will also be assessed.

The weight of this section will be 30%.

2. MOTIVATION AND IMPACT OF THE PROPOSAL

Summary: *in this section, the conceptual and methodological novelty of the submitted proposal, as well as its impact, understood in its broadest sense—capacity of the submitted project to contribute towards the transformation and improvement of fields such as economy and cre-*

ation of wealth, society, culture, science, citizens' quality of life, the environment or public policies—will be evaluated.

In this section, the following will be assessed:

- The submitted proposal must innovative and original, both conceptually and methodologically speaking. The proposals that involve risk and creativity will be looked upon favourably.
- The candidate must justify the interest and feasibility of the research project that they wish to carry out, as well as the suitability of the centre or centres where they propose to do so.
- The interests put forth by the candidate must be consistent and well structured, and the proposed project must be kept within a path with a broader scope. In this regard, the candidate must state—and the evaluator must rate—to what extent carrying out the proposed research project for which they are applying for the fellowship is a necessary step in the right direction.
- Determination to complete the project within the established period of time for the duration of the fellowship or justification for the duration of the project in the case of longer periods.
- The social return on investment—understood in its broadest sense—must also be evaluated: advancement of science and knowledge, creation of wealth and the possibility of transfer to third parties. Likewise, the estimated scientific, social or economic opportunity cost, should the candidate not be able to carry out the proposed project.
- They must be able to demonstrate sound knowledge of the risks involved in the proposed research, as well as identify prevention and mitigation measures for them.
- In this aspect, the potential impact of the fellowship on the candidate's future career will also be evaluated.

The weight of this section will be 20%.

3. ACADEMIC AND PROFESSIONAL CAREER

Summary: *the experts will assess the contributions made in the field chosen for the research project presented, as well as the coherence between the candidate's academic education and track record.*

In this section, the following will be assessed:

- Professional and academic experience, scientific findings and output, as well as prior training to carry out their proposal.
- Excellence, relevance and recognition for their contributions to the scientific discipline of their line of research, such as project management, doctoral theses, lectures at congresses, awards and other merits, as well as securing funding to carry out their research.
- Quality and the depth and breadth of their track record in relation to the candidate's possibilities. Thus, particular attention will be given to the years that have passed since they earned their doctoral degree, evaluating both their past achievements and possible future opportunities according to their potential. In this regard, younger candidates or who have made an interruption of their research career due to justified reasons cannot be penalised for having a relatively short track record.
- Stable and well-planned paths throughout their entire career. Should there be changes in their career path, these must be duly justified and supported in a coherent and reasoned manner. In case they have professional experience, its relevance in relation to the proposed project must be explained, if exists.

The weight of this section will be 50%.

Final score during shortlisting

The rating given to candidates during the shortlisting process will also be considered during the interview stage. The standardised rating given during the shortlisting process will be added—as an expert rating—to the rest of the ratings given by the evaluators during the interview to each application.

Committee members will also have access to the ratings and evaluations given to each application by the evaluators who were involved in the shortlisting process. Likewise, they will have additional information regarding the application's position within its group during the shortlisting stage, the number of applications evaluated in this group and the number of shortlisted applications, as well as any other data that the Fellowship Programme Office may consider to be relevant for the evaluation of applications.

To the extent that they deem appropriate, committee members may consider such information when evaluating and rating interviewed candidates.

Interview content

Face-to-face interviews make it possible to detect—based on more subjective, fine and subtle considerations—the quality and consistency of the candidate being evaluated.

During the interview, the candidate's theoretical knowledge may be probed, although this is not the main objective. First and foremost, interviews are used to judge the merit of the application, considering the above-mentioned aspects.

The face-to-face interview seeks to:

- Delve into the information provided in the application, particularly with regard to the applicant's academic, scientific or professional interests.
- Ask about matters which were not included in the application and which the committee considers particularly relevant to evaluate the suitability of the applicant to carry out the proposed research project.
- Evaluate the applicant's scientific and professional potential.
- Evaluate their all-round training, interests, concerns, and curiosity for the social, scientific, economic, cultural, or artistic context—although not directly related to the research project.
- Evaluate their personal and scientific maturity, their motivation to complete the proposed project and their capacity to clearly express their ideas and firmly defend them.

Formal aspects

- Each interview will last approximately 30 minutes.
- The evaluators comprising the committee will not introduce themselves to the candidate.
- A representative of "la Caixa" Foundation will begin the interview which, in all cases, will start off by giving the floor to the applicant to briefly summarize their research proposal in 10 minutes. Then, the other committee members will ask the questions that they deem relevant in order to properly assess the application.
- Interviewers should not ask questions about topics already discussed in the application, unless they are to clarify some aspects.
- Language: interviews will be conducted entirely in English.
- Committees must try to observe the established schedules and be as punctual as possible with the candidates called in for interviews.

There are no established protocols with regard to use of the formal form of address or the need to shake hands before or after the interview. These are aspects that are left to the discretion of the committee or the spontaneity of the candidates. Nevertheless, it goes without saying that the interviews must be characterized by their formality, propriety and the relevance of the questions.