

National evaluation exercises as implementation of research policy. A comparative study of Norway and Portugal

Jon Holm¹ and Ana Ramos²

¹*jon.holm@rcn.no*

Research Council of Norway, PO box 564, 1327 Lysaker, Norway

²*ana.ramos@fct.pt*

Fundação para a Ciência e a Tecnologia (FCT), Av. D. Carlos I, 126, 1249-074 Lisboa, Portugal

This work compares the main processes of research evaluation in two countries with R&D systems at different development stages, Norway and Portugal. It aims at shedding light on the factors that determine the specificities of research evaluation in both countries and how “quality” of SSH research is conceptualized and translated into evaluation guidelines and assessments. Furthermore, it will discuss how those evaluation exercises are used as tools to implement national research policies.

National contexts in brief

In the last 30 years, “R&D units” have been the organizational base of the research system in Portugal. These units are networks of researchers with significant autonomy from HEIs (some are formed by the association of two or more research groups from different HEIs). The R&D units are subject to periodical (every 5 years) assessments. In Norway, SSH research is organized within the ordinary university units and in independent research institutes. The higher education institutions receive a rather generous basic funding amounting to 70 to 80 percent of their R&D expenditure depending on the research intensity of the unit. The most research intensive units show a larger share of externally funded projects.

Evaluation methodology and criteria in Portugal

R&D units are evaluated by panels of international experts, based on the units’ reports and activity plans, as well as on the direct contact of the evaluators with managers and researchers during on-site visits. The assessment is organized at national level and comprises all research areas. The result of the evaluation determines the amount of funding. This funding (multiannual contracts with the national funding agency, Fundação para a Ciência e a Tecnologia, FCT) does not differentiate cost levels of research activities in different areas, which must be met by other instruments, such as R&D project grants.

The system was fairly stable from its beginning in 1996 until 2013, when substantial changes were introduced, reflecting a different political orientation at that time. But peer review has always been the main methodology and, although the formulation of the evaluation criteria varied between exercises, it is possible to group them around three main purposes: (a) Scientific quality (relevance, internationalization) and productivity of the unit; (b) Working environment, management, feasibility of the proposed activities; (c) Training of (young) researchers, knowledge transfer and outreach activities.

The evaluation criteria and guidelines are the same for all scientific areas and there are no SSH-specific adaptations. The concept of scientific quality is described in the evaluation guidelines: “publications in major research journals”, “multidisciplinary and relevance to other research areas”, “international publications”, etc. .

Evaluation methodology and criteria in Norway

RCN is mandated by the government to "ensure the evaluation of Norwegian research activities" (see above). Subject specific evaluations are carried out by international peers supported by professional evaluations agencies commissioned by RCN. The role of RCN is to define evaluation criteria and processes, assure the involvement of stakeholders in the planning and organise follow up activities.

In line with its mandate, RCN performs evaluations of institutions, thematic areas and research within specific academic subjects. The national subject specific-evaluations started out as qualitative peer reviews in the late 1990'ties. The first round of evaluations were limited to single disciplines. The second round of evaluations starting with evaluation of Biology, Clinical Medicine and Health Science (2011) has seen an increase in scope and disciplinary breadth. From an evaluation of single disciplines, RCN has moved towards evaluating broader areas of research like the Humanities (2017) and the Social Sciences (2018). These recent evaluations within SSH have been pathbreaking in adding the dimensions of societal impact and interplay between research and education to the evaluation of research quality which remains the central task of international peers who are invited to perform the actual evaluation.

Comparison of criteria

In Table 1 we have matched the evaluation questions (Norway) and criteria (Portugal) used in the respective exercises. We use this comparison, as well as the reports of Humanities panels in both countries, to establish the role of evaluation exercises in policy implementation and to ascertain the factors related to "quality" in the Humanities.

In Portugal, the evaluation determines the amount of annual funding that the R&D units receive in the subsequent years, but the panelists also make a global analysis of the scientific area under evaluation. Therefore, the exercise encompasses a "formative" aspect, providing decision-makers and unit managers with an overview of each scientific area and recommendations aimed at correcting deficiencies and improving the units' performance and competitiveness.

In the 2017 exercise, R&D units in Humanities were evaluated by five panels: Architecture and Urbanism (8 units); Arts & Design (14); History & Archaeology (15); Literary Studies (13) and Philosophy (10). The most frequent concept of research quality in every panel was, by far, *internationalization*, followed by *collaboration*, *cultural impact*, *multi/interdisciplinarity* and *relevance*. One can argue that most of these "quality" concepts are transversal to all scientific areas, except for *cultural impact*, which can be considered specific to Humanities and was valued by the panelists. There are no references to bibliometrics or any other quantitative indicators in the reports. We conclude that despite the fact that there are not SSH-specific guidelines/criteria, the panels took as a starting point the general criteria, adapting them to the accepted disciplinary practices.

The national evaluations of academic subjects in Norway were conceived as formative exercises at the end of the 1990ies. The formative purpose has allowed for a certain variation in the conceptualisation of research quality to adapt to the specific practices of each discipline on the one side and developments in research policies on the other. The definition of research quality has been holistic with little discussion on how research quality should be defined and which aspects of this notion should be privileged in the assessment. The evaluation criteria (see table below) simply refers to the 'quality norms of the discipline', which indicates that it

is left to the evaluation committee – representing the discipline – to agree on what to look for in the evaluation data in order to assess the research quality in each research group.

Research groups were evaluated by eight panels: Aesthetic Studies (6 groups); Nordic Languages and Linguistics (11); Nordic and Comparative Literature (3); Modern and Classical Languages, Literatures and Area Studies (16); Archaeology, History and Cultural Studies (22); Philosophy and Studies in Science and Technology (15); Religion and Theology (12); Media Studies (11). Also in the Norwegian case the most frequent concepts of research quality across all panels were *internationalization* followed by *productivity**, *prestige of publication channels[†]*, *interdisciplinarity**, *collaboration and theoretical advancement*. Many of these concepts were referred to explicitly in the formal criteria (*) and in the grading scale used for research groups (†). The concepts not explicitly mentioned by in the formal criteria are probably more telling of the disciplinary norms brought to the table by the panels. Among these we find concepts that could be seen are more specific for SSH, like: national relevance and cross-theoretical work.

Conclusion

We can conclude that, in Portugal, the overall process of R&D units evaluation is key to implement policy: “strengthen and diversify the landscape”, quality improvement”, “not intended to reduce the number of units or to channel the available funds preferentially to a given area or to a reduced number of excellent institutions”.

In the Norwegian case an increased political attention to the societal benefits of research has influenced the overall design of the evaluation exercise through the introduction of societal impact as a separate evaluation dimension. The assessment of research quality as such still seems to be quite unaffected by this policy change. Experts are invited to base their judgement on the quality standards of their discipline. That this adaptation also takes place is indicated by the presence of humanities specific notion of quality in the assessment texts.

Table 1 - Comparison of FCT 2017 (PT) and HUMEVAL 2015 (NO) evaluation criteria and data 16042019

		Assessment questions (NO)/Standards and guidelines (PT)	
Dimension	NO	PT	
Organisation, leadership, strategy and resources	<ul style="list-style-type: none"> • How well is the group led? • Does it have a good strategy and sensible goals (albeit perhaps demanding ones)? • Does it contribute to the institution's overall goals? • Does it make good use of external research funding? • Does the institution provide adequate resources and infrastructure? • Does the group make good use of these? 	<ul style="list-style-type: none"> • R&D Units should assemble a critical capacity to successfully accomplish their objectives and to promote work environments fostering scientific creativity. • Appropriateness of objectives, strategy, plan of activities (including ethical concerns if applicable), budget, Programmatic Funding request and organization for the following five year period (2018-2022). • The configuration and organization model of the R&D Unit should be suitable to its objectives and R&D activities, and not artificially constructed with unnatural associations or size, and should not result in inappropriate dispersion or concentration of means or resources. 	
Research production and quality	<ul style="list-style-type: none"> • How good is the quality of the research overall, in relation to the quality norms of the discipline? • Is the productivity good, given the norms of the discipline? • Has the group contributed to advancing the state of the art in its discipline(s)? • Does it make good use of interdisciplinary approaches, where these are relevant? 	<ul style="list-style-type: none"> • Quality, merit, relevance and extent of internationalization of the R&D activity in 2013-2017 of the Integrated Researchers in the application (especially those with PhD), assessed by international standards, considering originality, consistency and rigor. If applicable to the nature/objectives of the R&D Unit, also the technical, cultural or artistic impact. • For this purpose consider: contributions for knowledge advancement and/or application; publications; organization of conferences, colloquia and/or seminars; patents, prototypes or products; knowledge and technology transfer; spin-offs; preservation, curation and dissemination of R&D results and data, respecting the principles and practices of Open Science; promotion of scientific and technological culture (outreach); actions of special scientific, technological, cultural, artistic, social or economic relevance to society. 	
Recruitment and training	<ul style="list-style-type: none"> • Are the group's hiring and career development practices consistent with best practice? • Are PhD candidates and post-docs adequately trained and mentored? • Is there sufficient national and international mobility of researchers? 	<ul style="list-style-type: none"> • Advanced training (PhD and Postdoc level); initiation of undergraduate or Master students to research. • Talent attraction, hiring of new researchers and scientific careers development. 	
Networking	<ul style="list-style-type: none"> • Does the group make good use of collaboration, nationally and internationally, to advance its strategy and produce high-quality, relevant research? 	<ul style="list-style-type: none"> • Extent of internationalization of the R&D activities. 	
Impact on teaching	<ul style="list-style-type: none"> • To what extent is the research of the group relevant for the study programmes at the host institution? 	<ul style="list-style-type: none"> • Support to PhD Programs. 	

Assessment questions (NO)/Standards and guidelines (PT) cont'd	
Overall	<ul style="list-style-type: none"> • What is your overall judgement of this group? • Is it a 'star' group, which should be identified as such in the panel's national assessment of the area?
Feedback	<ul style="list-style-type: none"> • Justify in detail the 3 evaluation criteria ratings and the overall grade, and provide substantive comments and recommendations regarding R&D activities and results, the team of Integrated Researchers, the objectives, strategy, plan of activities, reasonability of funding and budget, organization, ethical concerns and other aspects that may be considered relevant. Address strengths and weakness of the R&D Unit as a whole and, whenever applicable, comment on Research Groups and/or Thematic Lines. • Include references to specific cases or situations of resources, competencies and teams of exceptional quality or value that may be useful for monitoring and steering the development of the National S&T System, including: (i) the response to specific problems of public interest or to challenges faced by society, (ii) the strengthening of internationalization and the use of special opportunities of high value international cooperation or partnership, (iii) the preparation of concerted initiatives aiming at attracting resources for R&D activities from private sources or from outside Portugal, (iv) the opening of new promising avenues of R&D.

±

Evidence		
Dimension	NO	PT
Organisation, leadership, strategy and resources	<ul style="list-style-type: none"> • Institutional self-assessment • Group self-assessment • RCN funding data 	<ul style="list-style-type: none"> • Institutional self-assessment
Research production and quality	<ul style="list-style-type: none"> • Group self-assessment • Group member short CV's • Access to group member publications in CRIS Stin • Sample outputs 	<ul style="list-style-type: none"> • Summary of up to 5 contributions that the R&D Unit considers more important (2013-2017) • A limited number of full-text publications (5, 10, 15, 20 for R&D Units with, respectively, <30, 30-59, 60-119, ≥120 Integrated PhD Researchers) • A limited number of "Nuclear CV" (3, 5, 10, 15 for R&D Units with, respectively, <30, 30-59, 60-119, ≥120 Integrated PhD Researchers) • Access to updated Curricula Vitae of all the Integrated Researchers reporting all the relevant R&D contributions in the period 2013-2017 and easy access to publications references through the ORCID platform. • Institutional self-assessment
Recruitment and training	<ul style="list-style-type: none"> • Group self-assessment 	<ul style="list-style-type: none"> • Institutional self-assessment
Networking	<ul style="list-style-type: none"> • Group self-assessment 	<ul style="list-style-type: none"> • Institutional self-assessment
Impact on teaching	<ul style="list-style-type: none"> • Group self-assessment 	<ul style="list-style-type: none"> • Institutional self-assessment
Overall	<ul style="list-style-type: none"> • All of the above 	<ul style="list-style-type: none"> • All of the above
Feedback	<ul style="list-style-type: none"> • All of the above 	<ul style="list-style-type: none"> • All of the above