ESF Report on the Evaluation Process of the Foundation for Science and Technology (FCT)

R&D Units

- *END OF PROJECT REPORT-*

2013-2014

Ms. Valérie Allspach-Kiechel, Dr. Janina Kancewicz-Hoffman, M. Nicolas Walter, Dr. Olivier White, Dr. Jean-Claude Worms.

24 September 2015

Contents

Introduction

The Foundation for Science and Technology (FCT) has played an important role in the implementation of research policy in Portugal since 1997. One of the key instruments of this policy is multiannual, institutional funding for so-called Research and Development Units (R&D Units, including the Associated Laboratories). R&D Units are constituted by groups of researchers from several research organisations (mainly universities) around a specific research field or theme and are affiliated to a university or another research institution. The level of their core funding (besides other funding instruments such as research projects or fellowships) for the following period is based on the results of an evaluation and therefore they are regularly evaluated by FCT, normally every five to six years. This report presents the 2013-2014 R&D Units evaluation which was in part implemented by the European Science Foundation (ESF).

The report was prepared by ESF based on:

- FCT and ESF documentation produced throughout the process;
- Reports prepared by ESF observers after site visits;
- Reports produced at the end of the process by Review Panels under the coordination of the panel chairs;
- Various consultations and exchanges with Review Panel members.

Background and timeline

The previous R&D Units evaluation took place in 2007 and was structured in a different manner as it was then organised by FCT. Associated Laboratories (AL) were instrumental in implementing the assessment process with the support of international experts. In the 2013-2014 evaluation, all R&D Units (including AL) were evaluated on the same basis and the ESF was in charge of implementing key elements of the process, in particular identifying and appointing expert evaluators and running the first phase of the evaluation.

Over the past decade, ESF has collected and compiled key information about peer review and evaluation through its Member Organisations' fora:

- *Evaluation in Research and Research Funding Organisations: European Practices*[1]
- *Evaluation in National Research Funding Agencies: approaches, experiences and case studies*[2]
- *European Peer Review Guide – Integrating Policies and Practices into Coherent Procedures*[3]

1 *http://www.esf.org/fileadmin/Public_documents/Publications/mof_evaluation.pdf*

2 *http://www.esf.org/fileadmin/Public_documents/Publications/moforum_evaluation.pdf*

3 *http://www.esf.org/fileadmin/Public_documents/Publications/European_Peer_Review_Guide_01.pdf*

This knowledge allowed ESF to obtain a comprehensive overview of European standards and best practice. Discussions to involve ESF in the evaluation process started in September/October 2013, during the period R&D Units registered through the FCT portal, and before the deadline for submission of applications on 30 December 2013. The involvement of ESF in the evaluation process was announced by FCT on 29 April 2014 with the document '*Evaluation Guide - Additional Information'* (for detailed schedule of the evaluation see Table 1).

When discussing the involvement of ESF in the evaluation, opportunities to improve the level of collegiality and interactions with the R&D Units in the process were identified. The evaluation procedure agreed following discussions between ESF and FCT and eventually implemented was adapted in order to take these opportunities into account. It, however, respected the boundary conditions defined in the *Regulation for Evaluation and Funding of Research Units and Evaluation Guide* published by FCT in July 2013.

This assessment exercise applied a peer review process very similar to the one implemented in the framework of competitive research project funding. ESF has a long record in implementing robust research project proposal peer review processes not only for its own instruments, but also for third parties, including ERA-Nets (EUROPOLAR, HERA), large European programmes (GRAPHENE Flagship, European Space Agency microgravity programme), national funding programmes (Italian Ministry for Research/University of Bologna FIRB and PRIN pre-selection, Irish IRCSET), private foundations and philanthropic organisations (AXA Research Fund, Campania di San Paolo in Italy). **Over the past three years, the ESF has performed the assessment of over 1,700 research proposals following various evaluation models**.

The FCT evaluation focused on research units, including their past achievements and future development plans (strategic programmes). The ESF experience in grant proposal evaluation was coupled with its experience in evaluating research performing and funding organisations (Institutes of the Bulgarian Academy of Sciences, 2009; Slovenian Research Agency, 2010; European Space Agency Microgravity research programme, 2000-2012; Research Council of Lithuania, 2013; Hungarian Scientific Research Fund - OTKA, 2014). The specificity and the novelty of the 2013-2014 FCT evaluation of R&D Units was its orientation more toward future strategy, plans and corresponding capacities than to past achievements. This element was kept at the centre of the evaluation process.

The contract clearly limited ESF's role and participation in the evaluation process to the scientific assessment aspects. While ESF was strongly involved in the process leading to the independent rating of the R&D Units and in particular the implementation of the first phase of the process, it was not involved in the resulting allocation of funding. ESF also provided an interactive IT platform to support the evaluation process.

The adaptation of the evaluation process following ESF involvement is described below and detailed in Annex 1 of this document.

Table 1. Main steps of the Evaluation.

Date	Milestone
2013	
28 February – 13 March	Public consultation on the proposed regulations for the 2013 Evaluation Exercise of R&D Units
05 July	Publication of the *Notice of the Call for the 2013 Evaluation Exercise of R&D Units*
05 July	Publication of the *Regulation for Evaluation and Funding of Research Units*
09 July – 31 October	R&D units registered for evaluation
31 July	Publication of the *Evaluation Guide, FCT Evaluation of R&D Units 2013*
September – October	Negotiations start with ESF
30 December	Deadline for submission of applications
2014	
February to mid-April	Collection of external reviews by ESF
29 April	Announcement of ESF's role with the updated *Evaluation Guide - Additional information*
02-15 May	Rebuttal, by applicants, of the individual review reports (two by external referees and secondary rapporteur report)
26-29 May	Stage 1 Review Panel meetings, Amsterdam
27 June	Official announcement of Stage 1 results
July	Appeals process Stage 1
02 October	Official announcement of Stage 1 results after appeals
June- October	Stage 2 site visits
24-26 November	Stage 2 Review Panel meetings - Final Review Panels consensus meeting, Lisbon
22 December	Official announcement of Stage 2 results
2015	
January	Appeals process
25 May	Overall results after Stage 2 evaluation and appeals

The Evaluation Structure

Defining the Approach

According to the procedure set out by FCT in its initial document, "*All applications will be subjected to scientific evaluation distributed by four panels, which are responsible for the preliminary remote reviewing of all applications. This distribution is in accordance with the four major scientific domains under the aegis of the Scientific Councils of FCT. The constitution of the evaluation panels will take into consideration the number of applications for each scientific domain, a good gender balance as well as a fair geographic and institutional distribution of evaluators. (…) The members of each panel will, in turn, be distributed by several workgroups of four elements each. Every workgroup will be responsible for the remote assessment of about 10 proposals in a given scientific area. One member of each workgroup will be designated the coordinator of the workgroup. Optimally the workgroup coordinator will also be a member of one of the evaluation panels of Stage 2 of the evaluation.*"[4]

Based on its experience in evaluation processes, ESF proposed improving the structure and process while keeping in line with the above conditions. The improvements aimed at:

(1) Increasing collegiality of decision at both stages. To this end, two Review Panel consensus meetings were organised, one at each evaluation stage, and each application was discussed by the whole panel instead of by smaller compartmented workgroups of four panel members discussing and each evaluating remotely approx. 10 applications without further integration at a higher disciplinary level. **This ensured calibration of evaluations across a wide research domain.**

(2) Strengthening interactions with the R&D Units. First, at Stage 1, a rebuttal phase was introduced before the first consensus meeting. This allowed corrections of any misunderstandings or errors in individual reviews before the panel meetings and the shortlisting of units for Stage 2. Second, panels were strongly encouraged to list in their Stage 1 consensus reports specific issues and questions to be addressed during Stage 2 site visits. **Both measures created a dialogue between the units and the Review Panels, increasing accuracy and fairness of panels' final judgement.**

Moreover, the high number of applications (322 eligible applications of which 44 were interdisciplinary) eventually submitted and their uneven distribution between research domains required adaptations of the panel structure. To respond to the very high number of applications in those domains, it was agreed to split the *'Humanities and Social Sciences*' panel into two panels, *'Humanities'* and *'Social Sciences*'. '*Exact Sciences and Engineering'* followed the same approach and was split into '*Exact Sciences'* and *'Engineering'*. **This brought the number of disciplinary panels to six**.

To ensure interdisciplinary evaluation of R&D Units that had explicitly indicated an interdisciplinary profile in the application form, ESF suggested setting up a seventh, multidisciplinary Review Panel. **The multidisciplinary panel, including 18 members who were also members of the six disciplinary panels, created improved conditions for a**

[4] *Evaluation Guide, FCT Evaluation of R&D Units 2013*, p.12

truly interdisciplinary evaluation, at the same time benchmarking those applications against disciplinary panels. The multidisciplinary panel replaced the structure proposed originally – of several working groups independently assessing the same interdisciplinary application, which would not create favourable conditions for establishing a consensus.

Setting up Review Panels

Six disciplinary panels and one multidisciplinary panel brought together 74 individuals of whom 18 were also members of the multidisciplinary panel. Chairs of all six disciplinary panels were members of Multidisciplinary Panel 7 *ex officio*. The number of members in each panel was adapted to the number and disciplinary spread of applications while keeping the overall size of panels manageable (Table 2).

Table 2. Number of Research Units and members for the seven panels. Total number of panel members does not include P07 members as those were part of other panels.

Panel	Number of Research Units	Number of Members
1. Exact Sciences	39	11
2. Engineering Sciences	46	13
3. Health and Life Sciences	24	9
4. Natural and Environmental Sciences	24	8
5. Social Sciences	73	17
6. Humanities	72	16
7. Multidisciplinary	*44*	*18 – Members from the other panels*
TOTAL	**322**	**74 individuals – 92 panel members (some joined two panels)**

Each Review Panel member had on average 8.8 applications to review as lead or secondary rapporteur.

ESF identified and invited members of the Review Panels independently from FCT and the Portuguese Ministry of Education and Science; the final panel composition was, however, endorsed by the Ministry. To respond to the challenge of evaluating strategic development programmes of R&D Units, a specific profile of candidates was sought:

- Panel members should be internationally experienced scientists of the highest standard, have a wide view of the research domain of their Review Panel, with knowledge cutting across the specific disciplines covered by the panel.
- Panel members should also have good knowledge and experience in management of scientific structures and/or be experienced in institutional evaluation. The candidates were therefore sought among heads and former heads of laboratories, institutes or departments, members of international committees and panels.

From the start, Review Panel members were briefed on the specificities of the exercise.

The main driver in identifying Review Panel members was to provide an adequate disciplinary coverage for each panel, ensuring fair treatment of all applications. Considering the number of panel members to be appointed (74) and the specificities of some R&D Units, one of the numerous challenges was to ensure that Review Panel members had a sufficiently broad knowledge to ensure an adequate disciplinary coverage of all applications. This was achieved by appointing to the panels primarily individuals in charge of implementing and/or monitoring a broad research portfolio (e.g., heads of university departments, members of international committees). In the next step, lead and secondary rapporteurs were appointed following the closest match between the panel members and the R&D Units' research area(s). All candidates were checked for the absence of any conflict of interest (see Annex 2 for Conflict of Interest guidelines).

In addition to disciplinary coverage, the following criteria were taken into consideration:

- **Gender balance** was sought at equal level of expertise but was challenging to achieve practically. Among 205 persons contacted for panel membership, 48 were women which constitutes 23.4%. However, they made up only 16% of panel membership. Moreover, only one chair out of six was female. Table 3 below details gender balance per panel.

Table 3. Number of male (M) and female (F) panel members in each panel. The last row indicates whether the chair was F or M.

	Exact Sciences Panel 1	**Engineering Sciences** Panel 2	**Health and Life Sciences** Panel 3	**Natural and Environmental Sciences** Panel 4	**Social Sciences** Panel 5	**Humanities** Panel 6
Female	0	0	1	0	6	5
Male	11	13	8	8	11	11
Chair	M	M	M	M	F	M

- **Geographical coverage.** Panel members represented 25 countries. However, stronger participation of researchers based in the UK was clear (Figure 1), especially as 4 out of 6 chairs were from this country. This was partly due to the fact that UK researchers have significant experience in institutional evaluation so they constituted a reliable pool of people to draw from. It is also general ESF experience that UK researchers are keener than researchers from other countries to participate in assessment exercises as it is their tradition to consider it a part of their research mission and a broader contribution to their disciplinary community. While Review Panel members were primarily from European countries, it has to be noted that having two specialist external referees allowed the geographical scope of the evaluation to be expanded, also beyond Europe. The detailed breakdown of geographical representation per panel, including chairs' countries, can be found in Table 14 in Annex 3.

Review Panels were each chaired by one of their members designated by ESF. The chairs were all experts in their field, not only with extensive managerial experience but also with experience in evaluation and panel work. The role of the chair was mainly to lead and

coordinate panel work. Specifically, the chair was responsible for conducting the meeting, ensuring fair treatment of all applications and respecting criteria and procedures of evaluation.

One of the challenges was to adapt to a couple of very late drop outs from some panel members due to personal problems (this was the case for eight panel members).

To appoint 74 panel members, 205 potential members were identified and invited. Taking into consideration that they were invited at relatively short notice, it is a very good result, confirming that first choice experts constituted the panels.

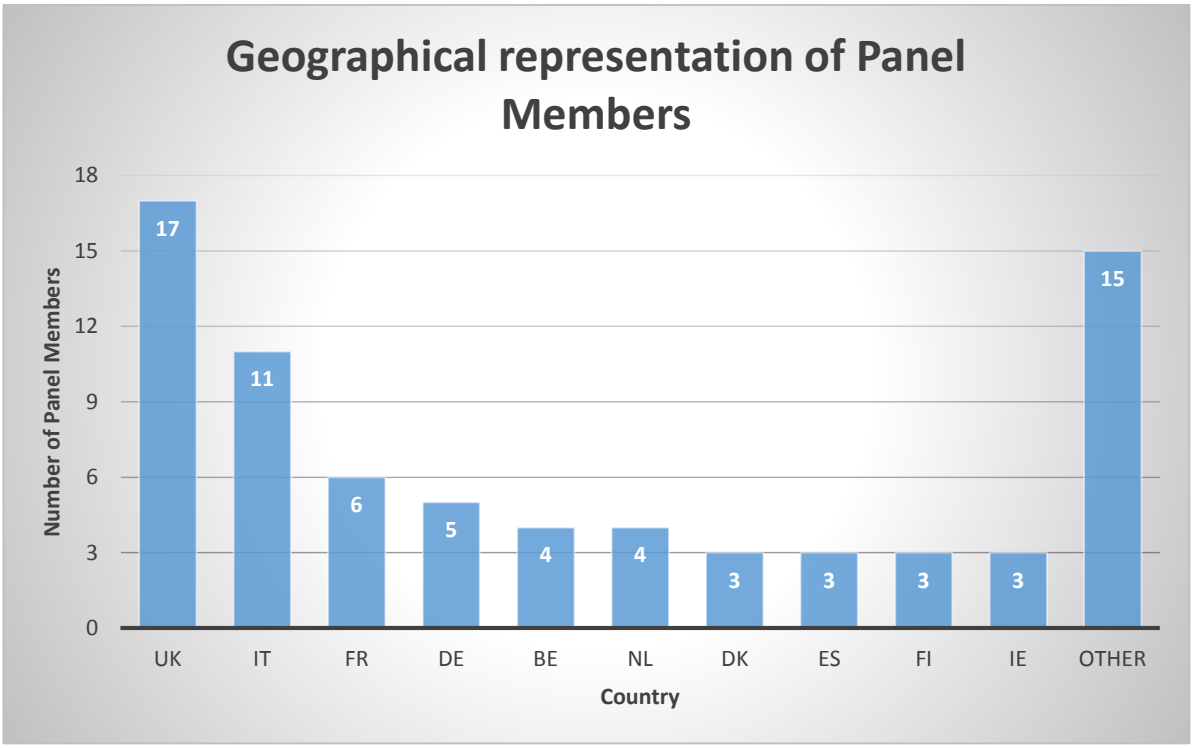

Figure 1. Histogram showing that 59 out of 74 Review Panel members (80%) represented 10 countries. Each of the remaining 15 panel members (right hand bar) represent one country (AT, CH, CY, CZ, EE, GR, HR, HU, IL, LU, NO, PL, SE, TR, US). Note that Portugal (PT) was not represented.

Stage 1 evaluation process, methodology and outputs

Stage 1 Approach and Concept

Stage 1 evaluation included the provision of four individual reviews for each application followed by a plenary discussion, scoring and grading at Review Panel meetings. This stage was fully and independently coordinated by ESF.

For each application, two independent reviews were submitted by experts not sitting in the Review Panels identified by ESF. These external, remote referees had a high level of expertise in fields indicated in the application (and significant research management experience, where possible). Two other assessments were provided by Review Panel members, lead and secondary rapporteurs. While the secondary rapporteur reviewed the application independently, the lead rapporteur's role was to integrate both his/her own evaluation and the

assessment reports from the two external referees and the secondary rapporteur. All four assessment reports were available to panel members before the meetings on the IT platform and constituted a starting point for panel discussion.

This procedure was in line with the procedure announced by FCT which provided for up to five individual reviews prepared for each application.

Having assessments from two specialist external referees and two Review Panel members combined the benefits of expertise in any given R&D Unit's field with the broader and more balanced perspective provided by comparison with other related disciplines, thus ensuring consistency and coherence across the evaluation. With 659 experts involved in the process (external referees and Review Panel members), **this approach considerably increased the ability of Review Panels to perform a robust and thorough evaluation**.

According to FCT procedures, R&D Units were invited to suggest up to three names of experts whom they considered to be qualified to assess their application. It was expected that one of the three experts indicated by each R&D Unit would be invited to carry out an individual review. This did not prove to be a feasible solution for a number of reasons. First, not all R&D Units suggested three names of potential referees; some provided one or two and 50 out of the 322 provided none. Second, of those suggested, in a large number of cases they were not appropriate as they had a conflict of interest and/or their profile did not match the ESF quality standards. A limited number of experts indicated by the R&D Units and fulfilling ESF criteria were invited to provide an assessment. For fairness and equality of treatment, it was decided that all applications would be assessed by the same number of individuals so for the remaining applications ESF identified experts.

External review

Suitable potential external referees were identified on the basis of their close matching expertise to a given R&D Unit proposal (determined by the content of the application), coupled to track record (e.g., h-index) and absence of conflict of interest. In addition to its own database – which counts over 35,000 researchers with a scientific profile – ESF staff also has access to Scopus, as well as other abstract and citation databases (e.g., Google Scholar).

To collect 644 valid reviews from external experts, the ESF identified and invited 2,799 potential targeted referees. The average success ratio was slightly over 4/1 – i.e., 4.35 reviewers had to be contacted to secure one review. While this is slightly over the ESF average for peer review exercises (between 3 and 4 contacted for one received), it is not unduly so and not unexpected considering the specific nature of the applications (R&D Units and their strategic programmes), their format and length (sometimes over 100 pages) and tight timeframe. In 71% of cases (230 R&D Units), ESF invited between two and ten targeted potential external referees to collect two assessments, in 26% of cases (82 R&D Units), the ESF invited between 11 and 20 potential external referees to collect two assessments, in only 3% of cases (10 R&D Units), ESF had to contact more than 20 (and up to 34) potential referees. Finally, and illustrating that external referees were narrowly targeted, in 91.5% of cases, external referees were invited to review only one application.

The process implemented by ESF involved a suite of check-points, the first one being that ESF checks the quality of the outputs from reviewers before these are circulated to the panels. Each review is checked for quality on the basis of the following criteria:

- Substantiation (are marks supported by meaningful comments?)
- Coherence between marks and comments
- Completeness (are all sections/criteria addressed?)
- Language (is it respectful of applicant and application alike; is it understandable?)

In this context, ESF rejected a round dozen of reviews as definitely below expectations and invited another external referee. For another approx. 10%, reviewers were requested to further clarify and substantiate their assessments before these were accepted. While no screening process is 100%, those steps did reduce the risks of inappropriate assessments (in content or form). However, it has to be noted that in some instances Review Panels were not fully satisfied with external reviews provided. The impact of assessments considered of suboptimal relevance was buffered by the fact that each unit was assessed in detail by four individuals (and later by the whole panel) as well as by the content of the applicants' rebuttal.

As mentioned above, three assessments were provided independently (i.e., without any input from, or interaction between, evaluators) for each application by the two external referees and by the secondary rapporteur. Applications were marked out of 20. Overall, the three independent reviews were consistent with each other as quantified by low average standard deviations between these three scores[5] (for 79% of applications, the standard deviation of remote assessment scores is below 3, Table 4).

Table 4. Percentage of proposals for which the standard deviations between remote assessments fall in 1-SD ranges.

Standard Deviation	Number of Applications	%
0 - <1	70	22%
1 - <2	89	28%
2 - <3	93	29%
3 - <4	45	14%
≥4	25	8%

Furthermore, deviation in the scoring behaviour of the secondary rapporteur compared to the two external referees was checked. To quantify this, a simple statistical test on mean scores was conducted. The average original cumulative scores of individual assessments were statistically not different between the external referees (15.1, SD=0.8) and the secondary rapporteur (14.7, SD=1.1) as revealed by a paired t-test across the seven panels (Figure 2), $t(6)=1.6$, $p=0.182$). This conclusion also holds for the dispersion of the scores as measured by their SDs (data not shown, SD=3.4, $t(6)=1.5$, $p=0.188$). This was further substantiated by the

[5] Dispersion, although low here, could still be explained by a systematic outlier. Therefore, we verified that no systematic bias could be attributed to the secondary rapporteur (SR). To do this, we compared the average of the unsigned difference between scores attributed by external referees (ER) and between each ER and the SR and found no difference, |ER1 – ER2|=3.4, |ER2 – SR|=3.1 and |ER1 – SR|=3.2.

fact that there was a solid correlation, for each panel, between scores attributed by the external referees and the secondary rapporteur both overall (r=0.35 to 0.57) and for each criteria taken separately (see Table 15 in Annex 3). **These elements demonstrate a good level of coherence and consistency across the evaluations provided for individual proposals and that there has been strong continuity between external assessments and panel discussions.**

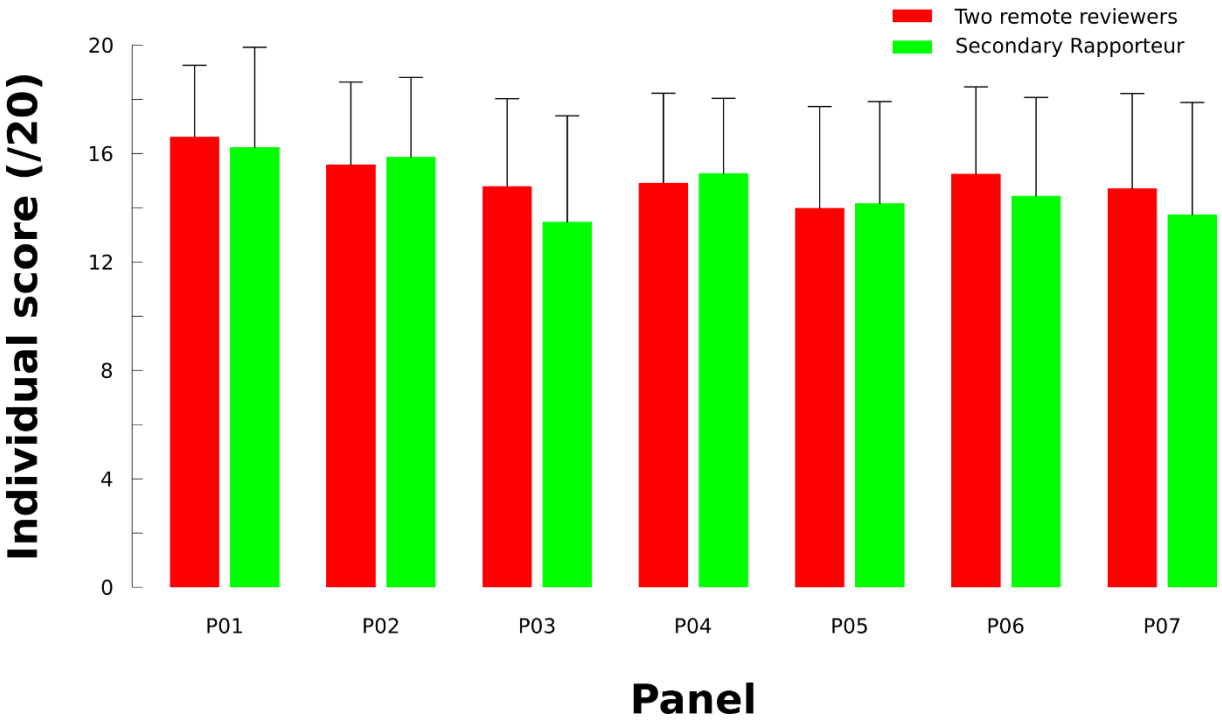

Figure 2. Mean scores across panels provided for the two external referees (red bars) and for the secondary rapporteur (green bars). Error bars are standard deviations.

A real challenge lay in securing all 644 reviews within 11 weeks (February to mid-April 2014), while giving reviewers sufficient time – three weeks – to perform their assessment. Although the timeline was tight, **each application was eventually provided with four individual reviews** (two from the external referees and two from the rapporteurs). The individual reviews were submitted in the Individual Reviewer Evaluation Form to the IT platform at ESF.

Rebuttal phase

The aim of this phase, part of FCT procedures, was to give all applicants the opportunity to prepare responses to the assessments and comments contained in the Stage 1 consensus reports. When applicable, these responses would have been taken into account by the Review Panels in Stage 2 of the evaluation. Following discussions between ESF and FCT, the rebuttal phase was implemented before the first Review Panel meeting, i.e., before any discussion took place. R&D Units had access to the three individual assessments (two from the external referees and one from the secondary rapporteur) and could provide their feedback online via the IT platform to correct errors and misunderstandings. This feedback was examined as part of discussion at the first Review Panel meeting.

The rebuttal step was a novelty compared to former FCT assessment exercises of R&D Units. The units did not always have a proper perception of its rationale and use and, as a result, the rebuttals were of variable quality, often void of content. In those cases, they did not fulfil their role as they were not a concrete response to the assessment reports providing additional evidence. Instead of correcting factual mistakes or misunderstandings, in some limited cases some R&D Units seemed to over-react to the content of the assessments provided and others provided feedback that could be perceived as inappropriate. **The Review Panels expressed an opinion that this is a useful step in every evaluation exercise**. However, they pointed out the misuse of the procedure, noting that rebuttal aims and criteria need to be clarified to make the rebuttals be based on factual information only.

It should be noted that, independently of the rebuttal phase and in accordance with Portuguese law, applicants had a right to formal appeals (hearings) on the panel deliberations at both stages of the procedure. The appeals followed a legal format as set out in Portuguese law and were handled by FCT. Review Panel members were requested by FCT to provide opinions on scientific aspects of the appeals.

Review Panel work, meetings and outputs

Members of the Review Panels were aware of this message from FCT throughout their work:

"*The 2013 R&D Units evaluation call is the FCT's foremost funding instrument for promoting quality in research performed in Portugal. Establishing R&D Units with long-term and stable funding gives the institutions an opportunity to restructure their research activities and develop new collaborative relationships to enhance their position on the international research front.*"[6]

Review Panel members were provided with all documentation FCT published on the evaluation. They were also provided with the documentation, prepared jointly by ESF and FCT, outlining their role and guiding them through the evaluation process, and detailing their tasks at each stage of the evaluation including instructions for the use of the ESF IT platform.

As mentioned above, to increase the collegiality of decisions, at each stage of evaluation one physical meeting of the seven Review Panels was organised. Those meetings were key moments of the work of the panels as collegial discussions and decisions regarding scoring, grading and consensus reports took place there.

For Stage 1, the Review Panel meetings took place in Amsterdam on 26-29 May 2014. At the opening, Professor Paulo Pereira, FCT Vice-President, presented the *FCT Evaluation Context*. He was followed by a presentation by Mr Nicolas Walter, ESF Senior Science Officer for Peer Review Services, entitled *Evaluation and Meeting Process*.

Panels were asked to review all applications and to rank each of the following:

- Criterion A. Productivity and contribution to the National Scientific and Technological System (NSTS);
- Criterion B. Scientific and technological merit of the research team;
- Criterion C. Scientific merit and innovative nature of the strategic programme;

[6] *Evaluation Guide, FCT Evaluation of R&D Units 2013*, p.6

- Criterion D. Feasibility of the work plan and reasonability of the requested budget.

Each criterion was weighted 25% in the final mark and was rated on a five-point scale following the description below (Table 5):

Table 5. Numeric score, corresponding wording and definition of categories.

Numeric score	Corresponding wording	Definition
5	Excellent	All relevant aspects of the assessment criteria successfully addressed. Any shortcomings are minor.
4	Very good	Assessment criteria very well addressed/met, although certain improvements are still possible.
3	Good	Assessment criteria well addressed/met, although improvements would be necessary.
2	Fair	Assessment criteria broadly addressed; however, there are significant weaknesses.
1	Poor	Assessment criteria addressed in an inadequate manner, or there are serious inherent weaknesses.

The addition of marks given led to an overall mark that resulted in the classification of evaluated units following the guidelines below (Table 6):

Table 6. Guidelines for providing coherent cumulative scores and grades.

Grade	Description	1st Stage Cumulative Score/20
Exceptional	R&D Unit recognised as an international reference for its scientific and technological output and exceptional contributions to its area of research	≥ 15[1]
Excellent	R&D Unit distinguished by the high quality and international merit of its scientific and technology output and with significant contributions to its area of research	≥ 15[1]
Very Good	R&D Unit with high quality and national merit and with significant contributions of international relevance in its area of research	≥ 15[1]
Good	R&D Unit with quality at the national level, reduced internationalisation and some contributions to its area of research	< 15 > 12[2]
Fair	R&D Unit without significant contributions to its area of research	≤ 12 > 11[3]
Poor	R&D Unit without contributions to its area of research and with other weaknesses	< 11

Notes:

1 *Additionally the application must score at least 4 points in each of the ratings of criteria A and C, and it must also score at least 3 points in each of the ratings of criteria B and D.*

2 *Additionally the application must score at least 3 points in any of the four evaluation criteria ratings.*

3 *Additionally the application must score at least 3 points in each of the ratings of criteria A and C, and it must also score at least 2 points in each of the ratings of criteria B and D.*

Before the meeting, the following steps were completed:

- Collection of assessments from four experts: two external referees and two Review Panel members;
- Submission of applicants' comments and feedback on the assessments provided by the external referees and the secondary rapporteur (rebuttal step).

During the meeting, Review Panel members reviewed in plenary all the applications submitted to their panel following the scenario outlined in the meeting guidelines. Each application was introduced by the lead rapporteur, then commented upon by the secondary rapporteur and an open discussion followed. Arbitration was made by consensus, taking into account all applications assessed by the panel.

Having plenary disciplinary panel meetings allowed a higher level of collegiality and a better 'normalised' perception of marks and classification than the structure designed initially (e.g., 'good' grade or 3 out of 5 scoring would naturally not have the same meaning for all 30+ workgroups originally planned). At the same time it followed the general spirit of the original procedure.

The discussion and the conclusions (expressed in scores and grades) were led and moderated by the Review Panel chair. Each chair was supported in her/his work by one ESF panel secretary. The role of the secretary was to assist the chair in conducting the meeting, providing information on procedures and criteria when necessary and keeping track of conclusions. Prior to the meeting, the secretaries identified candidates for their panel, allocated applications to panel members for assessment in accordance with their expertise and monitored the work of the panel so they were fully up to date with the process, character of the applications and panel expertise. Panel chairs acknowledge the importance of this support.

The Vice-President of FCT, Professor Paulo Pereira, and FCT science officers were present during the meeting and attended individual panel meetings, providing additional information and clarifications.

The final step of Stage 1 was the preparation of consensus reports which summarised individual reviews and the discussion at the Review Panel meeting, providing substantiation for final scores and grades. R&D Units which were *not* selected for Stage 2 received all results (external reviews with individual scores as well as Review Panel consensus reports with individual scores attributed to each evaluation criterion and automatically calculated qualitative grades) after the first Review Panel meeting. R&D Units which were selected for Stage 2 received only Review Panel consensus reports, without scores.

Each report was written by the lead rapporteur in collaboration with the secondary rapporteur and finally checked for consistency with regards to the panel meeting deliberations by the ESF

secretary to the panel in collaboration with the panel chair. The ESF IT platform allowed easy and efficient interactive work.

The consensus reports for R&D Units to be visited in Stage 2 all included a list of issues to be addressed by the unit coordinator during the site visit. This step created a dialogue between the panel and the units to be visited, allowing the units to be aware of and subsequently address the points of specific concern for the panel. It should be noted that science management and managerial capacities of the unit coordinators had importance for the panels and were discussed already at Stage 1. Questions for site visits reflected a need for managerial information which panels found missing in applications. All Review Panels requested a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis and meetings with PhD students and post-docs as a standard element of each unit presentation during the site visit (most units did provide and present this information).

As already mentioned, in accordance with Portuguese law, R&D Units had the right to submit an appeal within 10 days after notification of the Stage 1 results, which should be answered before the beginning of Stage 2 of the evaluation process. All appeals submitted were performed as required by law and coordinated by FCT. FCT sought opinions of Review Panel members in cases where complaints referred to the scientific content of the evaluation. While the appeal process was known to ESF, its unexpected magnitude required a very significant additional effort from panel members. **Despite that unforeseen additional workload, the panel members all collaborated with FCT on appeals and contributed significantly to the process, being aware of its importance to the overall success of the evaluation**. All appeals were answered in time to raise scores and grades of ten R&D Units above the threshold and include them in the schedule of site visits.

An important element of the procedure laid out by FCT was the provision of **bibliometric analysis** to the panels as a supporting tool. This analysis gathered statistics on R&D Units' production (publications and patents) in the period 2008-2012.[7] For this analysis, FCT adopted, as a unique identifier for the researchers, the ORCID iD and the database Scopus from Elsevier as a source to use in the study. According to FCT, with this bibliometric exercise, 97% of the integrated members registered their ORCID iD and, for the evaluation period (2008-2012), 90% of all publications from Portugal indexed in the Scopus database, were imported by researchers and considered in this bibliometric study. The research unit researchers had access to Scopus to update their publication lists and ORCID iD before Elsevier extracted the data and compiled the bibliometric files. Publications of researchers - integrated members of R&D Units - who did not register their ORCID iD in FCT's platform or who did not populate their profile with publications from the Scopus database were not considered for the bibliometric study.

Panel members were requested to use solely the analysis provided and avoid using any other bibliometric tools including their own analyses. FCT stressed in their document that bibliometric information should not replace peer judgment by panel members and that its use should be decided by each panel. The bibliometric analysis made available on the ESF IT platform was used by Review Panels as complementary to R&D Units applications and to individual

[7] http://www.fct.pt/apoios/unidades/avaliacoes/2013/analise_bibliometrica.phtml.pt

assessments provided in the frame of the evaluation, in particular for considering past achievements.

Some panel members found it more, some less useful. In their feedback, panels suggested that it would be more useful if presented in a condensed and synthesised format. They also regretted that FCT provided an updated set of bibliometric data – with corrected information on units' Full Time Equivalent – for Stage 2 evaluation. Although the panel thought this additional, revised data did not have significant impact on the final outcome of Stage 1 (e.g., relevant data, like h-index, was not affected) it would be preferable to have had the same set of data from the beginning.

As bibliometrics is used differently in the evaluation of different research domains, some panels raised specific issues. In particular, members of Panel 5 (Social Sciences) felt it would have been good to use more than one source for the bibliometrics as Scopus excludes many journals that are still recognised as important in the social sciences. From their point of view Google Scholar could be a good supplement to Scopus in future exercises. The same approach would be helpful for the humanities. Panel 6 (Humanities) used the bibliometric data to a limited degree, mainly for psychology; in fact, in other fields the analysis included little relevant data as humanities journals are rather poorly represented in Scopus.

Outcome of Stage 1

During Stage 1 there are three time points at which assessment scores can differ: before the Review Panel meeting, after panel consensus and after appeal. Table 7 reports mean scores at these three time points and highlights a correlation between scores before panel meetings (reflected by the integrated assessment of the lead rapporteur) and after consensus.

Table 7. Total mean score per panel at three time points during the first stage evaluation. 'Before': mean of raw scores; 'Consensus': mean of scores of consensus reports after the panel meetings; 'After appeal': mean of scores taking into account scores updated following the legal appeal procedure.

	Before	Consensus	After appeal
1. Exact Sciences	16.46	15.46	15.49
2. Engineering Sciences	15.66	15.20	15.35
3. Health and Life Sciences	14.33	14.00	14.08
4. Natural and Environmental Sciences	15.01	15.29	15.42
5. Social Sciences	14.02	13.86	13.92
6. Humanities	14.96	14.63	14.75
7. Multidisciplinary	14.37	14.02	14.16
Global	**14.97**	**14.64**	**14.74**

Figure 3 quantifies this by plotting the overall score before (x-axis) and after (y-axis) the panel meeting, for all panels and the four criteria.

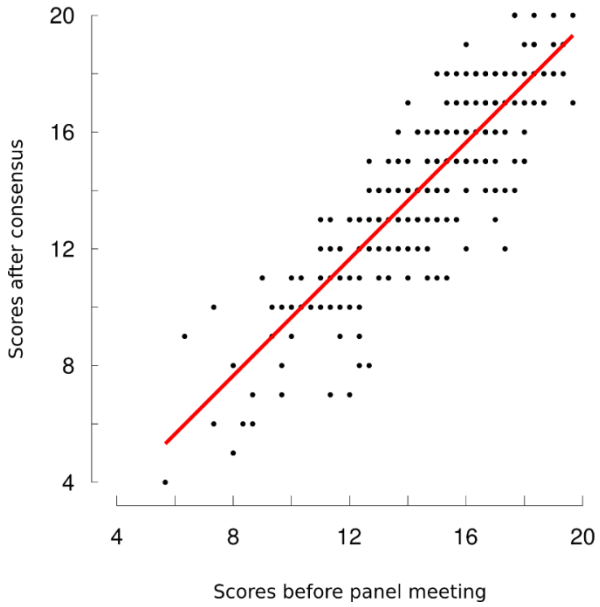

Figure 3. Correlation between total scores before any discussion and scores after panel consensus (r=0.86, p<0.001; $S_{after}=S_{before}$-0.35, linear gain=1).

This was also very consistent per panel and for each criteria (see Table 16 in Annex 3).

Finally, it is interesting to report that, when comparing the average remote assessment marks with the actual final marks (Table 8), the difference between those values increases (as does the standard deviation). This demonstrates the buffering and normalisation effect of the consensus made possible through the Review Panel meetings.

Table 8. Normalisation effect of the panel meetings on scores.

Standard Deviation	Average difference between the averaged remote assessments and the final scores
0 - <1	0.8
1 - <2	1.1
2 - <3	1.2
3 - <4	1.5
>4	1.9

Altogether, this analysis demonstrates that panel members (1) took into account external referee scores, (2) adopted a consistent strategy for all scores and (3) were consistent in their approach across panels.

In all panels, the appeal process translated into slightly higher mean average scores (Table 8). These marginal changes affected applications positively as 10 additional R&D Units proceeded to Stage 2 (Table 9). Table 17 in Annex 3 presents these data per panel and shows that the

number of proposals going to Stage 2 increased in five out of seven panels and remained constant in two panels.

Table 9. Number (and %) of applications falling into the different qualitative categories ('Classification') after panel consensus ('consensus') and after appeal ('After appeal').

Overall		
Classification	**Consensus**	**After appeal**
Poor	35 (10.9%)	32 (9.9%)
Fair	36 (11.2%)	31 (9.6%)
Good	83 (25.8%)	81 (25.2%)
Proceeds to Stage 2	168 (52.2%)	178 (55.3%)

Further information on the outcome of the Stage 1 evaluation provided by FCT to the R&D Units indicated that:

- The 55.3% of units proceeding to Stage 2 represent 71% of the Integrated Researchers;
- 77% of the units that restructured through a merger of previous units proceeded to Stage 2;
- 70% of previously existing units (same identity, irrespective of internal reorganisations) proceeded to Stage 2;
- Only 38% of totally new units proceeded to Stage 2.

Stage 2 evaluation process, methodology and outputs

Stage 2 was implemented and coordinated by FCT; ESF provided documentation for site visits based on Stage 1 documentation and was an observer to the process, providing advice when requested. In its observer role, ESF delegated representatives to attend at least one session of site visits per Review Panel (4-5 days); and delegated some ESF staff to attend the second Review Panel meetings.

Site Visits

During the site visits in Portugal 178 R&D Units were visited. The visits were organised in 22 sessions lasting from 4 to 6 days and visiting on average 8 units. Visiting teams consisted of 3-4 panel members on average, accompanied by an FCT scientific officer. As mentioned above, some teams were accompanied by an ESF representative acting as observer, who submitted an observer report after the session.

Each site visit was planned for 3 hours and consisted of a standard programme usually including:

- Welcome by head of the host institution
- Presentation of the unit by coordinator/director (max. 20 min.)
- Presentations by other staff if so planned by the R&D Unit coordinator/director

- Discussion with coordinators/directors and sometimes staff
- Visit of facilities if relevant
- Meeting with PhD students and post-docs (max. 60 min.).

Each visiting team started with a briefing meeting at FCT headquarters in Lisbon where FCT Vice-President Professor Paulo Pereira introduced the site visit phase. The meeting in a small group was an opportunity to ask questions and share observations on the evaluation process. Those meetings were considered useful by most teams, although maybe taking place too late in the process.

In the meeting, evaluating teams were reminded that in their evaluation during site visits they should apply also criterion E – *Impact of the scientific, technological and cultural output*, not applied in Stage 1 evaluation. Although the *Evaluation Guide* mentioned this criterion, its actual implementation was not discussed in Stage 1 and its application was not understood in the same way by all visiting teams which led to some confusion during the final Review Panel meetings.

The programme of site visits was suggested by FCT and respected by the units, although they were also flexible if visiting teams requested adaptations due to changes in their schedule (e.g., delays or number of facilities to be visited). Evaluators were generally satisfied with this format, suggesting, however, that in future the duration of site visits should be proportionate to the size of the R&D Unit and number of laboratories. This is important for research domains where laboratories and facilities play a significant role. In the case of humanities and social sciences this time could be used for other purposes (e.g., studying digital infrastructures like databases or lexicons mentioned in applications).

The units were, globally, well prepared for the site visit. As panels and ESF observers reported, some visits took longer, up to 4 hours, mainly in units with significant laboratories and/or with more than one location.

The schedule of all sessions was very intense. Usually there were two visits per day, one before and one after lunch. Since each session covered several cities, the group had to travel quite long distances in addition to time spent on actual site visits.

All panels in their reports noted that they felt uncomfortable sharing lunch after the meeting with the board of directors/coordinators of the R&D Unit visited (in some cases also with unit staff), mainly for two reasons. On the one hand, this could have been the occasion for the unit to extend the discussion which would give some units more time with the evaluators than others. Second, lunch time would have been a very good opportunity for the visiting team to wrap up the visit.

An important part of all visits were meetings with PhD students and post-docs without the presence of the unit coordinator. Some teams asked roughly the same set of questions each time which allowed for a very good basis of comparisons across units. Evaluators remarked that it would be commendable if all units limited participation in those meetings to PhD students and post-docs only, not mixing them with master and undergraduate students as happened in some cases. Some also suggested that meetings could be structured somehow and poster sessions by PhD students could be considered in the future.

For each unit visited, one of the panel members was designated as a lead rapporteur, chairing the visit and the discussion and elaborating the final consensus report. This was in line with the procedure provided for by FCT. Ideally the lead rapporteur from Stage 1 should have been in each case designated as the chair for Stage 2. This was, however, not feasible as it was not possible to coordinate site visits and schedules of panel members to achieve this. The logistics of planning the visits with the participation of 74 researchers was very complex. It had to accommodate their already fixed commitments.[8] In some cases, last minute cancellations occurred for personal reasons. It should, however, be noted that Review Panel chairs participated in a high number of Stage 2 site visits, replacing absent colleagues and acting as lead rapporteurs and ensuring continuity and coordination.

Each site visit was followed by a wrap up session of the visiting team summarising their observations and conclusions (the teams found different ways of incorporating them into their busy travelling schedules). Then, the lead rapporteur/chair of the site visit prepared a site visit report in consultation with all members of the team and submitted it to FCT via the ESF IT platform and made it available to all panel members before the second Review Panel meeting.

All panels considered site visits as a crucial element of the evaluation even if the visits usually confirmed assessments made in Stage 1; revisions of grading were made only in a few cases. However, the visits enabled 'live' verification of opinions worked out on the basis of paper documentation in Stage 1. Despite demanding schedules and sometimes tiring, long days visiting teams were enthusiastic about this experience. The ESF observers stressed that panel members took their task very seriously and acted with high professionalism and efficiency. This was also helped by the fact that most of them had previous experience of similar evaluation exercises and knew how to make best use of the visits. This is how one of Review Panel chairs reported on the visits:

"*The site visits were invaluable; though intense (we recognise that this was inevitable in terms of efficient use of panel members' time) with packed schedules, they were of sufficient duration to allow panel members to get at the academic core behind the paper submissions, to ask challenging questions of Centre members, where relevant to view any specialist facilities and to see for themselves the life blood of the centres concerned, from academic staff and Centre directors to PhDs and postdocs.*"

The teams were also generally satisfied with support provided by FCT staff during the site visits. A FCT representative was present during each visit and their assistance with the challenging and sometimes complicated practicalities of the visits was very much appreciated. It was felt that there were occasions, however, when the secretaries could have given stronger procedural guidance, to ensure that all aspects of the evaluation were covered (e.g., the application of criterion E).

[8] It should be kept in mind that invitations to Review Panel members were issued early in 2014 when many of them already had fixed commitments for the summer; moreover, the visits could only be planned after results of Stage 1 were known, that is in June.

Final Review Panel consensus meetings and outputs

All seven panels reconvened physically at the final consensus meetings (Lisbon, 24-26 November 2015) to discuss and finalise the evaluation of all R&D Units visited and reach consensus on their final classification. As already mentioned, this step allowed a better level of collegiality and integration of results.

The administrative support was provided by FCT. ESF was represented by observers visiting meetings of individual Review Panels.

All panels were welcomed by Professor Paulo Pereira who gave a presentation on the aims of the final evaluation step in the context of research strategy and policy of Portugal.

In their proceedings all panels followed the scenario outlined in the meeting guidelines. Each application was discussed individually; observations and conclusions from the site visits were the key factor in considerations on the final scores and grades. The main approach was peer review and bibliometric analysis was only a supporting tool. It is justified to say that panels discussed each application with care and insight. Their aim was to ensure that research plans and strategies presented in the application will result in excellent science and are robust and feasible. They were also concerned that the feedback provided to units in consensus reports is useful for their future development. In this context, they paid substantial attention to the science management aspects and the managerial capacities of unit coordinators. This was a continuation of the approach from Stage 1 when the panels requested a SWOT analysis from each R&D Unit to be visited.

At Stage 2 of the evaluation, starting with site visits, a fifth criterion, *E: Impact of the scientific, technological and cultural output,* had to be considered. The introduction of an additional evaluation criterion E, which also resulted in a change of the dedicated criteria weighting rule, caused some confusion. Firstly, not all site visit teams provided provisional scores for criterion E in their site visit reports. As already mentioned, this issue was not clearly enough discussed at the meetings with the FCT Vice-President at the beginning of each site visit session, nor was it clear for all FCT secretaries supporting visiting teams. Panel members felt that the criterion should have been introduced from the beginning and that the concept of impact, especially with regard to basic research, is not straightforward and requires clarification. Discussions over the interpretation and application of this criterion and then scoring of individual units took substantial time and effort in some panels during the meeting. Panel 6 (Humanities) especially found the concept difficult to apply. On the other hand, Panel 5 (Social Sciences) noted that "*the capacity for Centres to choose their own basic/applied ratio for research was a good feature, as was differentially weighting criterion E on the basis of this ratio. We also did not find much difficulty in dealing with criterion E in the site visit reports and at the Stage 2 meeting*". Another issue with the introduction of a new criterion in Stage 2 was the fact that panel members considered it a bit challenging to transition from an assessment out of 20 to an assessment out of 25. This also necessitated adapting to a seemingly new cut-off threshold for units falling in categories below 'very good'.

Criterion E also induced relative weighting between criteria A and E. The resulting criteria weighting was:

- Criterion A – 20% to 35%
- Criterion B – 20 %
- Criterion C – 20%
- Criterion D – 20 %
- Criterion E – 5% to 20%

The relative weighting of Criteria A and E depended on the specific research profile(s) of the R&D Units (basic research or applied research/experimental development, Table 10). Therefore, R&D Units with a basic research profile were assessed with a lower weighting in criterion E (i.e., 5%), which will be balanced by a higher weighting in criterion A (i.e., 35%).

Table 10. Weight for criteria A and E according to research profiles.

Research Profile		Weight of criterion A	Weight of criterion E
Basic research	**Applied research**		
76-100%	0-24%	35%	5%
51%-75%	25%-49%	30%	10%
26-50%	50%-74%	25%	15%
0-25%	75-100%	20%	20%

Each criterion was assessed on a scale of five and, unlike during the first Review Panel meeting, the use of half points was introduced. The compilation of criteria marks resulted in the classification of units using the following guidelines (Table 11):

Table 11. Guidelines to provide coherent cumulative scores and grades.

Grade	Description	Stage 2 Cumulative Weighted Score/25[a]
Exceptional	R&D Unit recognised as an international reference for its scientific and technological output and exceptional contributions to its area of research	25.00
Excellent	R&D Unit distinguished by the high quality and international merit of its scientific and technology output and with significant contributions to its area of research	< 25.00[1] ≥ 23.00[1]
Very Good	R&D Unit with high quality and national merit and with significant contributions of international relevance in its area of research	< 23.00[1] ≥ 18.75[1]
Good	R&D Unit with quality at the national level, reduced internationalisation and some contributions to its area of research	< 18.75[2] > 15.00[2]

Fair	R&D Unit without significant contributions to its area of research	≤ 15.00[3] ≥ 13.75[3]
Poor	R&D Unit without contributions to its area of research and with other weaknesses	< 13.75

Notes:

1. *Additionally the application must score at least 4 points in each of the ratings of criteria A and C, and it must also score at least 3 points in each of the ratings of criteria B and D.*
2. *Additionally the application must score at least 3 points in any of the four evaluation criteria ratings.*
3. *Additionally the application must score at least 3 points in each of the ratings of criteria A and C, and it must also score at least 2 points in each of the ratings of criteria B and D.*

In each of these cases, units that do not fulfil the additional conditions were downgraded one level.

Another challenging issue was the request from FCT to give an opinion on the alignment of the R&D Units' research themes with the Smart Specialisation Strategy for Portugal. Although it was mentioned in the *Evaluation Guide* as a part of criterion C, it was not considered in Stage 1 as the Smart Specialisation Strategy for Portugal was not yet published. At the final consensus meetings in Lisbon, the panels were asked to consider this issue. However, they were also informed that it would have no influence on the scoring. They felt that the meeting schedule was demanding already, when the task is not directly related to the evaluation. They also did not feel they had the expertise to assess 'eligibility to access EU structural funds' as it was phrased in the meeting guidelines. Not all of them were familiar with EU regulations and even less with Portuguese national policies for structural funds.

On the more practical side, panels welcomed the fact that FCT introduced the use of half points when assessing individual criteria from 1 to 5 at the final consensus meeting as it was helpful in fine-tuning individual criteria scores. It was, however, felt by panel members that use of half points should have been introduced earlier in the process and certainly for site visit reports. Similarly, panel members also felt that the above guidelines for classification (what mark for what category) should have been known by them before the site visits.

The Multidisciplinary Panel 7 was a complex one. In the case of this panel, it happened more often that site visit teams were composed of people who were not lead or secondary rapporteur at Stage 1 of the R&D Unit being visited, which was related to the difficulty in coordinating the presence in Portugal of many researchers at the same time. The chairs dealt with this problem, nevertheless, very well and ensured smooth deliberations.

All panels were supported by FCT panel secretaries, in most cases the same who attended site visits. They provided diverse levels and types of support: some provided only technical support (e.g., keeping an electronic record of scores), while some provided more procedural support, responding to questions related to the evaluation procedure and criteria, making sure that all steps were taken, etc. The Vice-President of FCT, Professor Paulo Pereira, was present most of the time and visited panels during their deliberations, responding to all questions and comments.

Outcome of Stage 2

The output of the meetings were lists of R&D Units visited with final scores and grades as well as consensus reports.

After Stage 2, results were provided to R&D Units and again a round of formal appeals followed. Panel members duly and patiently participated in them, supporting FCT in cases where complaints referred to the scientific content of the evaluation. Table 12 summarises how the second appeal process influenced the final outcome of the evaluation and Figure 4 presents the final outcome in terms of Integrated Researchers in each category. In short, it skewed the distribution toward better outcomes (Table 18 in Annex 3 for details per panel).

Table 12. Number (and %) of R&D Units falling into the different qualitative categories ('Classification') after each phase of evaluation. Italic lines highlight changes. Covers all 322 units.

Overall			
Classification	**After Stage 1 meetings**	**After Stage 2 meetings**	**After Stage 2 appeals - FINAL**
Poor	32 (9.9%)	32 (9.9%)	32 (9.9%)
Fair	31 (9.6%)	*33 (10.2%)*	33 (10.2%)
Good	81 (25.2%)	*90 (28%)*	*89 (27.6%)*
Very good	178 (55.3%)	104 (32.3%)	*97 (30.1%)*
Excellent		52 (16.1%)	*60 (18.6%)*
Exceptional		11 (3.4%)	11 (3.4%)

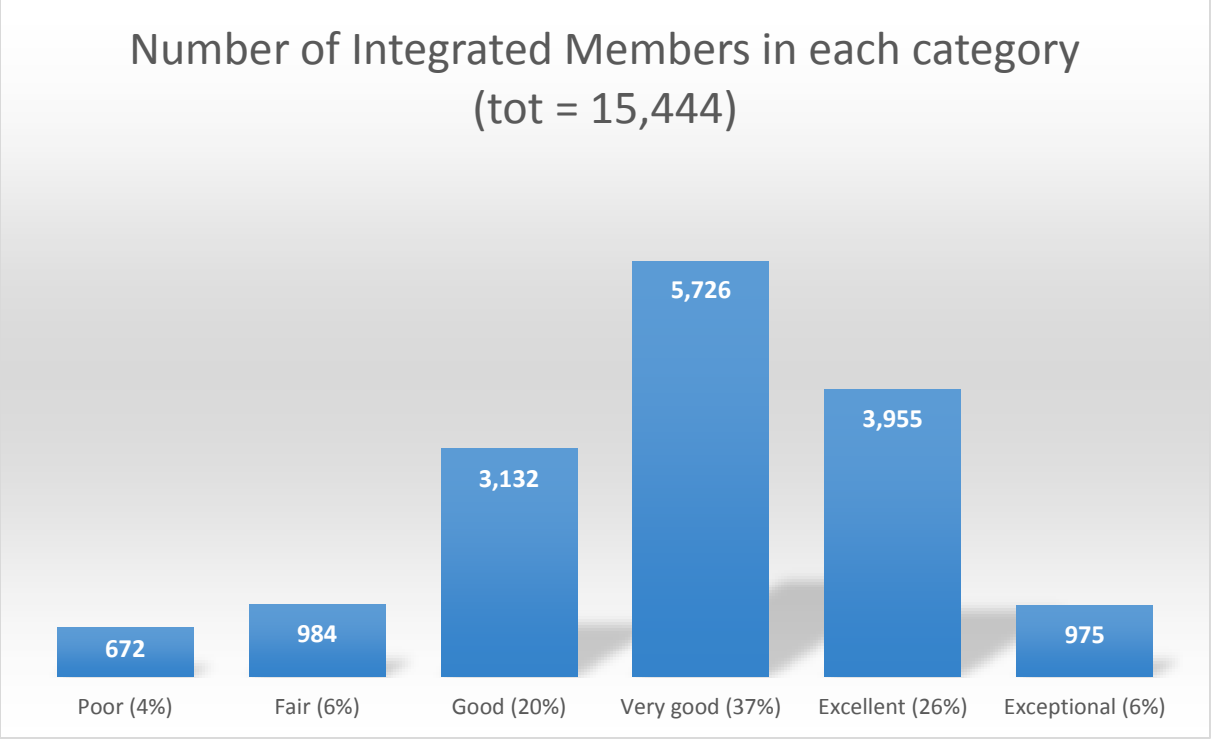

Figure 4. Number of Integrated Researchers employed in all 322 applicant R&D Units according to the evaluation outcome categories. FCT data.

The final results of the evaluation (after Stage 2 and appeals) and the funding granted to each unit were announced by FCT on 25 May 2015.

Ensuring quality of process and evaluators

ESF applied quality assurance mechanisms at all stages under its control in order to reduce the risks of inappropriate or unfair assessments:

- **Selection of Review Panel members** according to agreed profiles: researchers with a very high academic profile and extensive experience as research managers and/or Review Panel members, e.g., heads of laboratories, institutes or departments, members of international committees and panels. Gender balance and fair geographic distribution of panel members were additional criteria for constituting the panels; they did not, however, have priority over quality of candidates.

  Selection of external referees according to agreed profiles: researchers with a high level of expertise closely matched to the field of the application and some significant research management experience.

  The standards applied by ESF in the selection of the evaluators are supported by a bibliometric analysis of profiles of Review Panel members for four of the panels: Panel 1 (Exact Sciences); Panel 2 (Engineering Sciences); Panel 3 (Health and Life Sciences) and Panel 4 (Natural and Environmental Sciences). The analysis was limited to these panels because bibliometric information has more relevance to the disciplines they cover compared to humanities or social sciences, while covering 41% of all R&D Units evaluated (133 units). This analysis can be considered as a good proxy to illustrate the principles guiding the appointment of evaluators. On average, the 41 members of Review Panels 1 to 4 have an h-index of 28 according to Scopus records (as of 24 April 2015). On average, the coordinators of the R&D Units assessed by these panel members have an h-index of 17.3. A similar analysis for external referees shows the same picture, i.e., on average, the 266 external referees involved in assessing proposals from Review Panels 1 to 4 have an h-index of 24.9, still higher than an h-index of 17.3 of the coordinators. In fact, even when broken down to the level of disciplines and sub disciplines, it can be shown that Review Panel members and external referees have a higher h-index than the average of unit coordinators in the vast majority of cases. In only three cases this could not be achieved due to the very high profile of the coordinator. However, these units have not been impacted by this gap as all three have been visited and have been classified excellent (2) and exceptional (1). Detailed statistics are available in Annex 3.

- Following the **Conflict of Interest** issue in a strict way: ESF staff checked in each case that candidates for external referees and for Review Panel members had no conflict of interest (no history of collaboration with the Portuguese R&D Unit coordinator and main staff, no common publications, projects, etc.). Each panel member was also asked to sign a Conflict of Interest declaration provided in the FCT documentation. Usually panel chairs brought the issue up at the beginning of the panel meeting and any cases of doubt were documented in CoI Management reports submitted to FCT. Detailed CoI guidelines are available in Annex 2.

  It was an explicit recommendation of the FCT and ESF to avoid any bias for or against the focus on Portuguese culture and its specificities, relevant especially in humanities and social sciences. This created an additional challenge for Panel 6 (Humanities) to appoint

an expert with knowledge of Lusophone studies who at the same time has no recent connections or collaborations in Portugal. This was solved by appointing specialists in comparative literature and translation studies who ensured knowledge and understanding of national languages and cultures and their study.

- **Checking of external reviews:** ESF peer review standards include checking the quality of the outputs from reviewers before these are circulated to the units and panels. As already mentioned, each review is checked for quality on the basis of four specific criteria.
- **Checking of consensus reports:** At the end of Stage 1, all consensus reports prepared by the lead rapporteur and the secondary rapporteur were checked for consistency with regards to the panel meeting deliberations and the above mentioned criteria by the ESF secretary to the Review Panel in collaboration with the panel chair.
- **Collegiality of scoring and grading decisions:** the evaluation of all applications took place in plenary meetings of six Review Panels, each covering a wide domain of disciplines instead of four-person workgroups focusing each on a small number of applications from one specialised field. This collegiality allowed specificities to be buffered and interests of individual research fields to be grouped into larger research domains. It also normalised the marks given by external referees (who each worked with a different framework of value and typically only reviewed one application).
- **Ensuring four assessments per application,** combining in this way specialist expertise of two external referees and a wider perspective on the domain from two rapporteurs (Review Panel members) constituted another quality assurance step.

Additional quality control steps safeguarding the fairness of the evaluation process were:

- **Applicants' rebuttal** as it allowed any factual mistakes or misunderstandings included in the assessment reports to be clarified and addressed.
- Although this can be considered as impacting the level of collegiality (as not all panel members were involved), another layer of quality control was provided by the **two appeal procedures** that allowed, at each stage, the conclusions of the Review Panels to be reconsidered based on factual information and comments provided by the R&D Units.
- Finally, **site visits** provided another opportunity to verify assessments made by Review Panels.

Strong points, areas of improvement and recommendations for the future

Overall assessment of the evaluation

In the opinion of ESF and the Review Panels, the evaluation process worked well despite its complexity.

As one of the Review Panel chairs put it in their final report:
"The overall approach to the evaluation was clear, systematic and very well planned. The ESF team were excellent in both their development and support of the overall methodology. The methodology itself is highly likely to be used by other countries in the future as its approach is broader and more comprehensive than many existing research excellence methodologies."

The results should be considered solid and trustworthy, and should help FCT to achieve goals set out in its mission:

"FCT supports, funds and assesses the brightest minds, ground-breaking ideas and internationally competitive research centres. FCT aims to create a talent-base of researchers through sustainable advanced training and science careers of excellence; foster international competitiveness and visibility of scientific research and innovation carried out in Portugal; encourage knowledge transfer between R&D centres and businesses; allow access of the scientific community to state-of-the-art infrastructures and support the development of internationally leading research centres."[9]

Joint management of the evaluation by a European organisation such as ESF and a national funding body such as FCT could be considered an advantage as it brought together knowledge and experience of the European scene and the national context. On the practical side, it worked well although some perceived it initially as potentially confusing that two different organisations implemented different parts of the process. Excellent communication between the two bodies throughout the evaluation is key to success.

Strong points

The general structure of the evaluation was good, with a good balance between remote assessment of R&D Units, discussion in Review Panel meetings at Stage 1, the site visits and Stage 2 Review Panel meetings. This structure allowed a continuum from targeted/specialist subject-specific evaluation to consensual collegial domain-specific ranking at scientific discipline level while integrating feedback from the applicants. As noted above, the fact that larger groups of experts assessed disciplinary groups of applications increased objectivity and fairness of the outcomes.

[9] *Evaluation Guide, FCT Evaluation of R&D Units 2013*, p.4

The Review Panels also appreciated **the balance between evaluating both past performance and promised future performance**, noting that many other research excellence evaluations (e.g., UK RAE/REF) have focused mainly on the past and pay more limited attention to the future. FCT explicitly pointed to the orientation of the evaluation more toward future strategy of R&D Units (strategic programmes) and corresponding capacities and structural developments than to past achievements through the structure of the application and the evaluation criteria as presented in the *Evaluation Guide*.

At Stage 1, criteria A, B, C and D applied:

- Criterion A was clearly relevant to the past;
- Criterion B was relevant to both the past (what the team had delivered?) and the future (is it able to deliver what is planned? – this included management aspects);
- Criteria C and D were clearly relevant to the future.

With all criteria having the same weight, one can see that the strategic programming had a higher weight in the assessment than past performance.

At Stage 2, the impact criterion (E) was introduced, with a different weight depending on the profile of the research (to be) performed. Criterion E was also relevant to both past performance and future plans – the relative importance of the strategic programme with regards to past performance was preserved. This new approach of FCT was not always well understood by applicants and as a result some units with very good past performance but a weaker strategic programme did not proceed to Stage 2 while others with more modest past achievements but a more convincing strategic programme succeeded.

Quality and commitment of panel members was exceptional. This was partly due to the fact that ESF has privileged access to a wide research community which trusts ESF process, procedures and ethos/ethics. Diversity of experiences, approaches and perspectives among evaluators was ensured: if Review Panel members and external referees are included, 659 international experts from 46 countries took part in the evaluation process.
ESF applied stringent **quality assurance** at all stages under its control: selection of referees and panel members according to high standards; checking of external reviews and consensus reports; following the Conflict of Interest issue in a strict way; collegiality of the evaluation.

Evaluation procedures were meticulously planned down to detailed instructions like panel meeting format; main points in consensus reports, etc. Panels were supported by professional and experienced ESF (Stage 1) and FCT (Stage 2) secretaries. Furthermore, experienced administrative staff were dedicated to this activity.

IT infrastructure supporting the process and provided by ESF was sound and reliable. ESF staff responsible for this infrastructure were constantly available, reacting very quickly to all inquiries or problems.

Areas of improvement and recommendations for the future

The panels and observers pointed out the following issues for consideration in future evaluations:

Application forms

The main recommendation concerns application forms. All Review Panels pointed out that forms need to be changed. The main criticism can be summarised as follows:

- Panels strongly suggested simplification of the form and its limitation in size (setting a maximum number of words for the whole text and its sections) to make it 'user-friendly': easier for applicants to prepare and at the same time easier for evaluators to read and find necessary information. What panels were looking for was a clear, specific and convincing overall picture of past achievements and future plans.
- The forms were very detailed and had too many sections, the differences between which were not always clear either to applicants or to external rapporteurs and panel members. The applications were thus very long (sometimes more than 100 pages and up to 300 pages in extreme cases), they often included numerous repetitions and this made it hard to focus on essential information. FCT should be more specific about the information it wanted the applicants to supply, and in what form; and, equally importantly, what it does not want applicants to include, at least not in the main text (annexes are a good solution in such cases).
- The evaluation criteria and the content of the form should be more closely aligned, as sometimes it was necessary to search in different places in order to find details related to a single criterion.
- Encouraging the use of figures could help drastically reduce the repetition of material across themes and groups and shorten and improve the text.
- The core concepts used in the forms (e.g. 'research group' and 'thematic line') should be better defined as in some cases they were interpreted quite differently in different applications; in other cases they were not clear to the applicants themselves. This also led to a lot of repetition.
- Applicants should overall have made a greater effort to articulate the added value accruing to their research as a result of interaction and collaboration within the R&D Units.
- In order to judge productivity, information on numbers of staff should be based on actual research time (in Full Time Equivalents - FTE), preferably broken down according to number of researchers in different career stages (post-doc; junior; senior). As it is now, large numbers of staff are listed but, since they spent only a part of their time on research in the centre (up to 30% of FTE according to information provided to panels), it is difficult to evaluate the efficiency of the unit.
- Moreover, the presentation of publications of Integrated Members (IMs) and how to relate them to their career stage should be reconsidered. The wide inclusion of younger, 'non-tenured' staff publications makes lists of publications longer but, paradoxically, obscures the performance of the real group leaders in the units, and it also makes it tough to tease out the number of publications and other outputs that are

shared. This needs some attention, because units might be unfairly penalised for a long tail of apparently semi-productive IMs when in fact that tail represents very young staff who are recent PhD graduates with just the right number of papers to qualify for inclusion.

- When listing relevant publications with more than one author it was important to ascertain *where* the work was done, mainly in the Portuguese R&D Unit under evaluation, or in the collaborator's institution. This could be clarified by insisting that FCT is acknowledged in a publication, citing the grant number.
- It might have been helpful if the form was formatted for a particular font, type size and line spacing, as some were quite difficult to read.

The site visits confirmed some of those observations and strengthened the feeling that the forms did not facilitate a clear, consistent and compact presentation of a unit's real achievements and plans. In several cases, presentations made by unit coordinators during the visits were much more informative than the application submitted. The coordinators and panel members alike noted that, on one hand, there was not enough space in the application for issues like a presentation of research topics while, on the other hand, many sections required repetition of the same information.

It would be helpful if the guidelines for applicants are more explicit in how applications should be prepared. Providing training in writing applications and providing successful applications as models ('best practice' example) could be helpful as well.

All Review Panel chairs provided specific advice and many detailed recommendations on how to improve the forms and the preparation of applications in their reports to FCT and one panel member even provided a concrete draft proposal (attached as an annex to the chair's report). They also suggested for the future that it would be very helpful for the Review Panels to have access to *all* the guidance given to applicants when they were completing the application forms.

Strategic planning

As already stressed, the request to applicant R&D Units to submit for evaluation and funding their strategic programmes for the next funding period was, in the opinion of Review Panels, an excellent approach. However, they observed that R&D Units and their managers should be better prepared for presenting their future plans. The units often seemed unclear as to the meaning of the term 'strategic programme'. Too often there was confusion between 'management' and 'strategy' and not sufficient attention was paid to the importance of governance and management structures and procedures for the success of the outlined strategy. Also, strategic programmes tended to be all-inclusive, overambitious and often couched in vague terms without a concrete and realistic indication of the resources required to carry them out and the means by which goals would be achieved. The panels were looking for manageable activities in respect of well-defined projects, depending on available resources. As the first step, SWOT (Strength, Weaknesses, Opportunities and Threats) analysis should become a standard part of the strategic programme. It was suggested that more training in developing SWOT analysis and strategic programmes would be helpful.

The programmes as presented too often left the panels uncertain as to what the units would actually do in the future which impacted on the evaluation results and prohibited some units with very good past performance from qualifying to Stage 2.

Evaluating budgets

The Review Panels found it very difficult to judge budgetary information. Evaluating budgets of institutions requires substantial knowledge of the national funding system and of the legal and financial national context. Budgets were not presented in a clear and consistent way. Moreover, there was an enormous range in the budgets that applicants were requesting, not always related to the size of the unit. It became clear only late in the evaluation process that this was a deliberate rule set out by FCT not to limit to requested funding. This, however, increased differences in the final funding levels which in the opinion of the panels were unjustified (see chapter 'FCT funding decisions').

All this made evaluating budgets very complex. In the opinion of panels, either in the future this task should be made simple with a clear, standardised form of presenting budgets and a set of specific questions to the panel, explained well at the beginning, or should not be requested from panels of foreign researchers. This is all the more important as criterion D, of which the budget issue was a part, was given a very important role in the final funding decision.

Evaluating interdisciplinary applications

Evaluating interdisciplinary applications poses a challenge. Panels do not question the idea of setting up a multidisciplinary panel for this purpose. This is a solution often used in other evaluation exercises as well. They felt, however, that it must be better defined which units are evaluated as multi- or interdisciplinary. Applicants indicating as their panel the multidisciplinary panel should justify in the application that their unit has a multidisciplinary character. This requires defining criteria for multidisciplinarity. The choice of the 'discipline' (panel) to apply to should also be better taken into account to avoid some typically multidisciplinary units being reviewed only by a disciplinary panel, while monodisciplinary units apply for evaluation under a multidisciplinary panel. In some evaluation procedures the designation of the discipline (panel) by applicants is verified at the beginning of the evaluation process by the panel or the supporting staff (this was the case in ESF procedures of grant proposal evaluation).

Site Visits

The use of site visits merits further consideration. As already mentioned, the Review Panels were clear that these were extremely valuable, allowing assessments made on the basis of paper documentation in Stage 1 to be verified. Although they recognised that it would not be cost effective and not feasible to visit every unit, panel members wonder if more site visits could have been made to units on the borderline between grades 'good' and 'very good', with a view either to raising them or to reaching a better understanding of what that threshold of 15/20 applied for qualification to Stage 2 evaluation actually signifies.

Logistics of Site Visits

The complexity of logistics of site visits was underestimated by FCT and ESF. Taking into consideration that panel members are high level researchers and research managers with schedules planned well ahead, it was unrealistic to expect sufficient flexibility on their part at

very short notice (visits could be planned in June only, after the results of Stage 1 were known). As a result it was not always possible to group specialists from a given group of research disciplines in one site visit session. In 15% of cases, neither the Stage 1 lead rapporteur nor secondary rapporteur for a given application could take part in the visit. Ensuring their participation proved especially difficult in the case of Multidisciplinary Panel 7 as visits to those units required the parallel presence in Portugal of evaluators from different panels. FCT took special care to coordinate the teams in a way which would ensure specialised expertise and continuity but they were not successful in all cases. In those cases other evaluators present in Portugal at the time replaced Stage 1 rapporteurs. At the final consensus meeting, the Multidisciplinary Panel 7 chair made sure that all evaluators involved in the evaluation of a given R&D Unit, rapporteurs of Stage 1 and rapporteurs of the site visit were included in the discussion and in delivering the consensus report.

Grade 'exceptional'

Some panels pointed out that the grade 'exceptional' would require more of a steer on what FCT wanted in terms of definitions of exceptionality and excellence, otherwise different panels could have used different approaches. Some perhaps went for exceptionality as what might be achieved by the best centres in their field in Portugal given the state of scientific development in Portugal at the time, others went for exceptional by world standards and this variation may give a misleading picture of the state of some fields. In terms of outcomes the number of units graded as exceptional diverged between 0 and 3 per panel; however, in terms of percentage the spread was 0-8%. One practical suggestion could be to adapt a range of marks for 'exceptional', e.g., from 24 to 25, with 'excellent' from 22 to 23.75, and a smaller range for 'very good' (18.75 to 22.75) to facilitate panels' decisions. It might also be recommended to create an opportunity for the panel chairs to confer on calibrating exceptional units before their panels made final decisions on this category.

Contextual information

Panel members felt that they could have been provided with more in the way of an introduction to the Portuguese research and higher education environment. They would have benefited also from a more detailed explanation of the background, objectives and rationale behind the evaluation at the outset; this was provided later in the process, e.g., in presentations by the FCT President and Vice-President preceding site visits and at the start of the Stage 2 Review Panel consensus meeting and was excellent but would have been welcome earlier in the process. This could also be done in the form of supporting documentation explaining some specificities of the Portuguese system (e.g., the role of host universities toward R&D Units, or what constitutes a doctoral programme).

Review Panels' strategic concerns

FCT funding decisions

Although panels felt responsible only for the scientific evaluation of R&D Unit applications they were well aware that its outcomes translated into funding for the units. They expressed their concern to the FCT leadership that rules for funding were difficult to comprehend and lacked

clarity. They would have valued more explanations from FCT and a clearer picture of what could be expected in the final results in terms of funding from the agreed grades. Some panel members thought that the methodology/formula used for the final weighting was not sufficiently explained to them, nor were the effects derived from each of the scores clearly explained. Thus, for example, in some cases some minor differences in grade seemingly resulted in very significant differences in the average *per capita* funding that units were offered. In the end, some units will benefit financially comparatively more than expected by panel members and, conversely, others are going to be impacted comparatively much more than panel members had expected, preventing them from achieving the objectives of their strategic programmes. The situation was intensified in their opinion by the fact that the funding rules for strategic funding were not available before panel grades were finalised, i.e., during the evaluation process.

Recommendations for units below 'very good' grade

Panel members expressed their concern about R&D Units which did not pass to Stage 2, especially those with scores just below 15/20 at Stage 1. They suggested that those units are offered funding for reorganisation as there is hope that they will be able to improve significantly. The choice of centres able to apply for such funds should be carefully considered and should depend on scores for individual criteria and not just on overall grades.

Recommendations on national strategy and policy

In their deliberations over individual applications, panel members, selected also for their policy expertise, came up with recommendations regarding national research strategy and policy. To provide only a few examples:

- External pressure to merge several centres should be exerted with care. In many cases, in the view of panel members, the mergers don't seem rational, resulting from a real need of the merged units, but rather a result of external recommendations or pressures.
- The degree of collaboration between centres in the same field at a national level is too low or their distribution over the country is uneven. There is a need for strategic decisions in this area.
- Panel members noticed in some cases that all centres in a given field are rated low and there is a need for a strategic approach to improve a whole discipline at the national level, e.g., dedicated funding for a restructuring plan.

Such observations and suggestions do not really belong to a report on the evaluation of individual units. Also, they should not always be addressed to FCT as some decisions may belong to other instances, e.g., the Ministry. It is recommended that in future a special channel of communication for the strategy and policy recommendations is created for international evaluators, e.g., the possibility to submit a separate report with strategic recommendations.

Some aspects of panels' work

The evaluators were very committed and worked with enthusiasm, professionalism and efficiency. However, in their final reports, they pointed out some weaknesses of the evaluation exercise impinging on their work:

- They noted that the evaluation process was overall too long and complex, including too many steps. They indicated specifically their participation in the formal appeal process which they were not informed about at the beginning and which constituted a heavy additional workload. ESF admits that integrating the appeal process, the magnitude of which it was not aware of until June 2014, was a real challenge and contributed significantly to the already tight timeline and the workload of the evaluators.
- They also considered that panels should not be involved in the Smart Specialisation Strategy question as they didn't feel experts on EU and national policies and regulations related to EU structural funds.
- Stressing that guidelines in general were clear and adequate they also noted that there were probably too many of them and sometimes the same document was distributed in different, slightly changed versions more than once.

Last but not least, taking into consideration that such exercises often raise some concerns in the community, in the opinion of Review Panels, their membership should only be made public after closure of the review process, not in the midst of the review. Having their identity disclosed before the end of Stage 1 appeals exposed panel members to direct contact with members from units being evaluated and to personal attacks, some being very virulent. This complicated the achievement of their tasks. It has to be noted, however, that although this was very unpleasant and impacted the serenity of the process, such behaviour was marginal as it came from a very limited number of individuals.

Additional issues

There are a number of additional issues that have raised some concern for a part of the Portuguese scientific community. It seems important to address these in this document.

Did the rules change during the process?

As mentioned above in this report, when discussing the potential involvement of ESF in support of FCT, it appeared clear that the structure and process laid out in the *Evaluation Guide* lacked collegiality at both stages (32-37 workgroups of four members looking each at approximately ten applications without further integration at higher disciplinary level) and could also benefit from stronger interactions with R&D Units at Stage 1 (no feedback foreseen from the latter before the shortlisting of units going to Stage 2) and at Stage 2 (no provision for panels to point out specific issues to be addressed during the site visits).

The structure eventually implemented was adapted in order to take these potential improvements of the process into account. **It, however, respected the boundary conditions expressed in the *Evaluation Guide*** (the 'workgroups' were eventually composed of two external referees and two Review Panel members). The structure and process effectively implemented was presented to the community in *Evaluation Guide - Additional information* (29 April 2014[10]).

The structure implemented allowed a continuum from targeted/specialist subject-specific evaluation to consensual collegial domain-specific ranking at scientific discipline level while integrating feedback from the applicants. The fact that larger groups of experts assessed disciplinary groups of applications also allowed individual behaviours in the panels to be buffered and normalised the marks given by external referees (who each worked with a different framework of value and typically only reviewed one application).

The main adaptations made are listed in Annex 1.

Was it agreed that 50% of all units would not proceed to Stage 2?

It should be made clear here that it was never agreed between ESF and FCT that only 50% of R&D Units would proceed to Stage 2. It was, however, estimated that approximately 50% of the units would proceed to Stage 2. The '50%' figure was actually an estimate allowing to scale and scope the effort required for Stage 2 for contractual purposes and to get a better picture of the whole process. For these reasons it was mentioned in the contractual document. However, it was awkwardly written so that it could be interpreted as a hard fact.

This estimate was based on past experience. This 50% figure was based on the outcome from the previous (2007) evaluation exercise. Final figures of the 2007 evaluation provided by FCT are in Table 13 below:

[10] https://www.fct.pt/apoios/unidades/avaliacoes/2013/docs/EvaluationGuide_Additional_Info.pdf

Table 13. Distribution of R&D Units falling into the five categories during the 2007 evaluation exercise.

Category	Number of units	%
Poor	7	2%
Fair	58	15%
Good	112	30%
Very Good	139	37%
Excellent	62	16%
Total	***378***	***100%***

Out of 378 units evaluated, 53% were classified 'very good' and 'excellent'.

Notes:

- *the 2007 evaluation did not use an 'exceptional' category*
- *these figures do not cover the Associated Laboratories.*

This estimate was a good approximation. When looking at the 966 assessments (three per unit) provided totally independently (i.e., without any interaction, consultation or normalisation between the three evaluators) by the two external referees and the secondary rapporteurs before the Stage 1 Review Panel meetings in Amsterdam, 53% of the assessments called for 'proceeding to Stage 2' (i.e., provided marks for a classification 'very good' and above).

This estimate was not a contractually binding obligation. After Stage 1 (including appeals), 55.3% of R&D Units eventually proceeded to Stage 2 (178 out of 322 units), which is consistent with the 2007 results and **is undoubtedly and unequivocally higher than an arbitrary 50%** cut off. Furthermore, one can see that the share of R&D Units proceeding to Stage 2 varies significantly between panels (between 51% for social sciences and 63% for natural sciences).

Did ESF choose reviewers with non-matching and/or weak competence?

The process implemented by ESF aimed at identifying and mobilising specific expertise required to assess each of the R&D Unit applications. This is true for remote experts (targeted specialists sometimes with unique and/or complementary expertise) as well as for Review Panel members (senior enough to allow a broad view of the domains covered by their panel combined with strong experience in research management). Comparing disciplinary expertise of the Review Panel members with thematic focus of units under their evaluation allows us to conclude that scientific disciplines were adequately covered and represented to ensure professional and fair evaluation. In addition to being internationally renowned scientists, Review Panel members have at the same time been active as heads of departments of established universities, chairs and members of ERC and national panels and members of international expert and advisory committees.

The analysis provided in Annex 4 illustrates the fact that both the Review Panel members and remote experts had an appropriate academic level to review the R&D Unit applications.

Concluding remarks

The evaluation of the FCT Research Units performed in 2014 was a complex and challenging initiative that required adaptability from the actors involved in the process, including the investigators associated with the R&D Units.

The commitment and professionalism of Review Panel members was exemplary throughout the process; they all provided a thorough, balanced evaluation fully independent with regards to any of the stakeholders with an interest in this process, whether institutional or individual.

It can be considered that the integrated outcome of the evaluation process as presented by i) the final classification of research units, and ii) the consensus reports produced by the Review Panels (complemented by external assessment reports and in some cases reports from the appeal) is robust, balanced and fair. This consensual outcome results from a well-balanced, coherent, consistent and independent process involving open and free discussions at panel level with the support of external reviews and, in some cases, site visits.

Glossary

AL:	Associated Laboratory
CoI:	Conflict of Interests
ERA-Net:	European Research Area Net
ESF:	European Science Foundation
EUROPOLAR:	The European Polar Consortium
FCT:	Fundação para a Ciência e a Tecnologia (Foundation for Science and Technology)
FIRB:	Futuro in Ricerca – Bando (Call "Future in Research", IT)
GRAPHENE:	Future and Emerging Technology (FET) Flagship on Graphene
HERA:	Humanities in the European Research Area
IRCSET:	Irish Research Council for Science, Engineering and Technology
PRIN:	Progetti di Ricerca di Interesse Nazionale (Research Proposals of National Interest, IT)
R&D Units:	Research and Development Units
RAE:	Research Assessment Exercise (UK)
REF:	Research Excellence Framework (UK), *replaced RAE in 2014*
SWOT:	Strengths, Weaknesses, Opportunities and Threats

Annexes

Annex 1: Adaptation of the evaluation process details

Annex 2: Conflict of Interest guidelines

Annex 3: Detailed tables and statistics

Annex 4: Detailed statistics for bibliometric analysis of profiles of members of Review Panels 1 to 4 and external experts involved in the assessment of applications considered by Review Panels 1 to 4

Annex 1: **Adaptation of the evaluation process details**

The table below describes the adaptations made to the process announced by FCT in the document *EVALUATION GUIDE - FCT Evaluation of R&D Units 2013* (31 July 2013)

Evaluation Panels and Working Groups		
From the Evaluation Guide text	**Structure eventually implemented**	**Comments**
Applicants may propose up to three names of experts whom they consider to be qualified to assess the application.		Not all research units did suggest three names of experts. As for those suggested, there was a significant number of conflicts of interest and/or experts not at the right standard.
All applications will be subjected to scientific evaluation distributed by four panels, which are responsible for the preliminary remote reviewing of all applications. This distribution is in accordance with the four major scientific domains under the aegis of the Scientific Councils of FCT.	Six disciplinary panels and one interdisciplinary panel (composed of members of the six disciplinary panels) were set up. Disciplinary coverage of panel members was appropriate. Gender balance was sought at equal level of expertise.	The high number of applications eventually submitted required the 'Humanities and Social Sciences' panel to be split into two panels, 'Humanities' and 'Social Sciences'. Similar issue for the 'Exact Sciences and Engineering' panel, split into 'Exact Sciences' and 'Engineering'. The evaluation guide did not plan any physical meetings of review panels.
The constitution of the evaluation panels will take into consideration the number of applications for each scientific domain, a good gender balance as well as a fair geographic and institutional distribution of evaluators. The composition of the evaluation panels will be published in the FCT website.		
The members of each panel will, in turn, be distributed by several workgroups of four elements each. Every workgroup will be responsible for the remote assessment of about 10 proposals in a given scientific area.	The seven Review Panels (bringing together 74 individuals) met in plenary sessions twice. They reviewed all the applications submitted to their panel (between 24 and 74 applications).	With 322 applications (of which 43 were interdisciplinary), the structure mentioned in the *Evaluation Guide* would have resulted in a maximum number of approx. 145-155 individual review panel members (if no member participated in more than one 'workgroup'). The seven panels set up by ESF brought together 74 individuals of whom 18 were also members of the interdisciplinary panel. The overall number of panel members was therefore 92 (74+18). This number was, however, augmented by the remote reviewers.

From the Evaluation Guide text	Structure eventually implemented	Comments
All workgroups members should consider possible conflicts of interest and observe the confidentiality statements (see section 5. 'Confidentiality and Conflicts of Interest').	Conflicts of interest were dealt with and documented in an appropriate way.	
One member of each workgroup will be designated the coordinator of the workgroup. Optimally the workgroup coordinator will also be a member of one of the evaluation panels of the 2nd stage of the evaluation.	Review Panels were each chaired by one of their members. Chairs participated in a high number of Stage 2 site visits.	Stage 2 site visits were organised and implemented by FCT.
R&D Units that have explicitly indicated an interdisciplinary profile in the application form will be remotely reviewed by more than one workgroup.	R&D Units that explicitly indicated an interdisciplinary profile were assessed by the multidisciplinary review panel composed of members from all disciplinary panels.	The structure proposed originally would have led to several groups looking at the same application, possibly resulting in several differing positions with no plan to resolve the issue.
Individual Reviews		
From the Evaluation Guide text	**Structure eventually implemented**	**Comments**
Up to 5 individual reviews will be remotely prepared for each application, according to the evaluation criteria of the 1st stage of the evaluation (see section 3. 'Evaluation Criteria and Scoring System').	Four reviews were provided for each application and discussed during the first Review Panel meeting: two external (mail) referees (specialists in the topic addressed in the application) were identified and provided an assessment; two Review Panel members provided an assessment, the secondary rapporteur independently and the lead rapporteur with a deeper level of information, integrating both his/her own evaluation and the assessment reports from the two external referees and the secondary rapporteur.	Including Review Panel members and external referees, 659 international experts from 46 countries took part in the evaluation process. Having two specialist external referees allowed to stick very closely to any given proposal subject, bringing in complementary expertise and specific expertise that may have been relevant to only one (or a very limited number of) application(s). Having two specialist external referees also allowed the geographical scope of the evaluation to be expanded (panel members were primarily from European countries).
All 4 members of the workgroup will remotely elaborate an individual review for each one of the 10 proposals assigned to the workgroup.	Each of the Review Panel members had on average 8.8 applications to review as lead or secondary rapporteur. All applications given to a Review Panel were discussed by all panel members.	

From the Evaluation Guide text	Structure eventually implemented	Comments
One of the three experts indicated by each R&D Unit will be invited by FCT to elaborate an individual review.	A limited number of experts indicated by the Research Units provided a review as external referee	Not all research units provided names of potential referees; some provided one or two – 50 out of the 322 provided none. In a large number of cases, these potential referees were not appropriate as they had a conflict of interest and/or their profile did not match the ESF quality standard. For fairness and equality of treatment it was decided that all applications would be assessed by the same number of individuals. Experts suggested by units were invited to provide an assessment only when considered relevant.
Both workgroup members and experts must submit their individual reviews for each proposal in the Individual Reviewer Evaluation Form and lock them.	Done as planned but on the ESF platform.	
The coordinator of the workgroup will be in charge of arbitrating the discussions of each application and the corresponding consensus reports.	All applications were discussed by all panel members during a physical meeting, arbitration was made by consensus, taking into account all applications assessed by the panel.	The adapted structure allowed a higher level of collegiality and a 'normalised' perception of marks and classification (e.g., 'good' or 4 out of 5 would naturally not have the same meaning for all 30+ workgroups originally planned).
Results and Rebuttal		
From the Evaluation Guide text	**Structure eventually implemented**	**Comments**
All applicants will receive the consensus reports comments, regardless of being pre-selected or not to the 2nd stage of the evaluation process. The candidates whose applications will not be selected for the 2nd stage of the evaluation will also receive the individual ratings attributed to each evaluation criteria in their corresponding consensus report.	Done as detailed but processed through ESF platform. A specific field was added to the consensus reports: 'specific points to be addressed during the site visits'. This allowed the units to be aware of and subsequently address the points of specific concern for the panel during visits.	
The R&D Units whose applications are selected for the 2nd stage will only receive their qualitative overall grading at the end of the whole evaluation process. In order to be selected for the 2nd stage of the evaluation process, an application must: *[conditions for Stage 2]*	Done as detailed but processed through ESF platform.	

From the Evaluation Guide text	Structure eventually implemented	Comments
After the end of this 1st stage of the evaluation exercise, all applicants will have the opportunity to prepare responses to the assessments and comments contained in the received consensus reports (rebuttal phase). When applicable, these responses should be taken into account by the evaluation committees of the 2nd stage of the evaluation.	The rebuttal phase was implemented before the first panel meeting and deliberations, allowing research units to provide their feedback and clarifications on assessments before any discussion was made. Research units had access to three assessments (from the two external referees and the secondary rapporteur) and could provide their feedback online. This feedback was made available to the Review Panels in preparation of the first Review Panel meetings (and not only for the second stage).	
In accordance with the Portuguese law, the candidates will also have the right to submit a prior hearing, within 10 days after notification of the results, which should be answered before the beginning of the 2nd stage of the evaluation process.	Prior hearing (formal appeal) was performed as planned and coordinated by FCT.	
Second Stage		
From the Evaluation Guide text	**Structure eventually implemented**	**Comments**
One of the evaluation panel members will be designated by the chair to elaborate the final report for each R&D Unit; according to the evaluation criteria of the 2nd stage (see section 3. 'Evaluation Criteria and Scoring System').	Done as detailed but processed through ESF platform.	Second stage was implemented and coordinated by FCT, ESF was observer to the process.
If considered necessary, the general observer for the 2nd stage of the evaluation can still decide to schedule a final meeting with all the evaluation panel chairs in order to validate and ensure the consistency of the qualitative overall grading of all the assessed R&D Units.	All seven panels reconvened physically to discuss all units visited and reach consensus on a final classification of the research units.	Allowed a better level of collegiality and integration.

Annex 2: **Conflict of Interest guidelines**

All Review Panel members and external referees conformed to the following Conflict of Interest guidelines (from FCT's *Evaluation Guide*).

Circumstances that could be interpreted as a **disqualifying conflict of interest** are laid down in the following criteria:

1. First-degree relationship, marriage, life partnership, domestic partnership;
2. Personal interest in the application's success or financial interest by persons listed under no.1;
3. Current or planned close scientific cooperation;
4. Dependent employment relationship extending five years beyond the conclusion of the relationship;
5. The affiliation or pending transfer to the research unit or to a participating institution;
6. Researchers who are active in a council or similar supervisory board of the applying institution are excluded from participating in the review and decision-making process for applications originating from this institution;

A potential conflict of interest may exist, even in cases not covered by the clear disqualifying conflicts indicated above, in the following circumstances:

7. Relationships that do not fall under no. 1, other personal ties or conflicts;
8. Financial interests of persons listed under no. 7;
9. Participation in university bodies other than those listed under no. 6, e.g. in scientific advisory committees in the research environment;
10. Research cooperation within the last three years, e.g. joint publications;
11. Preparation of an application or implementation of a project with a closely related research topic (competition);
12. Participating in an on-going scientific or inter-personal conflict with the applicant(s).

Before starting the evaluation of each application, and in order to be able to access the evaluation form, experts completed a CoI Declaration.

During panel meetings, panel members were asked to inform the panel secretary of any potential CoI; in such cases, adequate measures were taken.

Annex 3: **Detailed tables and statistics**

Table 14. Panel Members' geographical representation per panel. Numbers in red indicate the country of the chair.

	P01	P02	P03	P04	P05	P06	Total
AT					1		1
BE		**2**	1			1	4
CH						1	1
CY					1		1
CZ	1						1
DE	1	1	2		1		5
DK		1			1	1	3
EE					1		1
ES	1				2		3
FI		1			1	1	3
FR	1	2	2	1			6
GR	1						1
HR						1	1
HU						1	1
IE				1		2	3
IL			1				1
IT	1	2	1	**3**	2	2	11
LU					1		1
NL	1			2		1	4
NO	1						1
PL						1	1
SE					1		1
TR					1		1
UK	**2**	4	**2**	1	**4**	**4**	17
US	1						1
Total	**11**	**13**	**9**	**8**	**17**	**16**	**74**

Table 15. Correlation between scores attributed by the two external referees and the secondary rapporteur by panel (columns) and for each of the criteria (QA, QB, QC, QD) applicable during Stage 1. The last row presents the same correlation but calculated for the total score (/20). All correlations are significant (p<0.001).

	P01	P02	P03	P04	P05	P06	P07
QA	0.50	0.41	0.38	0.30	0.33	0.23	0.54
QB	0.30	0.31	0.51	0.22	0.35	0.34	0.55
QC	0.22	0.17	0.44	0.21	0.37	0.29	0.58
QD	0.13	0.45	0.40	0.46	0.27	0.19	0.26
Overall	0.35	0.49	0.49	0.37	0.37	0.33	0.57

Table 16. Scores for each individual criterion QA, QB, QC and QD (/5) and total score per panel at three time points during Stage 1 evaluation. 'Before': mean of raw scores; 'Consensus': mean of scores after consensus reports; 'After appeal': mean of scores taking into account scores updated due to the legal appeal procedure.

		Before	Consensus	After appeal
P01	**QA**	4.12	4.03	4.03
	QB	4.17	3.92	3.92
	QC	4.08	3.92	3.92
	QD	4.09	3.59	3.62
	Cumulated	*16.46*	*15.46*	*15.49*
P02	**QA**	3.96	3.96	4.02
	QB	3.97	3.85	3.89
	QC	3.87	3.65	3.67
	QD	3.86	3.74	3.76
	Cumulated	*15.66*	*15.20*	*15.35*
P03	**QA**	3.74	3.63	3.63
	QB	3.63	3.67	3.71
	QC	3.58	3.50	3.54
	QD	3.39	3.21	3.21
	Cumulated	*14.33*	*14.00*	*14.08*
P04	**QA**	3.94	4.00	4.00
	QB	3.81	3.92	3.96
	QC	3.74	3.67	3.75
	QD	3.53	3.71	3.71
	Cumulated	*15.01*	*15.29*	*15.42*
P05	**QA**	3.56	3.62	3.62
	QB	3.45	3.36	3.36
	QC	3.49	3.55	3.58
	QD	3.52	3.34	3.37
	Cumulated	*14.02*	*13.86*	*13.92*
P06	**QA**	3.81	3.94	3.94
	QB	3.67	3.56	3.60
	QC	3.70	3.60	3.65
	QD	3.78	3.53	3.56
	Cumulated	*14.96*	*14.63*	*14.75*
P07	**QA**	3.67	3.80	3.86
	QB	3.59	3.59	3.61
	QC	3.55	3.43	3.45
	QD	3.57	3.20	3.23
	Cumulated	*14.37*	*14.02*	*14.16*

Table 17. Number (and %) of R&D Units falling into the different qualitative categories ('Classification') after panel consensus ('Consensus') and after appeal ('After appeal') shown separately for each panel.

Panel 01: **Exact Sciences**		
Classification	**Consensus**	**After appeal**
Poor	3 (7.7%)	3 (7.7%)
Fair	4 (10.3%)	3 (7.7%)
Good	9 (23.1%)	10 (25.6%)
Proceeds to Stage 2	23 (59%)	23 (59%)

Panel 02: **Engineering Sciences**		
Classification	**Consensus**	**After appeal**
Poor	1 (2.2%)	1 (2.2%)
Fair	8 (17.4%)	4 (8.7%)
Good	12 (26.1%)	14 (30.4%)
Proceeds to Stage 2	25 (54.4%)	27 (58.7%)

Panel 03: **Health and Life Sciences**		
Classification	**Consensus**	**After appeal**
Poor	5 (20.8%)	4 (16.7%)
Fair	1 (4.2%)	2 (8.3%)
Good	5 (20.8%)	5 (20.8%)
Proceeds to Stage 2	13 (54.2%)	13 (54.2%)

Panel 04: **Natural and Environmental Sciences**		
Classification	**Consensus**	**After appeal**
Poor	2 (8.3%)	2 (8.3%)
Fair	0 (0%)	0 (0%)
Good	9 (37.5)	7 (29.2%)
Proceeds to Stage 2	13 (54.2%)	15 (62.5%)

Panel 05: **Social Sciences**		
Classification	**Consensus**	**After appeal**
Poor	9 (12.3%)	9 (12.3%)
Fair	12 (16.4%)	12 (16.4%)
Good	18 (24.7%)	16 (21.9%)
Proceeds to Stage 2	34 (46.6%)	36 (49.3%)

Panel 06: **Humanities**		
Classification	**Consensus**	**After appeal**
Poor	7 (9.7%)	6 (8.3%)
Fair	7 (9.7%)	7 (9.7%)
Good	22 (30.6%)	20 (27.8%)
Proceeds to Stage 2	36 (50%)	39 (54.2%)

Panel 07: **Multidisciplinary**		
Classification	**Consensus**	**After appeal**
Poor	8 (18.2%)	7 (15.9%)
Fair	4 (9.1%)	3 (6.8%)
Good	8 (18.2%)	9 (20.5%)
Proceeds to Stage 2	24 (54.6%)	25 (56.8%)

Table 18. Number (and %) of R&D Units falling into the different qualitative categories ('Classification') after second meeting panel consensus ('Consensus') and after second appeal ('After appeal'), per panel. Italic lines in relevant panel table highlight changes. 'Poor' is not represented as no applications were categorised below 'fair'.

Panel 01: **Exact Sciences**		
Classification	**Consensus**	**After appeal**
Fair	0	0
Good	0	0
Very good	13 (56.5%)	13 (56.5%)
Excellent	7 (30.4%)	7 (30.4%)
Exceptional	3 (13%)	3 (13%)

Panel 02: **Engineering Sciences**		
Classification	**Consensus**	**After appeal**
Fair	0	0
Good	1 (3.7%)	1 (3.7%)
Very good	16 (59.3%)	16 (59.3%)
Excellent	9 (33.3%)	9 (33.3%)
Exceptional	1 (3.7%)	1 (3.7%)

Panel 03: **Health and Life Sciences**		
Classification	**Consensus**	**After appeal**
Fair	0	0
Good	0	0
Very good	7 (53.9%)	7 (53.9%)
Excellent	4 (30.8%)	4 (30.8%)
Exceptional	2 (15.4%)	2 (15.4%)

Panel 04: **Natural and Environmental Sciences**		
Classification	**Consensus**	**After appeal**
Fair	0	0
Good	0	0
Very good	*11 (73.3%)*	*10 (66.7%)*
Excellent	*4 (26.7%)*	*5 (33.3%)*
Exceptional	0	0

Panel 05: **Social Sciences**		
Classification	**Consensus**	**After appeal**
Fair	2 (5.6%)	2 (5.6%)
Good	1 (2.8%)	1 (2.8%)
Very good	*27 (75%)*	*23 (63.9%)*
Excellent	*6 (16.7%)*	*10 (27.8%)*
Exceptional	0	0

Panel 06: **Humanities**		
Classification	**Consensus**	**After appeal**
Fair	0	0
Good	4 (10.3%)	4 (10.3%)
Very good	*18 (48.7%)*	*17 (43.6%)*
Excellent	*14 (35.9%)*	*15 (38.5%)*
Exceptional	3 (7.7%)	3 (7.7%)

Panel 07: **Multidisciplinary**		
Classification	**Consensus**	**After appeal**
Fair	0	0
Good	*3 (12%)*	*2 (8%)*
Very good	*12 (48%)*	*11 (44%)*
Excellent	*8 (32%)*	*10 (40%)*
Exceptional	2 (8%)	2 (8%)

Annex 4: **Detailed statistics for bibliometric analysis of profiles of members of Review Panels 1 to 4 and external experts involved in the assessment of applications considered by Review Panels 1 to 4**

The information provided here is based on bibliometric information extracted from SCOPUS. As bibliometric information has less relevance for humanities and social sciences, the figures provided are given for the four following panels:

- *Panel 1 - Exact Sciences*
- *Panel 2 - Engineering Sciences*
- *Panel 3 - Health and Life Sciences*
- *Panel 4 - Natural and Environmental Sciences*

This represents 41% of all research units evaluated (133), and can be considered as a reliable representation of the standard applied by ESF in the evaluation.

The information is organised by panel and by main scientific disciplines of the research units evaluated.

To put things in context, bibliometric information of panel members and external referees is compared to the same bibliometric information of the units' coordinators (considered to be at least representative of the profile of their unit's members, or even above).

The h-index (as of 24 April 2015) over the whole career is taken as the most representative bibliometric reference.

Profile of Review Panel members

On average, the 41 members of Review Panels 1 to 4 have an h-index of 28. On average, the coordinators of the units assessed by these panel members have an h-index of 17.3.

The table below (Table 19) provides more detail on the h-index of panel members serving as rapporteur (2/proposal) and of research unit coordinators, classified by the main discipline of the research units assessed as declared by the coordinators.

One can see a clear difference between the indexes of Review Panel members and unit coordinators. However, the average profile of Review Panel members does not match that of the unit coordinator in two instances:

- Materials Science and Engineering: two of the four unit coordinators have a very high h-index (45 and 48). With few units, these high values - harder to match - lift the whole discipline.
- Neurosciences, Ageing and Degenerative Diseases: one of the two unit coordinators has a very high h-index (56). With few units, this high value - harder to match - lifts the whole discipline.

Table 19. Average h-index of panel members serving as rapporteur (2/proposal) and of research unit coordinators, by main discipline of the research units assessed as declared by the coordinators, for Panel 01 to 04. Italic lines highlight cases where average h-index of panel members is lower than average h-index of research unit coordinators.

Panel	Main Discipline	N Units	h-index of Rapporteurs from Review Panels (average)	h-index of Unit Coordinators (average)
01	Chemistry	10	31.7	27.1
	Materials Science and Engineering	4	*23.0*	*31.3*
	Mathematics	12	25.0	7.1
	Nanoscience and Nanotechnology	1	23.0	22.0
	Physics	12	27.7	19.7
	Panel 1 Total	**39**	**27.3**	**18.9**
02	Bioengineering	1	35.0	8.0
	Biotechnology	2	35.0	27.5
	Chemical Engineering	4	26.6	14.3
	Civil Engineering	9	12.8	10.1
	Computer Science and Engineering	10	16.6	10.3
	Electronics and Electrical Engineering	9	17.9	12.9
	Mechanical Engineering and Engineering Systems	11	18.7	11.3
	Panel 2 Total	**46**	**18.7**	**12.0**
03	Biochemical Sciences	1	56.5	31.0
	Biomedicine	8	44.8	26.4
	Clinical Research	3	38.0	26.7
	Diagnostic, Therapeutics and Public Health	10	39.3	20.4
	Neurosciences, Ageing and Degenerative Diseases	2	*32.0*	*41.5*
	Panel 3 Total	**24**	**41.1**	**25.4**
04	Agricultural and Forestry Sciences	5	37.2	15.4
	Animal Science and Veterinarian Science	3	42.3	6.3
	Biological Sciences or Environmental Biology	6	39.2	21.3
	Environmental Sciences	3	29.3	24.7
	Geosciences	4	27.3	12.0
	Marine Sciences and Technologies	3	25.0	16.7
	Panel 4 Total	**24**	**34.2**	**16.5**

Profile of external referees

On average, the 266 remote reviewers involved in assessing proposals from panels 1 to 4 have an h-index of 24.9. On average, the coordinators of the units assessed by these panel members have an h-index of 17.3.

The table below (Table 20) provides more detail on the h-index of remote reviewers and of research unit coordinators, classified by the main discipline of the research units assessed as declared by the coordinators.

One can see a difference between the indexes of remote reviewers and unit coordinators. However, the average profile of remote reviewers does not match that of the unit coordinator in three instances:

- Biotechnologies and Environmental Sciences show a difference but it is relatively minor (in particular for biotechnology).
- As for Review Panel members and for the same reasons: Neurosciences, Ageing and Degenerative Diseases.

Table 20. Average h-index of remote reviewers (2/proposal) and of research unit coordinators, by main discipline of the research units assessed as declared by the coordinators, for Panel 01 to 04. Italic lines highlight cases where average h-index of remote reviewers is lower than average h-index of research unit coordinators.

Panel	Main Discipline	N Units	h-index of Remote Reviewers (average)	h-index of Unit Coordinators (average)
01	Chemistry	10	30.9	27.1
	Materials Science and Engineering	4	38.1	31.3
	Mathematics	12	13.9	7.1
	Nanoscience and Nanotechnology	1	26.5	22.0
	Physics	12	31.5	19.7
	Panel 1 Total	**39**	**26.5**	**18.9**
02	Bioengineering	1	38.0	8.0
	Biotechnology	2	*27.3*	*27.5*
	Chemical Engineering	4	29.4	14.3
	Civil Engineering	9	19.2	10.1
	Computer Science and Engineering	10	12.1	10.3
	Electronics and Electrical Engineering	9	18.6	12.9
	Mechanical Engineering and Engineering Systems	11	22.2	11.3
	Panel 2 Total	**46**	**19.9**	**12.0**
03	Biochemical Sciences	1	48.5	31.0
	Biomedicine	8	37.7	26.4
	Clinical Research	3	31.2	26.7
	Diagnostic, Therapeutics and Public Health	10	26.2	20.4
	Neurosciences, Ageing and Degenerative Diseases	2	*31.5*	*41.5*
	Panel 3 Total	**24**	**32.0**	**25.4**

Panel	Main Discipline	N Units	h-index of Remote Reviewers (average)	h-index of Unit Coordinators (average)
04	Agricultural and Forestry Sciences	5	19.6	15.4
	Animal Science and Veterinarian Science	3	20.2	6.3
	Biological Sciences or Environmental Biology	6	31.8	21.3
	Environmental Sciences	3	*21.0*	*24.7*
	Geosciences	4	28.1	12.0
	Marine Sciences and Technologies	3	24.2	16.7
	Panel 4 Total	**24**	**24.9**	**16.5**

Integrated profiles of all evaluators (Review Panel members and external referees)

Considering all evaluators (panel and remote), the average h-index of individuals providing assessments for applications is 26.5 (17.3 for unit coordinators). Details provided in the table below (Table 21):

- The difference in Environmental Sciences has been buffered by the combination of the remote and the review panel assessments.
- Because of the very high profile of the coordinators of three units, the difference in Materials Science and Engineering and Neurosciences, Ageing and Degenerative Diseases remains. However, these units have not been impacted by this gap as all three have been visited and have been classified 'excellent' (2) and 'exceptional' (1).

Table 21. Average h-index of all evaluators (panel and remote combined – 4 per proposal), and of research unit coordinators, by main discipline of the research units assessed as declared by the coordinators, for Panel 01 to 04. Italic lines highlight cases where average h-index of evaluators is lower than average h-index of research unit coordinators.

Panel	Main Discipline	N Units	h-index of All Evaluators (average)	h-index of Unit Coordinators (average)
01	Chemistry	10	31.3	27.1
	Materials Science and Engineering	4	*30.6*	*31.3*
	Mathematics	12	19.4	7.1
	Nanoscience and Nanotechnology	1	24.8	22.0
	Physics	12	29.6	19.7
	Panel 1 Total	**39**	**26.9**	**18.9**
02	Bioengineering	1	36.5	8.0
	Biotechnology	2	31.1	27.5
	Chemical Engineering	4	28.0	14.3
	Civil Engineering	9	16.0	10.1
	Computer Science and Engineering	10	14.3	10.3
	Electronics and Electrical Engineering	9	18.3	12.9
	Mechanical Engineering and Engineering Systems	11	20.5	11.3
	Panel 2 Total	**46**	**19.3**	**12.0**

Panel	Main Discipline	N Units	h-index of All Evaluators (average)	h-index of Unit Coordinators (average)
03	Biochemical Sciences	1	52.5	31.0
	Biomedicine	8	41.3	26.4
	Clinical Research	3	34.6	26.7
	Diagnostic, Therapeutics and Public Health	10	32.8	20.4
	Neurosciences, Ageing and Degenerative Diseases	2	*31.8*	*41.5*
	Panel 3 Total	**24**	**36.6**	**25.4**
04	Agricultural and Forestry Sciences	5	28.4	15.4
	Animal Science and Veterinarian Science	3	31.3	6.3
	Biological Sciences or Environmental Biology	6	35.5	21.3
	Environmental Sciences	3	25.2	24.7
	Geosciences	4	27.7	12.0
	Marine Sciences and Technologies	3	24.6	16.7
	Panel 4 Total	**24**	**29.5**	**16.5**