CRITICAL ANALYSIS OF VQR 2015-2019 RESULTS

Edited by the Research Quality Assessment Organizational Unit
and the Third Mission Organizational Unit

Quotation allowed only with source reference.

Table of contents

1 Objectives of the report

The 2015-2019 Research Quality Assessment exercise was one of the key activities carried out by ANVUR, which ended in July 2022. Such activity was fully shared with the entire ANVUR Governing Board, composed by Proff. Alessandra Celletti, Marilena Maniaci, Menico Rizzi, Massimo Tronci, Antonio Uricchio, with ANVUR Director Dr. Daniele Livon, and with ANVUR Senior Manager for Research evaluation Dr. Marco Malgarini. In particular, Alessandra Celletti, besides being responsible for the entire exercise, was in charge of the areas of Mathematics and Computer Sciences, Physics, Third Mission/Impact; Menico Rizzi of the areas of Chemistry, Earth Sciences, Biology, Medicine, Agricultural and Veterinary Sciences; Massimo Tronci of the areas of Architecture, Civil Engineering, Industrial and Information Engineering; Marilena Maniaci of the areas of Classics, Philology, Literary Studies, Art History, History, Philosophy and Pedagogy, Political and Social Sciences; Antonio Uricchio of the areas of Psychology, Law Studies, Economics and Statistics, Business Studies, Third Mission/Impact.

The VQR 2015-2019 evaluated the research produced by 98 state and non-state universities, 14 research institutions supervised by MUR (Ministry of University and Research) and 22 institutions that voluntarily underwent the evaluation exercise. Overall, more than 182,000 scientific products (journal articles, contributions in conference proceedings, contributions in book, monographs, patents and other types of products) and 676 case studies for Third Mission/Impact activities have been submitted for evaluation, with more than 65,000 researchers involved. The results of the assessment exercise have been published in an extensive report available on the ANVUR website (https://www.anvur.it/attivita/vqr/vqr-2015-2019/rapporto-finale-anvur-e-rapporti-di-area/).

Having concluded the evaluation exercise, the objective of the present report is to carry out an "assessment of the assessment", namely to analyse the main characteristics of the exercise and to understand its strengths and weaknesses, also in the light of the most recent international debate on evaluation. Such work is preparatory to launch the next evaluation exercise, that will cover the period 2020-2024. In this regard, it is important to emphasise the need to define criteria and methodologies as early as possible in order to provide prompt guidance to the participating institutions.

The Report is divided into 3 further chapters: Chapter 2 is devoted to an in-depth analysis of the VQR 2015-2019 results, as regards both the evaluation of the research products, and that of the Third Mission/Impact case studies; Chapter 3 presents the results of two user satisfaction surveys on the quality of the procedures established by ANVUR; finally, Chapter 4 provides a brief outline of the ongoing European and global debate on the reform of research assessments.

The Report was draft by the VQR and Third Mission Organizational Units, composed by Dr. Brigida Blasi, Dr. Paola Costantini, Dr. Francesca Macrì, Dr. Irene Mazzotta, Dr. Francesca Pentassuglio, Dr. Sandra Romagnosi, Dr. Scipione Sarlo, with the collaboration of Dr. Marco Malgarini and Prof. Alessandra Celletti. The content of the report was shared and discussed with the members of the Governing Board and the ANVUR Director.

2 The VQR 2015-2019

The VQR 2015-2019 is the third evaluation exercise carried out by ANVUR; it follows the previous exercises referred to the periods 2004-2010 and 2011-2014. In this chapter of the report, the main regulatory references that have guided the evaluation exercise will be provided (paragraph 2.1); followed by an in-depth analysis of the VQR Call, aimed at highlighting the main innovations introduced with this VQR compared to the previous ones (paragraph 2.2), and by a critical examination of the results, referred to the evaluation of the research products (paragraph 2.3) and Third Mission/Impact case studies (paragraph 2.4).

2.1 Regulatory framework

The Presidential Decree no. 76 of 1 February 2010 defined ANVUR's structure and organisation; in particular, paragraph 1, lett. i-bis of the Presidential Decree, introduced under art. 1, paragraph 339, of the law no. 232 of 11 December 2016, provided that the Agency "carries out, every five years, the evaluation of research quality of universities and research institutes, on the basis of a special decree issued by the Ministry of Education, University and Research, enacted by 31 March of the year following the five-year period under assessment, and aimed at identifying the guidelines for carrying out the assessment itself and the economic resources required". The new evaluation exercise, covering the period 2015-2019, was therefore launched with the publication of the Ministerial Decree No. 1110 of 29 November 2019, which set out the general criteria for the evaluation exercise, specifying the fulfilments required for the institutions, the Expert Groups in charge of the evaluation and ANVUR. In particular, ANVUR was entrusted with the task of defining in detail the criteria and the modalities of the evaluation with a specific Call, approved by the Agency on 3 January 2020. The evaluation procedures then slowed down due to the pandemic; some adjustments to the VQR Guidelines have been introduced with the Ministerial Decree No. 444 of 11 August 2020, also as a result of the change of the ministerial leadership. The new Ministerial Decree was followed by an update of the ANVUR Call, approved by the Agency on 25 September 2020. Subsequently, the Ministerial Decree no. 289/2021 (General guidelines of the Universities Planning 2021-2023) established in art. 6 paragraph 4 that the products evaluated were assigned the following scores, corresponding to an assessment category:

- A – Excellent and extremely relevant: score 1
- B – Excellent: score 0,8
- C – Standard: score 0,5
- D – Sufficient relevance: score 0,2
- E – Low relevance or not eligible: score 0

2.2 The VQR Call: changes from previous evaluation exercises

The new Ministerial Decree and the new VQR Call present elements of novelty compared to previous evaluation exercises carried out by the Agency, referred to the periods 2004-2010 and 2011-2014, respectively. The general methodological framework of the evaluation, based on the peer review method – or, where appropriate, on peer review informed by citation indicators – is adopted. It is also maintained that the assessment is based on the three criteria of originality, methodological rigour and impact, defined in

much the same terms as the previous evaluation exercise. However, the combined provisions of the new Ministerial Decree and the new Call also introduce some important innovations.

First, in line with the provisions of Presidential Decree no. 76/2010, as amended by Law no. 232/2016, the evaluation period is set at five years, instead of the four of the previous VQR (seven in VQR 2004-2010); as a result, the number of products to be submitted for each researcher rises to three, compared to the two of the VQR 2011-2014. Moreover, unlike previous years, the number of products to be submitted is the same for university researchers and for those affiliated to research institutes. A major innovation concerned the possibility for the institutions to submit for each researcher a varying number of products, ranging from zero to four, provided that the products to be submitted at Department level were equal to three times the number of researchers belonging to it (without prejudice to any reductions provided for by art. 5 paragraph 6 of the Call). Such innovation may have significantly affected the assessment results themselves, allowing universities and Departments to make appropriate choices: indeed, according to this rule, it has been possible to exclude or limit the participation of researchers deemed less productive in the evaluation, while compensating with a higher contribution from researchers considered to be better ones. This effect may have been greater in the STEM+LS areas (STEM stands for Technology, Engineering, Mathematics and LS for Life Science), where simulation tools of achievable results were possibly available. This should also be taken into account in the interpretation of the results contained in paragraph 2.3.2. With regard to the number of products, the new Call also introduced the requirement to submit additional products if the same product was presented by a number of institutions higher than the threshold set by the Call.

Important innovations concerned the structure and selection procedures of the Evaluation Expert Groups (hereafter GEVs, acronym stemming from the Italian translation of "Evaluation Expert Groups") as well. Taking into account the increase in the number of products to be evaluated, the number of GEV members was increased to 600, including 30 members of the Third Mission/Impact interdisciplinary GEV. The Call also provided the possibility for GEVs to request, on the basis of specific and justified needs, a further addition of members. As a result, the total number of GEV members at the end of the VQR was 668 (including replacements and integrations during the exercise). An important change also concerned the structure itself of the disciplinary GEVs, with the division of the CUN[1] Area 13 – Economics and Statistics, into two distinct GEVs, 13a and 13b, responsible for the evaluation in the fields of Economics and Statistics, and Business Studies, respectively.

Table 2.2.1 lists the disciplinary areas of the GEVs that participated in the VQR 2015-2019.

---

[1] CUN (acronym for Consiglio Universitario Nazionale) is the Italian National University Council, an advisory board of the Ministry of Universities and Research which is responsible, among other things, for the definition of Academic fields.

Table 2.2.1 - List of disciplinary areas of the VQR 2015-2019.

GEV	Naming
Area 1	Mathematics and computer sciences
Area 2	Physics
Area 3	Chemistry
Area 4	Earth sciences
Area 5	Biology
Area 6	Medicine
Area 7	Agricultural and veterinary sciences
Area 8a	Architecture
Area 8b	Civil engineering
Area 9	Industrial and information engineering
Area 10	Classics, Philology, Literary Studies, Art History
Area 11a	History, Philosophy and Pedagogy
Area 11b	Psychology
Area 12	Law Studies
Area 13a	Economics and Statistics
Area 13b	Business Studies
Area 14	Political and Social Sciences
Interdisciplinary	Third mission

For the first time, GEV members have been selected by drawing lots, starting from a list of candidates considered to meet the scientific requirements set out in the Call and respecting certain rules regarding the final composition of the groups: these rules, set out in the Call, were intended to ensure a balanced composition in terms of gender, academic position and disciplinary composition, with special regard to the adequate coverage of the Academic Recruitment Fields and Academic Disciplines of each Area. The Call also established the incompatibility of the appointment as GEV member with the main academic offices and precisely defined the rules for the management of conflicts of interest that may arise in the assessment exercise, also taking into account the policy act of the Minister of Education, University and Research, no. 39 of 14/5/2018. An important role in the entire evaluation procedure was played by the GEV assistants, selected by a committee appointed by ANVUR among the applications submitted in response to a specific Call. Such figures performed organisational and secretarial tasks with great professionalism, thus facilitating the running of the entire exercise.

While using an informed peer review as a method of evaluation, the new Call has included some important innovations regarding the assessment categories and the scores assigned. In particular, the 5 assessment categories have been redefined compared to previous VQRs, as well as the scores assigned to each category, as shown in Table 2.2.2.

Table 2.2.2 - Assessment categories in VQRs 2004-2010, 2011-2014 and 2015-2019.

Assessment exercise	Assessment categories	Definition	Scoring
VQR 2004-2010	Excellent	The publication ranks in the top 20% of the value scale shared by the international scientific community. Excellent products are those recognised as excellent at international level for originality, methodological rigour and interpretative relevance; or those that have significantly innovated the field of studies at national level.	1
	Good	The publication ranks in the 60-80% segment. Good products are those of international and national importance, recognised for originality of the results and methodological rigour.	0,8
	Acceptable	The publication ranks in the 50-60% segment. Products of an acceptable level are those with international or national circulation that have increased to some extent the body of knowledge in the relevant fields.	0,5
	Limited relevance	The publication ranks in the bottom 50%. Products of a limited level are those with national or local circulation, or of no particular importance in the international context, which have made a modest contribution to the body of knowledge in the relevant fields.	0
	Not eligible for evaluation	The publication belongs to types excluded from the assessment exercise or presents attachments and/or documentation inadequate for evaluation or has been published in years before or after the seven-year period under evaluation.	-1
	Plagiarism or fraud	In proven cases of plagiarism or fraud.	-2
VQR 2011-2014	Excellent	The publication reaches the highest levels in terms of originality and methodological rigour, and has achieved or is likely to achieve a strong impact in the relevant scientific community at international and/or national level. Ideally, it ranks in the first 10% of the distribution of the international scientific production of its area.	1
	Good	The publication reaches good levels in terms of originality and methodological rigour, and has achieved or is likely to achieve a significant impact in the relevant scientific community at international and/or national level. Ideally, it ranks in the segment 10-30% of the distribution of the international scientific production of its area.	0,7
	Fair	The publication reaches fair levels in terms of originality and methodological rigour, and has achieved or is likely to achieve an appreciable impact in the relevant scientific community at international and/or national level. Ideally, it ranks in the 30-50% segment of the distribution of the international scientific production of its area.	0,4
	Acceptable	The publication reaches sufficient levels in terms of originality and methodological rigour, and has achieved or is likely to achieve a limited impact in the relevant scientific community at international and/or national level. Ideally, it ranks in the 50-80% segment of the distribution of the international scientific production of its area.	0,1
	Limited relevance	The publication achieves a low level of originality and methodological rigour, and has achieved or is likely to achieve a very limited impact in the relevant scientific community at international and/or national level. Ideally, it ranks in the 80%-100% segment of the distribution of the international scientific production of its area.	0
	Not eligible for evaluation	The publication belongs to types excluded from the assessment exercise, or presents attachments and/or documentation inadequate for evaluation or has been published in years before or after the four-year period under evaluation. This category also includes expected non-submitted publications.	0
VQR 2015-2019	Excellent and extremely relevant	The publication achieves the highest levels in terms of originality, knowledge and use of literature, methodological rigour and clarity of presentation, impact in the scientific community.	1

Assessment exercise	Assessment categories	Definition	Scoring
	Excellent	The publication reaches excellent levels in most of the following aspects: originality, knowledge and use of literature, methodological rigour and clarity of presentation, impact in the scientific community.	0,8
	Standard	The publication, compared to international standards, reaches a good level in terms of originality, knowledge and use of literature, methodological rigour and clarity of presentation, impact in the scientific community.	0,5
	Sufficient relevance	The publication, compared to the standards of the relevant scientific community, has a sufficient relevance in terms of originality, methodological rigour and clarity of presentation, even if it has limited impact in the scientific community.	0,2
	Low relevance or not eligible	The publication is of low relevance in terms of originality, knowledge and use of literature, methodological rigour and clarity of presentation, impact in the scientific community. This category also includes publications which belong to types excluded from the assessment exercise, or presenting attachments and/or documentation inadequate for evaluation.	0

As regards the evaluation of the Third Mission/Impact activities, please refer to paragraph 2.4 for a detailed examination of what is new in this assessment exercise compared to previous ones.

Important changes also concerned the calculation of the quality profiles into which the evaluation results were divided. In particular, the Ministerial Decree provided, as already in the previous VQR, the calculation of separate quality profiles for permanent staff and for those who were recruited or promoted during the 2015-2019 period. In addition to these profiles, a training-in-research profile has also been calculated to assess the institutions' ability to train staff capable of producing quality research: in this case, the results were calculated by aggregating the evaluations of products submitted on behalf of the researchers in service in the period 2015-2019 who obtained their doctoral degree in the period 2012-2016 at a university or institution participating in the evaluation exercise. More specifically, for the purposes of calculating the R3 and IRAS3 indicators, the result of the evaluation was attributed to the university where the PhD degree was awarded (and not to the university to which the researcher was affiliated as of 1/11/2019). In the case of doctoral programmes which involved several institutions, the result of the evaluation was proportionally attributed to the partner institutions.

A further important innovation brought about by the Ministerial Decree was the introduction of a limit to the division of the products evaluation into categories, with each category indicatively being assigned at least 5% and not more than 25% of products. Moreover, a quality profile dedicated to the Third Mission/Impact activities has been introduced for the first time, aimed at evaluating the activities of exploitation and dissemination of research (more on this topic in the section devoted to the Third Mission). On the other hand, some elements that appeared in the previous evaluation exercise, such as indicators X, IRAS3, IRAS4, IRAS5, have been dropped.

A final important change concerned the requirement to indicate which products were available in open access: the institutions were required to indicate at the time of submission whether the publications were open access, and to provide the link to the document (at the time of submission or later, but in any case by the end of the assessment exercise). In order to ensure full transparency, at the end of the assessment exercise the Agency published the complete list of metadata of the evaluated products on its website, by also making available the link to the open access whenever possible.

2.3 Analysis of VQR 2015-2019 results related to the evaluation of research products

2.3.1 Types of products submitted for evaluation

Overall, the number of products submitted for evaluation by universities and research institutes participating in the VQR was 181.280; if we also take into account the monographs, for which a double value (1.368 products) was indicated by the institutions, in accordance with the VQR Call, the total number of submitted products amounts to 182.648. In what follows, however, for the analysis of the submitted products by type, data have been considered as net of the double value attributed to monographs. Figure 2.3.1 shows the distribution of products by type, considering all the types envisaged by the VQR Call: as it can be noted, the Call allowed for the possibility of submitting a considerable variety of types of research products, but the attention of researchers was ultimately focused only on some of them, namely journal articles (which alone account for 87,2% of the total number of publications submitted), contributions in book and monographs.

Figure 2.3.1 - Submitted products by type of publication.
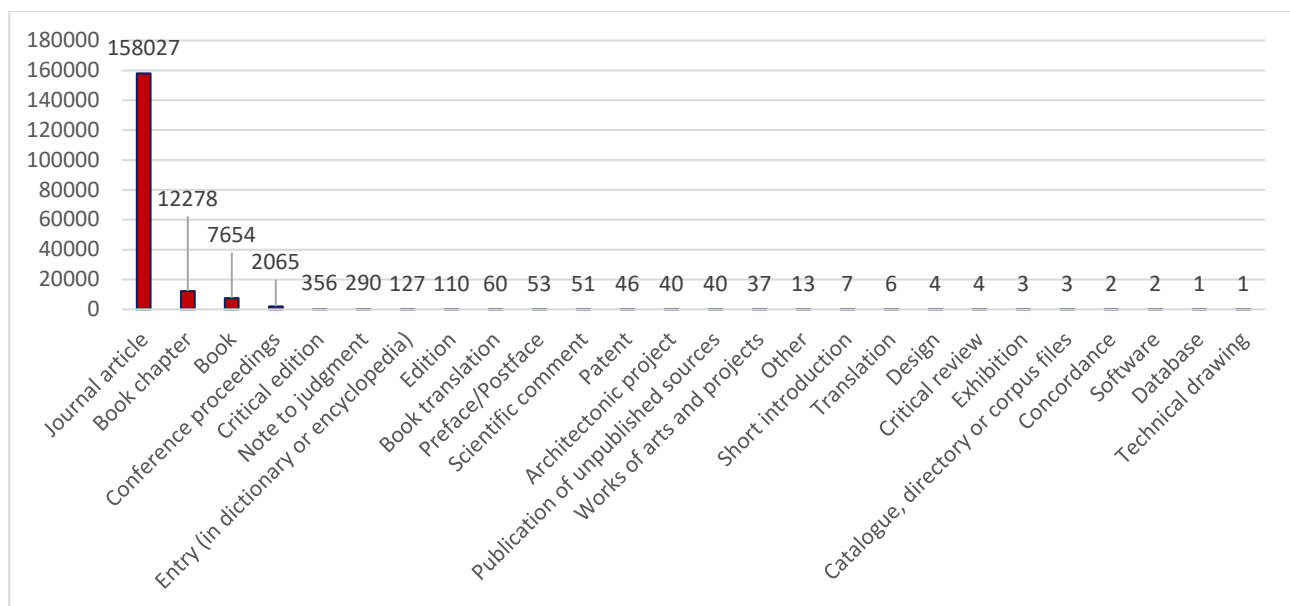
Figure 2.3.2 shows in detail the distribution of the products by type according to the researcher's area, grouping the products according to the main categories defined in the VQR Call (art. 5, paragraph 2). The areas have been grouped into two sub-categories, distinguishing between the areas of Technology, Engineering, Mathematics and Life Sciences (STEM+LS in the Anglo-Saxon acronym) and those of Social Sciences and Humanities (SSH). Journal articles are the predominant dissemination channel for scientific research in all areas. In the Social Sciences and Humanities, however, contributions in books and monographs (or similar products) are also significantly represented. It should be noted from the outset that areas 13a and 13b share common elements with both STEM+LS and SSH areas.

Figure 2.3.2 - Submitted products by type and area of researcher.
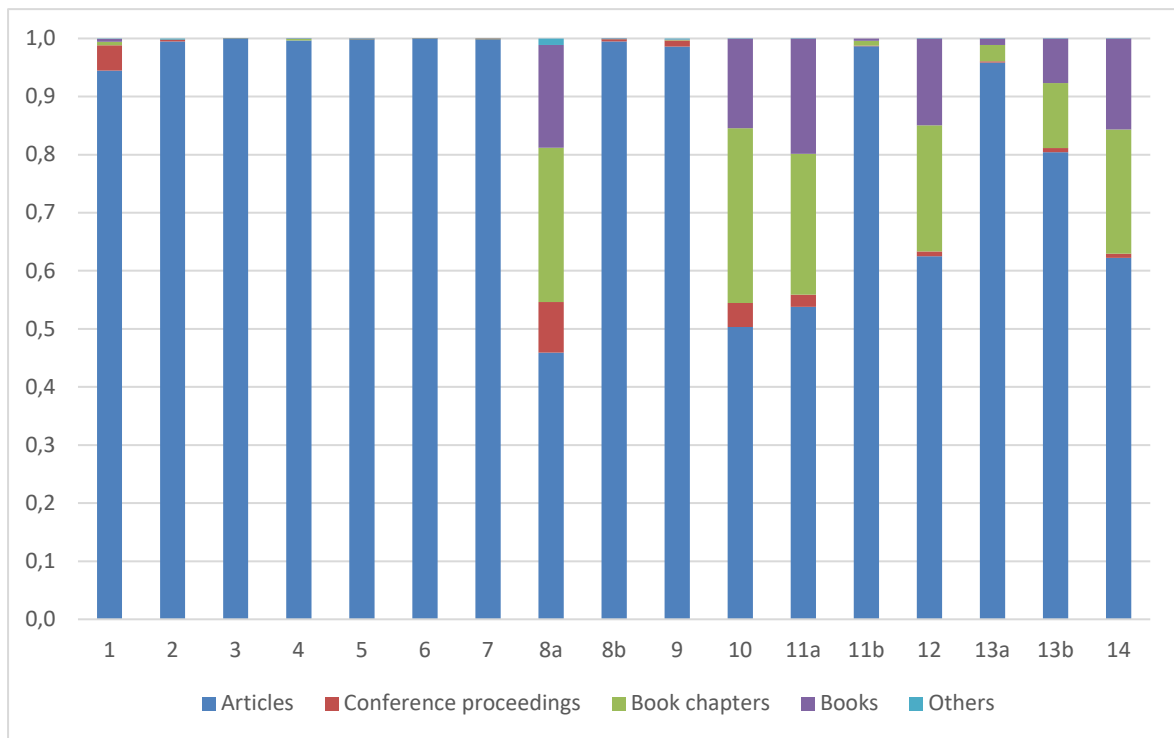
Figure 2.3.3 presents the evolution over time of the types of publication submitted for evaluation, based on data from VQR 2004-2010, 2011-2014 and 2015-2019.

Figure 2.3.3 - The evolution of the types of publication submitted in the 3 VQR exercises.
   a.   STEM+LS areas
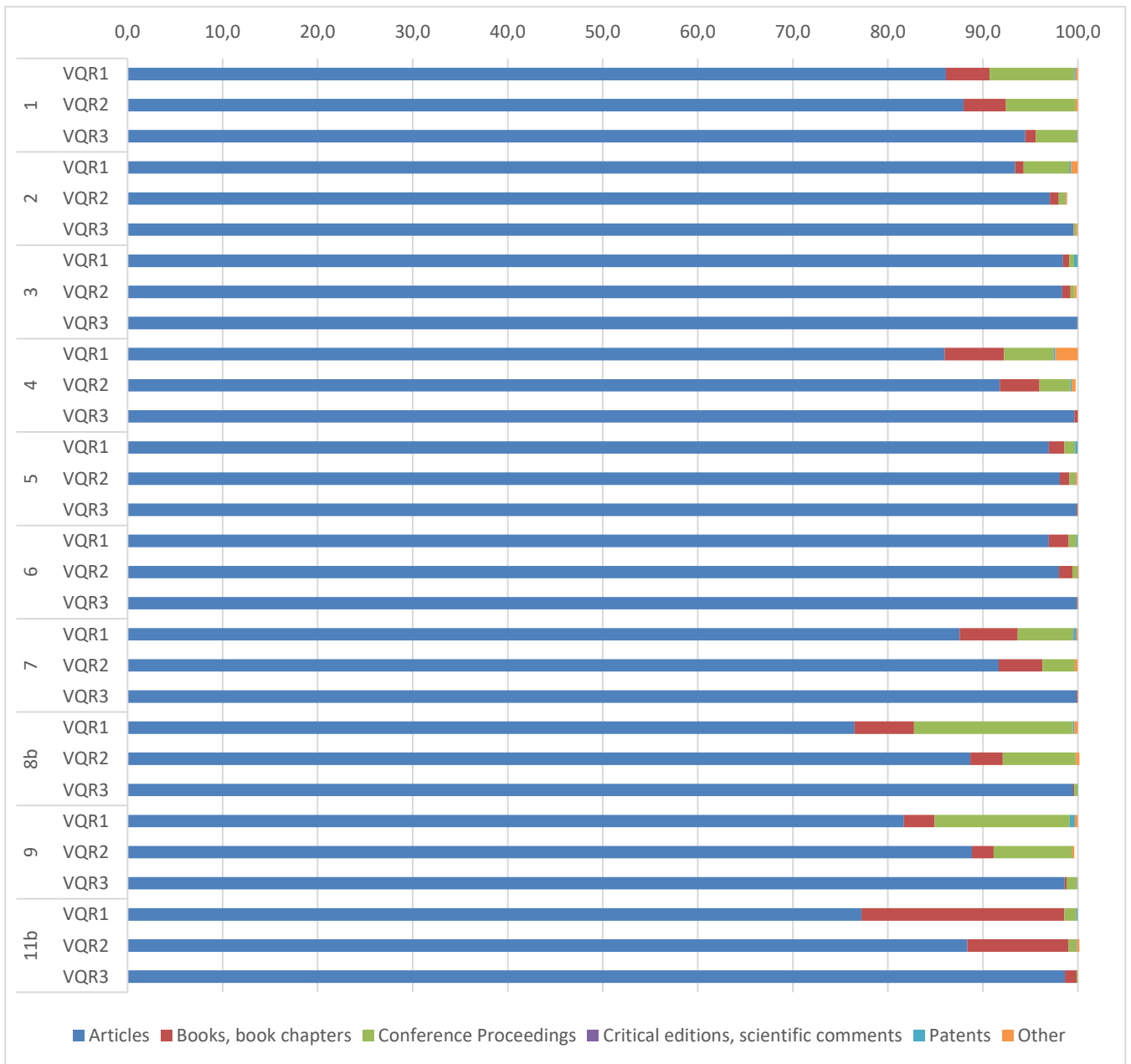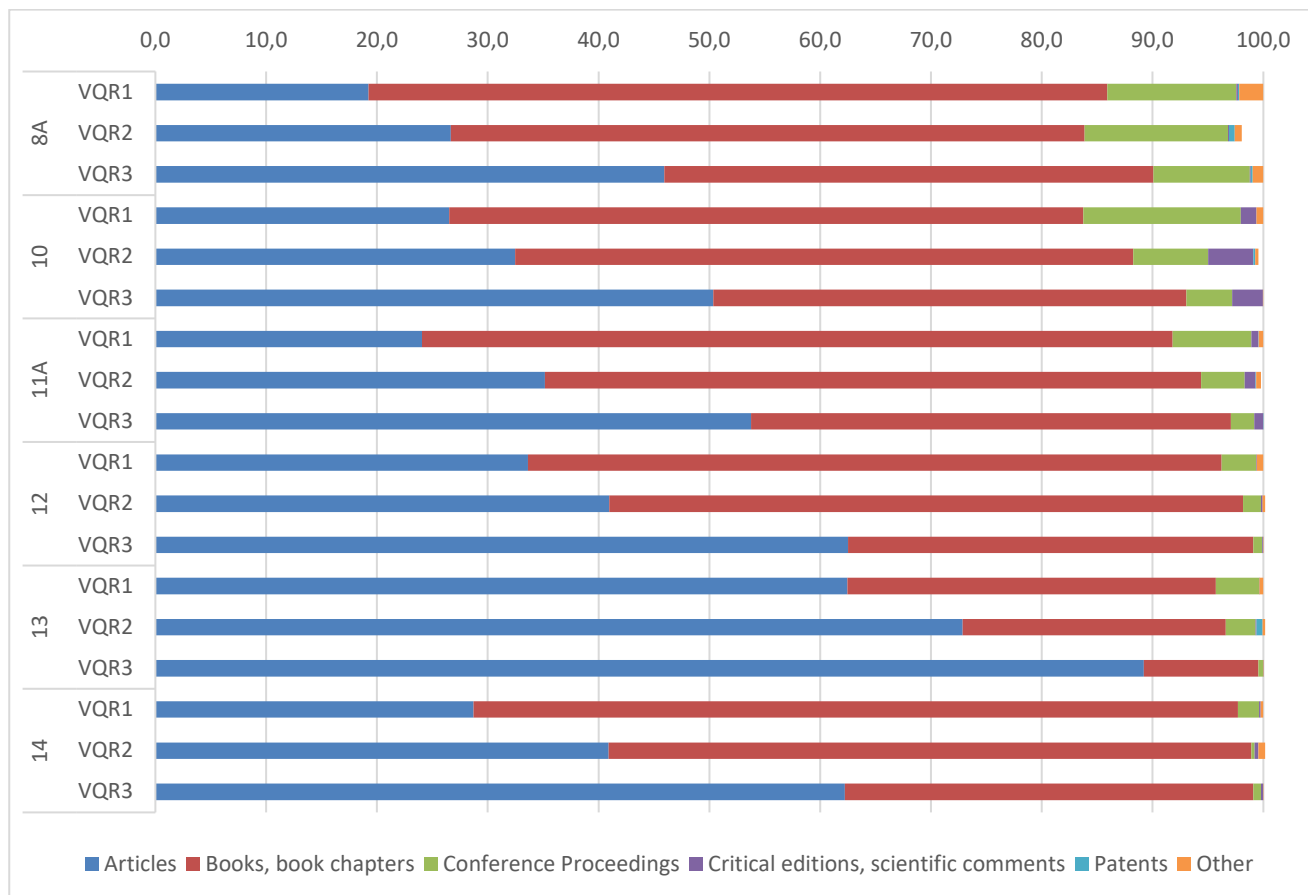
The data show a clear trend, between the first and third VQRs, towards an increase in the "journal articles" type among the publications submitted in all areas. Monographs, although decreasing, still retain an important role in SSH, while their incidence is marginal in STEM+LS areas, even in those, such as area 11b, in which they played a significant role in the past. It should be noted that in previous VQRs areas 13a and 13b were considered jointly; they have therefore been re-grouped for the latest VQR in order to allow comparison.

Finally, it should be emphasised that the STEM+LS areas differ from the Social Sciences and Humanities also with respect to the language of the research products submitted for evaluation. Indeed, in STEM+LS areas, nearly all products are written in English, whereas in the Social Sciences and Humanities there is a share of products (sometimes the majority) written in Italian and considerable shares of products written in other languages (Figure 2.3.4).
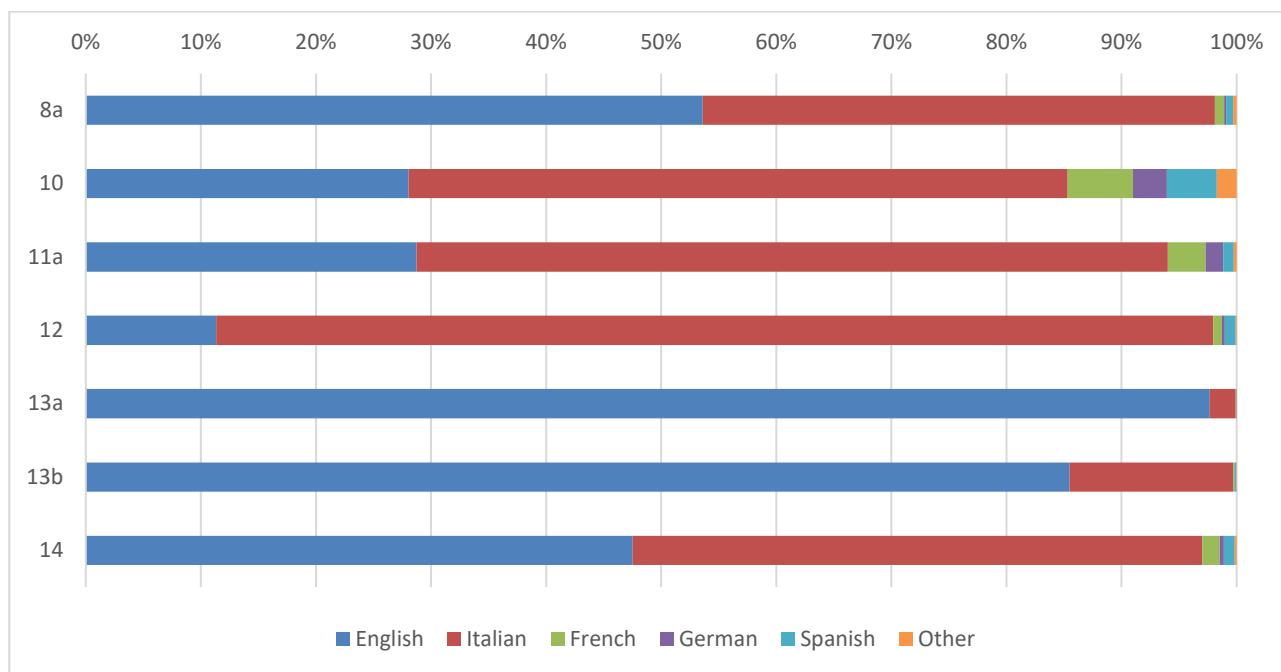
2.3.2 In-depth analysis of the results with respect to the characteristics of the submitted products, the submitting institutions and the researchers

This section presents an in-depth analysis of the relationship between the results of the evaluation, the characteristics of the submitted products, and those of the institutions and researchers who have submitted them. Such an analysis may provide useful information to understand merits and shortcomings of the evaluation exercise, and to gain useful indications for the future of the evaluation process. In the considerations that follow, the results of the evaluation will be expressed for each individual product in terms of its placement in one of the five categories envisaged by the VQR Call; in some cases, the analysis will be limited to the product's placement in the two categories of excellence A and B. Table 2.3.1 lists the characteristics that will be analysed below.

Table 2.3.1 - Characteristics of products, institutions, and researchers.

Characteristics of products	Characteristics of the institutions	Characteristics of researchers
Type (Journal article; Contribution in conference proceedings; Contribution in book; Monograph; Other)	Institution (state; non-state; with special regulations)	Gender (Male; Female)
Year of publication	Geographical area (North West; North East; Center; South)	Academic Position (Full Professor; Associate Professor; Researcher; fixed-term Researcher)
Language (only for SSH: Italian; English; other language)		Disciplinary area (STEM+LS; SSH)
Availability in open access (YES/NO)		Profile (Permanent; Newly recruited)

To consider simultaneously the effect of all the above characteristics on the product's placement in the assessment categories, an ordered multinomial logistics model taking all five categories into account was first estimated; subsequently, attention was focused on the probability of obtaining an overall excellent evaluation (indicated by placement in categories A and B). The multinomial model has the advantage of simultaneously studying how the placement in one of the five assessment categories (ordered from best to worst) relates to the above characteristics; estimates can detect the direction of the effect (positive or negative), but not the size of the effect itself. The model has therefore also been estimated on the binomial variable, represented by the placement of a product in the two categories of excellence or in the remaining three: in this case, the estimated coefficients can be immediately interpreted as the effect of a given characteristic on the probability of belonging to the category of excellent products.

More formally, we assume that the probability for the i-th product submitted to the VQR to receive an evaluation equal to x ∈ {A; B; C; D; E} for the ordered multinomial model, or to x ∈ {A∩B; ≠A∩B) for the binomial model, is related to the above-mentioned characteristics as follows:

$$P \; (Evaluation_i = x) = \; F(characteristics \; of \; the \; product_i, characteristics \; of \; the \; researcher, \\ characteristics \; of \; the \; institution_i)$$

The function F is the logistic function; it is assumed therefore that the relation between the variables is non-linear. The method used to estimate the regression parameters is that of maximum plausibility. In other words, the method involves choosing parameters that maximize the probability (or plausibility) of observation in the considered sample. Table 2.3.2 presents the results of the ordered multinomial model; the appraisal was made only on the products submitted by universities, for which all the information on the above-mentioned characteristics is available. A total of 158.463 products were considered in the analysis. The table shows the reference categories for each characteristic used; the second column contains the value of Student's t statistics for each category; the third column shows the effect (positive - in green, or negative - in red) of each considered category (compared to the reference category) as regards the probability of receiving a less favourable assessment. For example, being "male" has a positive influence (positive sign of Student's t value) on the probability of having a more favourable assessment.

Table 2.3.2 - Results of the ordered multinomial model, t-value for characteristics of the products, institutions and researchers.

Variable	t-value	Probability of a more favourable assessment compared to the reference category
Characteristics of the researcher (Reference categories: Female; Full Professor; Newly recruited; STEM+LS areas)		
Male	-2,219	✓
Associate Professor	37,002	✓
Fixed-term researcher	26,763	✓
Researcher	61,671	✓
Permanent profile	27,653	✓
SSH areas	55,206	✓
Characteristics of the institution (Reference categories: State universities; Geographical area Centre)		
Non-State universities	6,283	✓
Schools of advanced Studies with special regulations	-17,936	✓
SOUTH	12,855	✓
NORTH-EAST	-17,973	✓
NORTH-WEST	-12,185	✓
TELEMATIC UNIVERSITY	25,908	✓
Characteristics of the product (Reference categories: No open access available; Year 2015; Type Journal article)		
Open Access Availability	-6,716	✓
YEAR 2016	2,602	✓
YEAR 2017	2,717	✓
YEAR 2018	2,473	✓
YEAR 2019	8,845	✓
Other type of product	7,190	✓
Contribution in conference Proceedings	17,354	✓
Contribution in book	27,483	✓
Scientific monograph	-4,261	✓

We can summarise the results of the table as follows. The probability of obtaining more favourable evaluations is greater for:

- Men compared to women
- Full professors compared to the other categories
- Professors who have been recruited or promoted during the five-year assessment period
- Professors in the STEM+LS areas
- Professors at Schools of advanced Studies with special regulations with respect to those at State universities, and the latter with respect to those at the non-State universities
- Professors at universities of the North-West and North-East compared to those of the Centre, and the latter with respect to those of the South and the telematic universities
- Products published in 2015 compared to other years
- Products available in open access

- Journal articles compared to the other types, with the only exception of monographs, which have a more favourable assessment even compared to journal articles.

As mentioned above, the multinomial ordered logistics model allows us for a first assessment of the effect of the considered variables on the evaluation results. In order to be able to obtain an assessment of the magnitude of these effects, the model was estimated by considering as the dependent variable the binary variable representing the probability of obtaining an overall excellent rating, i.e. the probability that a product falls into the rating classes A and B. Table 2.3.3 shows the results, obtained estimating the model both on the entire universe of available products, and considering separately the products submitted in the STEM+LS and SSH areas. The available observations are 99.782 for STEM+LS areas and 58.681 for SSH areas, respectively. Moreover, compared to the previously estimated model, "traditional" non-State universities and telematic universities have been considered separately, and the latter have been attributed the geographical area to which they belong according to their registered office. To better grasp the differences that occur on a geographical basis, the Southern regions of the South were also considered separately from those of the Islands.

Table 2.3.3 - Results of the logistic model, odds ratio for characteristics of the products, institutions and researchers: total, STEM+LS areas and SSH areas.

Variables	Total		STEM+LS		SSH	
	ODDS RATIO	% compared to the reference category	ODD RATIO	% compared to the reference category	ODD Ratio	% compared to the reference category
Characteristics of the researcher (Reference categories: Female; Full professor; Newly recruited; STEM+LS)						
Male	0,985	NS	0,967	-3,31%	1,056	5,56%
Associate Professor	0,626	-37,39%	0,692	-30,78%	0,524	-47,64%
Fixed-term researcher	0,619	-38,05%	0,706	-29,35%	0,489	-51,12%
University researcher	0,383	-61,68%	0,430	-56,96%	0,294	-70,61%
Permanent profile	0,736	-26,38%	0,722	-27,79%	0,808	-19,18%
SSH areas	0,481	-51,86%				
Characteristics of the institution (Reference categories: State universities; Geographical area Centre)						
Non-State universities (excluding telematic universities)	0,826	-17,36%	0,742	-25,80%	0,837	-16,27%
Schools of advanced Studies with special regulations	4,217	321,70%	5,118	411,83%	2,859	185,88%
Telematic universities	0,306	-69,41%	0,304	-69,57%	0,264	-73,59%
ISLANDS	0,736	-26,37%	0,768	-23,16%	0,799	-20,15%
NORTH-EAST	1,319	31,93%	1,416	41,58%	1,111	11,07%
NORTH-WEST	1,199	19,90%	1,229	22,93%	1,129	12,94%
SOUTH	0,924	-7,59%	0,977	NS	0,807	-19,29%
Characteristics of the product (Reference categories: No open access available; Year 2015; Type: Journal article; Language: other language)						
YEAR 2016	0,960	-4,04%	0,956	-4,39%	0,973	NS
YEAR 2017	0,965	-3,50%	0,937	-6,26%	1,022	NS
YEAR 2018	0,960	-4,00%	0,929	-7,06%	1,002	NS
YEAR 2019	0,859	-14,09%	0,789	-21,06%	0,960	NS
MACRO-TYPE: Other	0,318	-68,16%	0,180	-82,03%	0,445	-55,54%
MACRO-TYPE: Contribution in conference Proceedings	0,485	-51,55%	0,613	-38,67%	0,434	-56,61%
MACRO-TYPE: Contribution in book	0,623	-37,66%	0,411	-58,86%	0,628	-37,24%
MACRO-TYPE: Scientific monograph	1,119	11,90%	0,910	-8,99%	1,278	27,76%
Open Access: YES	1,037	3,70%	1,221	22,06%	0,833	-16,66%
Language Other					1,202	20,20%
Language English					1,535	53,54%

The table also directly shows the difference in the estimated probability of obtaining an excellent evaluation between the category under consideration and the reference category. Please note that each variable is considered *ceteris paribus*, i.e. when it is said, for example, that the results for full professors are better than those for associate professors and researchers, this is to be understood with all other variables (gender,

geographical location, type of university, etc.) being equal. All this considered, the probability of obtaining an excellent evaluation (A+B):

- overall, is essentially not influenced by gender; in STEM+LS areas alone, men are about 3% less likely to achieve excellence than women; in SSH on the contrary, men are 5,5% more likely than women to be excellent;
- is significantly higher for full professors, than for other categories;
- is significantly higher for professors who have been recruited or promoted during the five-year assessment period than for permanent researchers. The overall model was also estimated by considering the interaction effect between academic position (full and associate professor, permanent and fixed-term researcher) and stability/promotion in the role over the considered period. Taking the category of full professors who were not recruited/promoted during the VQR period as a reference, the analysis shows that the category "Full professors-newly recruited" has a better performance compared to the reference category, while all other categories achieve worse results than "Full Professors-permanent position";
  The role played by the remaining variables is unchanged from what is shown in the first two columns of Table 2.3.3;
- is significantly higher for professors in the STEM+LS areas;
- is significantly higher for professors at Schools of advanced Studies with special regulations than for those at State universities, and for the latter compared to the non-State non-telematic universities, which in turn have a better performance than telematic universities;
- is significantly higher for professors at universities in the North-West and North-East than for those in the Centre, and for the latter than for those in the South and the Islands;
- is slightly higher for products published in 2015 than for all other years, and for products published in the years 2016-18 than for those in 2019; however, this effect is only found in STEM+LS areas, whereas in SSH areas the evaluation result is essentially not affected by the year of publication;
- is greater overall for products available in open access; however, the effect is opposite for STEM+LS areas compared to SSH areas, since in the latter the products available in open access are less likely to obtain an excellent evaluation;
- is higher for journal articles than for other types of publication, with the only exception of monographs, which have a more favorable assessment even compared to journal articles. Here, too, a significant difference emerges between STEM+LS and SSH areas, since in the latter monographs are more likely to receive an excellent evaluation than journal articles, whereas the opposite is true in the STEM+LS areas;
- is higher, in SSH areas only, for products written in English than for those in languages other than English and Italian, and for the latter than for those written in Italian.

The results show clear differences, sometimes substantial, between STEM+LS and SSH areas. It was therefore considered appropriate to further deepen the analysis, so as to take into account the different areas of evaluation. However, it is important to note that some differences can be explained by referring to specific publication practices (monographs, use of English) and evaluation methods (the use – albeit collateral – of bibliometry impacts on the differences per year of publication and on the higher average evaluations for STEM+LS areas).

2.3.3 The training-in-research profile

As already mentioned, one of the main innovations brought about by the Ministerial Decree no. 289/2021 concerns the introduction of a new profile, related to the scientific production of those who are now accredited within the Italian academic system and have obtained a PhD degree in the 2012-16 period. Indeed, the Ministerial Decree no. 289/2021 establishes, under Art. 6, paragraph 5 that "The Ministry upon ANVUR's proposal, shall define the indicators referred to in paragraph 2, letters a) and b), and in paragraph 3, based at least on the average score attributed to research products", i.e. the indicators relating to the quality profiles mentioned in section 2.2. Among these, the third one refers to the quality profile of the products submitted by researchers who, as of 1 November 2019, were affiliated to one of the institutions participating in the VQR having obtained their PhD in the period 2012-2016 from one of the institutions participating in the VQR. Going into the detail of the calculation of the R3 and IRAS3 indicators, the result of the evaluations associated to the products of PhDs was attributed proportionally to the institutions (more than one in case of a joint title) at which the researcher associated with the product had obtained the degree (in addition to the institution to which the researcher was affiliated as of 1/11/2019). In this section we present an analysis of the data used for the construction of the R3 and IRAS3 indicators defined in the Call, paying particular attention to the origin of the researchers considered for the calculation of the indicators.

The following tables illustrate the flows (accreditation vs. degree attainment) of researchers among the various institutions, types and geographical areas; in the analysis, in the case of joint PhD programmes each PhD is attributed to all participating institutions. Table 2.3.4 shows that PhDs from Schools of advanced Studies were the most subject to mobility at the time of accreditation. Barely 16% of them were accredited by the same institution where they had obtained their degree, compared to 28,6% of PhDs from non-State universities and 31,7% of those from state universities.

Table 2.3.4 - Distribution of PhDs by subset of the university where they have obtained the degree (in the case of joint PhD programmes, each PhD accredited in VQR is assigned to all partner institutions).

Subset of the institution	number of PhDs from the same institution	number of PhDs from other institution	total number of accredited PhDs in VQR	number of PhDs accredited by Schools of advanced Studies	number of PhDs accredited by State universities	number of PhDs accredited by non-State universities	number of PhDs accredited by Research institutes	Number of PhDs accredited by other institutions	number of doctors same institution values % [compared to the total accredited by institution]	number of doctors other institution values % [compared to the total accredited by institution]	number of doctors accredited by Special Schools values % [compared to the total accredited by other institution]	number of doctors accredited by State universities values % [compared to the total accredited by other institution]	number of doctors accredited by non-state universities values % [compared to the total accredited by other institution]	number of doctors accredited by EPR values % [compared to the total accredited by other institution]	number of doctors accredited by institutions different values % [compared to the total accredited by other institution]
State	1.898	4.080	5.978	22	2.775	379	802	102	31,7	68,3	0,5	68,0	9,3	19,7	2,5
Non-State	106	264	370	3	214	31	15	1	28,6	71,4	1,1	81,1	11,7	5,7	0,4
Schools of advanced Studies	21	110	131	4	81	6	17	2	16,0	84,0	3,6	73,6	5,5	15,5	1,8

In terms of territorial mobility, PhDs from the Islands were those who were accredited to a lesser extent by the same institution from which they had obtained their degree, a little over a quarter of the total: 26,1% (see Table 2.3.5). They are followed by PhDs from the Centre with 27,4% and those from the North-West with 29,8%. The lowest mobility was recorded among PhDs from the South and those from the North-East, that in 35% of the cases were accredited by the same institution from which they had obtained their degree.

Table 2.3.5 - Distribution of PhDs by geographical area of the university at which they obtained the degree (in the case of joint PhD programmes, each PhD accredited in VQR is assigned jointly to all affiliated institutions).

Geographical area of the institution	number of PhDs from the same institution	number of PhDs from other institution	total number of accredited PhDs in VQR	number of accredited PhDs in the NORTH WEST	number of accredited PhDs in the NORD-EST	number of accredited PhDs in the CENTRE	number of accredited PhDs in the SUD	number of accredited PhDs in the ISLANDS	number of PhDs from the same institution % [compared to the total number of accredited researchers per institution]	number of PhDs from other institution % [compared to the total number of accredited researchers per institution]	number of accredited PhDs in the NORTH WEST % [compared to the total number of accredited researchers by other institution]	number of accredited PhDs in the NORD-EST % [compared to the total number of accredited researchers by other institution]	number of accredited PhDs in the CENTRE % [compared to the total number of accredited researchers by other institution]	number of accredited PhDs in the SOUTH % [compared to the total number of accredited researchers by other institution]	number of accredited PhDs in the ISLANDS % [compared to the total number of accredited researchers by other institution]
NORTH WEST	585	1.085	1.670	521	228	249	37	50	35,0	65,0	48,0	21,0	22,9	3,4	4,6
NORTH EAST	408	961	1.369	217	391	235	83	35	29,8	70,2	22,6	40,7	24,5	8,6	3,6
CENTER	444	1.179	1.623	192	186	587	164	50	27,4	72,6	16,3	15,8	49,8	13,9	4,2
SOUTH	447	829	1.276	104	69	311	297	48	35,0	65,0	12,5	8,3	37,5	35,8	5,8
ISLANDS	141	400	541	106	49	146	61	38	26,1	73,9	26,5	12,2	36,5	15,2	9,5

By examining the data in more detail, and by looking at the level of the individual institution, we can observe a range from institutions that did not accredit any of the PhDs who have obtained the title at their institution to extreme and opposite cases of their total uptake, see Table 2.3.6.

Table 2.3.6 - Distribution of PhDs by university where they have obtained the title (in the case of doctorates in agreement, each doctor accredited in VQR is assigned to all affiliated institutions).

Institution	number of PhDs same Institution	number of PhDs other Institution	total number of accredited PhDs in VQR	number of PhDs accredited by Schools of advanced Studies	number of PhDs accredited by State universities	number of PhDs accredited by non-State universities	number of PhDs accredited by Research institutes	number of PhDs accredited by other institutions	Number of PhDs from the same institution % [compared to the total number of accredited researchers per institution]	number of PhDs from other institution % [compared to the total number of accredited researchers per institution]	number of PhDs accredited by Schools of advanced Studies% [compared to the total number of accredited reseacrhers by other institution]	number of PhDs accredited by State universities % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by non-State universities % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by Research institutes % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by other institutions % [compared to the total number of accredited researchers by other institution]
Aosta	0	9	9	0	9	0	0	0	0	100	0	100	0	0	0
Bari	48	60	108	0	37	1	20	2	44,4	55,6	0	61,7	1,7	33,3	3,3
Bari Politecnico	13	23	36	0	11	1	10	1	36,1	63,9	0	47,8	4,3	43,5	4,3
Basilicata	6	28	34	0	16	1	10	1	17,6	82,4	0	57,1	3,6	35,7	3,6
Bergamo	15	72	87	0	58	7	7	0	17,2	82,8	0	80,6	I9,7	9,7	0
Bologna	105	212	317	0	152	20	37	3	33,1	66,9	0	71,7	9,4	17,5	1,4
Bolzano	11	9	20	0	7	0	2	0	55	45	0	77,8	0	22,2	0
Brescia	28	90	118	0	68	10	12	0	23,7	76,3	0	75,6	11,1	13,3	0
Cagliari	43	85	128	0	64	8	13	0	33,6	66,4	0	75,3	9,4	15,3	0
Calabria (Arcavacata di Rende)	24	33	57	0	20	2	10	1	42,1	57,9	0	60,6	6,1	30,3	3
Dressing room	3	16	19	0	14	1	1	0	15,8	84,2	0	87,5	6,2	6,2	0
Casamassima LUM	4	7	11	0	5	2	0	0	36,4	63,6	0	71,4	28,6	0	0
Cassino	4	24	28	0	21	1	2	0	14,3	85,7	0	87,5	4,2	8,3	0
Castellanza LIUC	2	13	15	0	11	2	0	0	13,3	86,7	0	84,6	15,4	0	0
Catania	26	87	113	0	59	4	23	1	23	77	0	67,8	4,6	26,4	1,1
Catanzaro	7	31	38	0	23	5	3	0	18,4	81,6	0	74,2	16,1	9,7	0
Chieti and Pescara	18	43	61	0	36	3	4	0	29,5	70,5	0	83,7	7	9,3	0
Enna Kore	5	5	10	0	5	0	0	0	50	50	0	100	0	0	0
Ferrara	32	75	107	0	45	6	13	11	29,9	70,1	0	60	8	17,3	14,7
Firenze	70	144	214	1	80	14	41	8	32,7	67,3	0,7	55,6	9,7	28,5	5,6
Foggia	8	28	36	0	24	3	1	0	22,2	77,8	0	85,7	10,7	3,6	0
Genova	63	122	185	0	82	5	34	1	34,1	65,9	0	67,2	4,1	27,9	0,8
Insubria	5	54	59	1	39	3	9	2	8,5	91,5	1,9	72,2	5,6	16,7	3,7
Aquila	27	27	54	1	18	2	6	0	50	50	3,7	66,7	7,4	22,2	0

Institution	number of PhDs same Institution	number of PhDs other Institution	total number of accredited PhDs in VQR	number of PhDs accredited by Schools of advanced Studies	number of PhDs accredited by State universities	number of PhDs accredited by non-State universities	number of PhDs accredited by Research institutes	number of PhDs accredited by other institutions	Number of PhDs from the same institution % [compared to the total number of accredited researchers per institution]	number of PhDs from other institution % [compared to the total number of accredited researchers per institution]	number of PhDs accredited by Schools of advanced Studies% [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by State universities % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by non-State universities % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by Research institutes % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by other institutions % [compared to the total number of accredited researchers by other institution]
Lucca - IMT	0	10	10	1	7	0	2	0	0	100	10	70	0	20	0
Macerata	1	31	32	0	27	4	0	0	3,1	96,9	0	87,1	12,9	0	0
Marche	15	23	38	0	12	6	5	0	39,5	60,5	0	52,2	26,1	21,7	0
Messina	19	69	88	0	47	10	12	0	21,6	78,4	0	68,1	14,5	17,4	0
Milano	79	147	226	2	99	17	27	2	35	65	1,4	67,3	11,6	18,4	1,4
Milano Bicocca	36	91	127	2	65	7	16	1	28,3	71,7	2,2	71,4	7,7	17,6	1,1
Milano Bocconi	2	16	18	0	13	3	0	0	11,1	88,9	0	81,2	18,8	0	0
Milano Cattolica	53	115	168	1	95	10	8	1	31,5	68,5	0,9	82,6	8,7	7	0,9
Milano IULM	1	5	6	0	4	1	0	0	16,7	83,3	0	80	20	0	0
Milano Politecnico	86	43	129	1	24	5	10	3	66,7	33,3	2,3	55,8	11,6	23,3	7
Milano San Raffaele	8	19	27	1	17	1	0	0	29,6	70,4	5,3	89,5	5,3	0	0
Modena and Reggio Emilia	21	60	81	0	51	4	3	2	25,9	74,1	0	85	6,7	5	3,3
Molise	3	44	47	0	37	4	3	0	6,4	93,6	0	84,1	9,1	6,8	0
Napoli Benincasa	1	4	5	0	4	0	0	0	20	80	0	100	0	0	0
Napoli Federico II	170	176	346	0	101	12	62	1	49,1	50,9	0	57,4	6,8	35,2	0,6
Napoli II	38	79	117	0	59	4	14	2	32,5	67,5	0	74,7	5,1	17,7	2,5
Napoli L'Orientale	10	7	17	0	6	1	0	0	58,8	41,2	0	85,7	14,3	0	0
Napoli Parthenope	12	63	75	0	43	5	12	3	16	84	0	68,3	7,9	19	4,8
Napoli Pegaso	3	0	3	0	0	0	0	0	100	0					
Padova	119	132	251	0	85	12	34	1	47,4	52,6	0	64,4	9,1	25,8	0,8
Palermo	35	99	134	0	65	6	26	2	26,1	73,9	0	65,7	6,1	26,3	2
Parma	22	118	140	2	97	8	10	1	15,7	84,3	1,7	82,2	6,8	8,5	0,8
Pavia	37	78	115	1	64	8	5	0	32,2	67,8	1,3	82,1	10,3	6,4	0
Pavia IUSS	1	7	8	1	6	0	0	0	12,5	87,5	14,3	85,7	0	0	0
Perugia	25	79	104	0	52	10	16	1	24	76	0	65,8	12,7	20,3	1,3
Perugia Stranieri	1	0	1	0	0	0	0	0	100	0					

Institution	number of PhDs same Institution	number of PhDs other Institution	total number of accredited PhDs in VQR	number of PhDs accredited by Schools of advanced Studies	number of PhDs accredited by State universities	number of PhDs accredited by non-State universities	number of PhDs accredited by Research institutes	number of PhDs accredited by other institutions	Number of PhDs from the same institution % [compared to the total number of accredited researchers per institution]	number of PhDs from other institution % [compared to the total number of accredited researchers per institution]	number of PhDs accredited by Schools of advanced Studies% [compared to the total number of accredited reseacrhers by other institution]	number of PhDs accredited by State universities% [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by non-State universities% [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by Research institutes% [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by other institutions% [compared to the total number of accredited researchers by other institution]
Piemonte Orientale	12	59	71	0	44	6	8	1	16,9	83,1	0	74,6	10,2	13,6	1,7
Pisa	70	93	163	3	55	4	27	4	42,9	57,1	3,2	59,1	4,3	29	4,3
Pisa Normale	1	48	49	3	30	3	11	1	2	98	6,2	62,5	6,2	22,9	2,1
Pisa S.Anna	19	25	44	20	1	3	0	1	43,2	56,8	80	4	12	0	4
Reggio Calabria	4	31	35	0	22	1	8	0	11,4	88,6	0	71	3,2	25,8	0
Roma Biomedico	13	24	37	0	18	2	4	0	35,1	64,9	0	75	8,3	16,7	0
Roma Europea	1	5	6	1	3	1	0	0	16,7	83,3	20	60	20	0	0
Roma Foro Italico	0	11	11	0	9	2	0	0	0	100	0	81,8	18,2	0	0
Roma La Sapienza	131	287	418	2	159	44	69	13	31,3	68,7	0,7	55,4	15,3	24	4,5
Roma LUISS	0	29	29	0	22	7	0	0	0	100	0	75,9	24,1	0	0
Roma LUMSA	1	3	4	0	0	2	1	0	25	75	0	0	66,7	33,3	0
Roma Marconi	0	1	1	0	1	0	0	0	0	100	0	100	0	0	0
Roma Tor Vergata	39	108	147	1	61	25	20	1	26,5	73,5	0,9	56,5	23,1	18,5	0,9
Roma Tre	22	65	87	0	32	8	25	0	25,3	74,7	0	49,2	12,3	38,5	0
Roma UNICUSANO	1	0	1	0	0	0	0	0	100	0					
Salento	13	43	56	1	26	3	10	3	23,2	76,8	2,3	60,5	7	23,3	7
Salerno	32	65	97	0	48	7	9	1	33	67	0	73,8	10,8	13,8	1,5
Sannio	3	18	21	0	10	2	5	1	14,3	85,7	0	55,6	11,1	27,8	5,6
Sassari	13	55	68	1	37	3	13	1	19,1	80,9	1,8	67,3	5,5	23,6	1,8
Siena	14	83	97	0	63	6	13	1	14,4	85,6	0	75,9	7,2	15,7	1,2
Siena Stranieri	1	2	3	0	2	0	0	0	33,3	66,7	0	100	0	0	0
Teramo	3	19	22	0	14	4	1	0	13,6	86,4	0	73,7	21,1	5,3	0
Torino	73	115	188	0	81	7	25	2	38,8	61,2	0	70,4	6,1	21,7	1,7
Torino Politecnico	84	30	114	0	11	2	12	5	73,7	26,3	0	36,7	6,7	40	16,7
Trento	19	45	64	2	34	4	1	4	29,7	70,3	4,4	75,6	8,9	2,2	8,9
Trieste	17	77	94	0	52	3	16	6	18,1	81,9	0	67,5	3,9	20,8	7,8

Institution	number of PhDs same Institution	number of PhDs other Institution	total number of accredited PhDs in VQR	number of PhDs accredited by Schools of advanced Studies	number of PhDs accredited by State universities	number of PhDs accredited by non-State universities	number of PhDs accredited by Research institutes	number of PhDs accredited by other institutions	Number of PhDs from the same institution % [compared to the total number of accredited researchers per institution]	number of PhDs from other institution % [compared to the total number of accredited researchers per institution]	number of PhDs accredited by Schools of advanced Studies% [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by State universities % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by non-State universities % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by Research institutes % [compared to the total number of accredited researchers by other institution]	number of PhDs accredited by other institutions % [compared to the total number of accredited researchers by other institution]
Trieste SISSA	0	20	20	0	16	0	4	0	0	100	0	80	0	20	0
Tuscia	8	38	46	0	21	3	11	3	17,4	82,6	0	55,3	7,9	28,9	7,9
Udine	10	71	81	0	56	9	5	1	12,3	87,7	0	78,9	12,7	7	1,4
Urbino Carlo Bo	4	30	34	0	28	1	1	0	11,8	88,2	0	93,3	3,3	3,3	0
Venezia Cà Foscari	16	51	67	0	32	6	9	4	23,9	76,1	0	62,7	11,8	17,6	7,8
Venezia Iuav	5	28	33	0	24	2	2	0	15,2	84,8	0	85,7	7,1	7,1	0
Verona	31	63	94	1	53	7	1	1	33	67	1,6	84,1	11,1	1,6	1,6

Preliminary analyses were also carried out to better understand whether there are significant differences in the quality of scientific production according to the geographical area in which the PhD degree was obtained and the area in which the researcher was working at the time of the VQR accreditation. According to preliminary results, researchers who have obtained their degree in the northern regions and who belong to northern universities are those who achieve the best scientific performance (higher probability of achieving an excellent assessment - A or B). However, researchers who obtained their degree in the southern regions and who are now affiliated to northern universities show a better performance, than those who have obtained their degree in the North and now work in the Centre or in the South. In other words, these preliminary analyses seem to show that in northern universities the probability of achieving a better assessment of research activity is higher, regardless of where the PhD programme was completed. This suggests that also those who obtained their degree in the South and now work in the North contribute to the average improvement of the institution.

2.3.4 Comparison between the evaluation results in the STEM+LS areas and the results of the relevant bibliometric algorithm

This section analyses the degree of agreement between the results of the bibliometric evaluation, carried out according to the algorithm described in the documents on GEVs' evaluation methods (available on

ANVUR's website: https://www.anvur.it/attivita/vqr/vqr-2015-2019/gev/) and the results of the peer review carried out by the expert evaluators. In this regard, it should be recalled that in VQR 2015-2019 the experts were provided with the results of the application of the bibliometric algorithm at the time of the evaluation. The analyses were carried out by comparing the evaluation results produced by the pure bibliometric algorithm with those resulting from the informed peer review approved by the GEVs. The results at issue refer to all submissions, including "duplicate products", i.e. products submitted for evaluation several times by different institutions or by several departments of the same institution, without disaggregating them according to the type of submitting institution. The underlying idea is that, in order to fully assess the impact of the two types of evaluation, it is appropriate to consider all the submissions on which these outcomes have actually spread. A distinction between subsets of institutions could be misleading, as the evaluations (especially those resulting from the algorithm) do not take into account the submitting institution, but spread equally across all duplicated submitted products.

Table 2.3.7 shows the percentage distribution (% values within the relevant GEV) and absolute distribution of the assessment categories resulting from the two types of evaluation: namely, that deriving from the application of the best outcome of the bibliometric algorithm and that referring to the final peer evaluation. It should be recalled that, according to the algorithm applied, category A included – on the basis of bibliometric data alone – the products ranking in the top 10% of the world distribution, obtained by combining the citation values of the published article and those of the indicator referring to the publisher; bibliometric category B included products ranking between 10 and 35% of the distribution, category C between 35 and 60%, category D between 60 and 80%, while category E included products from the bottom 20% of the world distribution.

Table 2.3.7 - Absolute distributions and percentages of assessment categories by GEV and evaluation methods: pure bibliometry resulting from the best outcome of the bibliometric algorithm and informed peer-review.

GEV	Evaluation method	Class assigned to the proposal					
		A	B	C	D	E	Total
1	Bibliometry	50,0% (3.920)	39,8% (3.121)	8,8% (692)	1,2% (98)	0,1% (10)	100,0% (7.841)
1	Informed peer-review	29,4% (2.308)	46,5% (3.648)	21,5% (1.685)	2,4% (187)	0,2% (13)	100,0% (7.841)
2	Bibliometry	78,5% (10.369)	19,2% (2.537)	1,7% (229)	0,3% (46)	0,2% (24)	100,0% (13.205)
2	Informed peer-review	42,7% (5.636)	39,5% (5.210)	15,9% (2.101)	1,7% (229)	0,2% (29)	100,0% (13.205)
3	Bibliometry	56,4% (5.021)	40,3% (3.585)	3,0% (265)	0,3% (30)	0,0% (3)	100,0% (8.904)
3	Informed peer-review	33,5% (2.984)	52,4% (4.662)	13,5% (1.203)	0,6% (53)	0,0% (2)	100,0% (8.904)
4	Bibliometry	45,8% (2.113)	41,0% (1.888)	10,1% (464)	2,7% (126)	0,4% (18)	100,0% (4.609)
4	Informed peer-review	35,2% (1.621)	44,3% (2.042)	17,0% (783)	3,0% (139)	0,5% (24)	100,0% (4.609)
5	Bibliometry	59,7% (10.093)	35,8% (6.057)	3,8% (639)	0,5% (90)	0,2% (27)	100,0% (16.906)
5	Informed peer-review	30,5% (5.151)	47,0% (7.946)	20,0% (3.387)	2,3% (384)	0,2% (38)	100,0% (16.906)
6	Bibliometry	65,5% (17.233)	29,3% (7.718)	4,0% (1.061)	0,9% (248)	0,2% (62)	100,0% (26.322)
6	Informed peer-review	15,8% (4.148)	35,6% (9.369)	37,4% (9.855)	9,1% (2388)	2,1% (562)	100,0% (26.322)
7	Bibliometry	48,1% (3.849)	43,9% (3.516)	6,8% (545)	1,0% (79)	0,1% (12)	100,0% (8.001)
7	Informed peer-review	31,5% (2.524)	48,3% (3.862)	18,0% (1.443)	2,0% (161)	0,1% (11)	100,0% (8.001)
8a	Bibliometry	47,6% (60)	38,9% (49)	11,1% (14)	0,8% (1)	1,6% (2)	100,0% (126)

GEV	Evaluation method	Class assigned to the proposal					
		A	B	C	D	E	Total
8a	Informed peer-review	43,7% (55)	39,7% (50)	15,9% (20)	0,8% (1)	0,0% (0)	100,0% (126)
8b	Bibliometry	53,8% (2.580)	40,2% (1.926)	5,1% (246)	0,8% (38)	0,1% (7)	100,0% (4.797)
8b	Informed peer-review	18,3% (878)	44,7% (2.146)	32,6% (1.565)	3,9% (187)	0,4% (21)	100,0% (4.797)
9	Bibliometry	56,9% (9.974)	37,1% (6.491)	5,2% (915)	0,7% (121)	0,1% (14)	100,0% (17.515)
9	Informed peer-review	38,7% (6.774)	37,8% (6.615)	20,6% (3.616)	2,7% (476)	0,2% (34)	100,0% (17.515)
11b	Bibliometry	51,2% (1.540)	39,4% (1.186)	6,6% (198)	2,2% (65)	0,7% (20)	100,0% (3.009)
11b	Informed peer-review	40,1% (1.206)	44,6% (1.342)	11,6% (348)	2,9% (87)	0,9% (26)	100,0% (3.009)

As also highlighted by Figure 2.3.5, apart from rare exceptions (notably GEV 8a) bibliometric evaluations tend to place in category A a share of products close to - if not higher than - 50% of those evaluated by each GEV. If products in category A are added to those in category B, the overall percentage of these first two categories (A + B) is close to 90%. In GEVs 2, 5 and 6, the overall percentages of the first two assessment categories identified by the algorithm are 98%, 96% and 95% respectively.

Overall, therefore, the pure bibliometric evaluation turns out to be more favorable than the results of the informed peer review. For some GEVs, especially 6, 5 and 8b, the final peer evaluation phase has reduced the share of products evaluated as A, with reductions that affected more than two thirds of the products that the algorithm would have placed in the highest category. Other GEVs, especially 11b, 8a and 4, agreed more with the bibliometric evaluation in the final phase of the informed peer review.

Figure 2.3.5 - Bar chart comparing the percentage distributions of the assessment categories assigned to the products by type of evaluation carried out by the GEV (B = Bibliometry; P = Informed peer review).
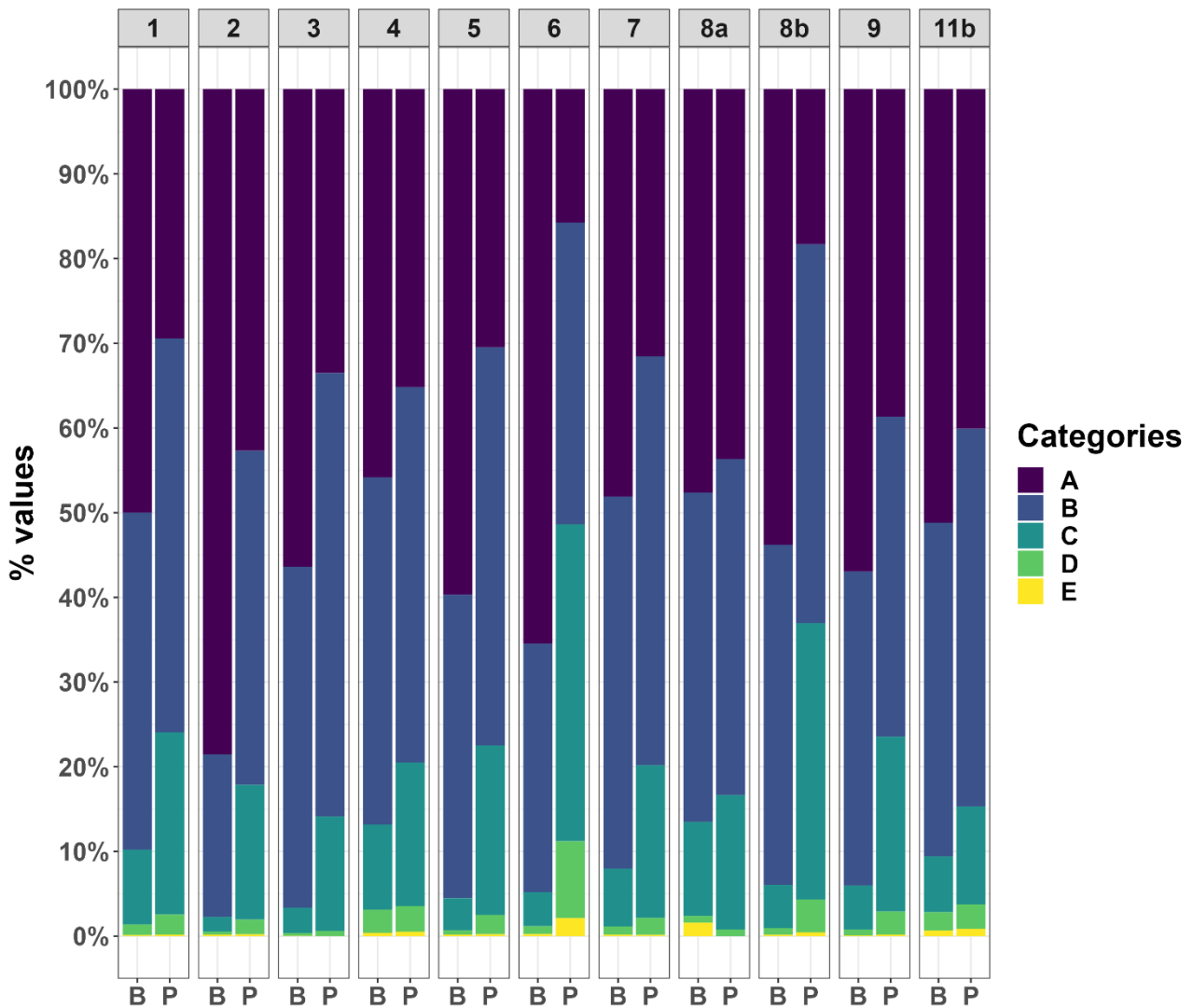
The different degree of agreement between the pure bibliometric evaluations and the informed peer evaluations is further illustrated in Table 2.3.8, which presents Cohen's Kappa values between these two types of evaluations for each GEV. Cohen's Kappa is a concordance coefficient that takes into account the probability of random concordance, varying between 0 (random concordance only) and 1 (maximum concordance). The concordance between the two types of evaluation is low in GEV 6 and modest in GEVs 2, 5 and 8b, while it is higher in GEVs 11b and 4, with the other GEVs having moderate levels of concordance.

Table 2.3.8 - Degree of agreement between bibliometric evaluation and informed peer review according to the evaluation area. Cohen's Kappa coefficient by GEV and percentage of products with n categories of difference between the evaluation resulting from the bibliometric algorithm and the informed peer review.

GEV	Kappa*	Lower limit Confidence interval	Upper limit Confidence interval	Total of products evaluated	0-category Difference %	1-category Difference %	2-category Difference %	3-category Difference %	4-category Difference %
1	0,490	0,476	0,505	7.841	60,74	37,06	2,12	0,08	0,00
2	0,272	0,261	0,282	13.205	51,66	42,23	5,88	0,20	0,03
3	0,455	0,442	0,469	8.904	64,16	35,00	0,84	0,00	0,00
4	0,720	0,704	0,736	4.609	76,91	22,76	0,26	0,00	0,07
5	0,373	0,364	0,383	16.906	53,83	42,29	3,76	0,11	0,02
6	0,138	0,134	0,143	26.322	22,69	51,22	23,81	1,94	0,35
7	0,567	0,553	0,581	8.001	69,48	28,65	1,80	0,05	0,02
8a	0,622	0,517	0,727	126	69,84	29,37	0,79	0,00	0,00
8b	0,249	0,234	0,265	4.797	37,92	52,91	8,88	0,27	0,02
9	0,461	0,451	0,470	17.515	62,40	33,31	4,12	0,14	0,03
11b	0,762	0,743	0,782	3.009	81,85	17,61	0,47	0,07	0,00

* negative Kappa values indicate that the concordance between the two types of evaluation is less than that expected as a result of random chance; values equal to 0 indicate that the degree of concordance observed is equal to that which would be obtained as a result of random chance; values between 0,01 and 0,20, indicate poor concordance; values between 0,21 and 0,40, modest concordance; values between 0,41 and 0,60, moderate concordance; values between 0,61 and 0,80, good concordance; values between 0,81 and 1, excellent concordance; Kappa values equal to 1 indicate perfect agreement between the two types of assessment.

2.3.5 Convergence and divergence measures of peer evaluation

This section analyses the degree of correlation between the two initial product assessments. It should be recalled that the VQR evaluation was carried out in two stages: in the first one, two experts independently assessed the product and assigned a score expressed in tenth to each of the three evaluation criteria (originality, methodological rigour and impact); overall, there was therefore a score in thirtieth expressed by each of the two evaluators. The evaluation could be carried out by two GEV members expert in the relevant disciplinary field, that is – where possible – belonging to the same Academic Discipline or a related one, or by external experts selected by them. Table 2.3.9 firstly shows that in some areas the use of external or mixed evaluation has been much more extensive than in other areas; for example, in Area 4, 96% of products were evaluated internally, while in Area 8a only 1,4% of products were evaluated internally. Table 2.3.9 also shows markedly high levels of correlation between the two reviewers in charge of evaluating the product, especially if the evaluation was entrusted to two GEV members; GEVs 3 and 11a represent an exception with a low degree of correlation, below 0,5 (see Table 2.3.9). On the other hand, the correlation between the scores assigned by the two reviewers is lower when one of the two referees is external to the GEV, and drops further when both reviewers are external.

Table 2.3.9 - Correlation between the overall scores attributed by the assigned reviewers (rev1 and rev2) to the products evaluated, by GEV and type of evaluation (completely entrusted to reviewers external to the GEV; completely entrusted to reviewers internal to the GEV; mixed). High correlations (i.e., with coefficient higher than 0,7) are highlighted in green. Evaluations made by the different pairs of reviewers, if referring to less than 100 products, are highlighted in red.

GEV	Correlations rev1 vs rev2 (external)	Correlations rev1 vs rev2 (internal)	Correlations rev1 vs rev2 (mixed)	Products evaluated by external experts	Products evaluated internally	Products evaluated with mixed evaluation (internal/external)
1	0,220	0,702	0,327	305	6.890	1.799
2	0,195	0,653	0,287	532	10.302	4.588
3	-0,033	0,392	0,124	84	8.948	1.071
4	0,904	0,793	0,224	< 10	5.460	223
5	0,996	0,682	0,191	< 10	16.923	305
6	0,635	0,690	0,353	38	25.300	1.499
7	0,161	0,708	0,160	242	7.755	2.058
8a	0,247	0,876	0,312	4.483	75	777
8b	0,248	0,655	0,387	702	2.222	1.964
9	0,384	0,733	0,394	329	15.629	3.010
10	0,280	0,627	0,362	8.461	847	4.118
11th	0,290	0,495	0,303	2.702	1.748	4.482
11b	0,432	0,850	0,441	86	3.689	73
12	0,251	0,788	0,433	9.076	522	3.538
13a	0,377	0,803	0,315	558	3.895	3.091
13b	0,456	0,913	0,583	1.668	2.023	2.150
14	0,236	0,729	0,405	1.723	1.017	2.293
All	**0,310**	**0,722**	**0,392**	**30.996**	**113.245**	**37.039**

2.3.6   Open access products: a typological analysis by subject area

The VQR Call, in compliance with art. 1, paragraph 3 of the Ministerial Guidelines, states that the products evaluated in the framework of the 2015-2019 VQR exercise must be made available to the public in open access (OA), if financed for a share equal to or greater than 50% with public funds; this provision also applies to scientific products published in open access journals. In other cases, the publication of the metadata is required. Overall, out of the 181.280 products submitted for evaluation, over 54% were made available to the public in open access; information can be found on the ANVUR website at https://www.anvur.it/attivita/vqr/vqr-2015-2019/elenco-dei-prodotti-e-dei-casi-di-studio-valutati/. In 22% of cases, universities have indicated that the product is still subject to an embargo, while a further 22% of the products were not made available because they were not at least 50% publicly funded. In a very limited number of cases, no information was provided by the institutions regarding the open access possibility of the submitted products (cf. Figure 2.3.6).
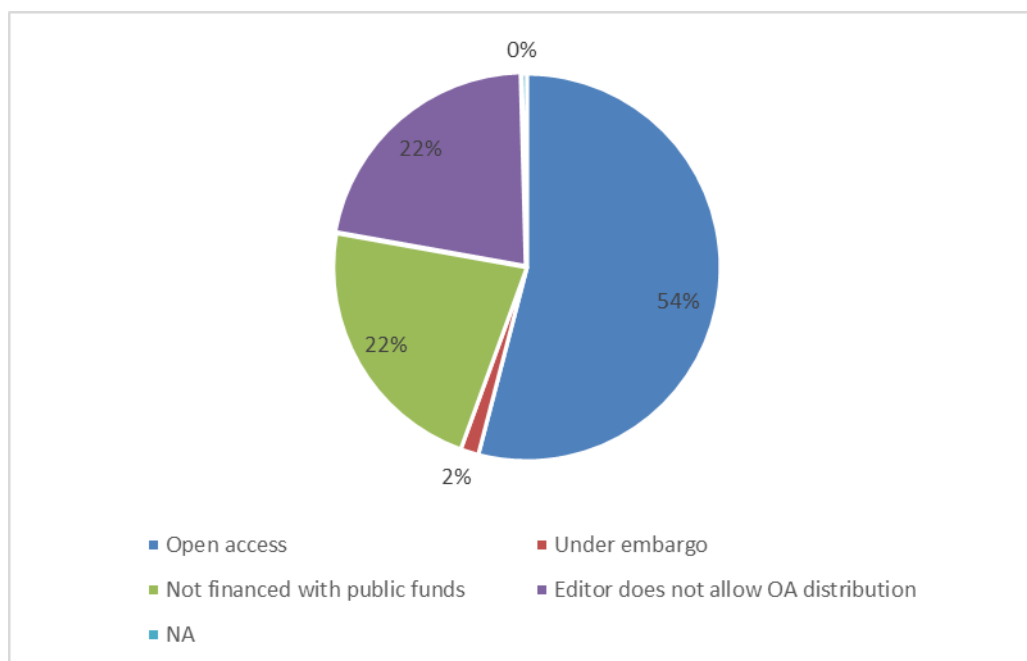
It should be noted that, according to the Call, products considered to be open access could be made available in at least one of the following channels:
- a) in an open access journal or volume,
- b) in a university open repository,
- c) in a disciplinary open repository,
- d) as a working paper,
- e) on personal websites.

Furthermore, the Call stated that the product had to be available in one of the following versions:
- a) version of Record,
- b) author Accepted Manuscript,
- c) version sent to the journal for publication.

Based on the information provided by the universities, it is not possible to classify open access products according to the access mode or to the version of the document available. Instead, it is possible to analyse in detail the open access availability of the products submitted to the VQR, looking in particular at how the open access mode is distributed at the level of disciplinary area, type of institution and type of product, as well as the evolution of open access availability over the five-year evaluation period. A first analysis of the relationship between evaluation results and the products' open access availability will conclude this part of the analysis.

Figure 2.3.7 shows the distribution of open access products by subject area. The share of products available in open access varies greatly by discipline, ranging from a maximum of 86,7% in physics to a minimum of 22% in law studies. In general, the share of products available in open access is higher in the STEM+LS areas, lower in the social sciences, even lower in humanities and law studies.

Figure 2.3.7 - Availability of open access products by disciplinary area.
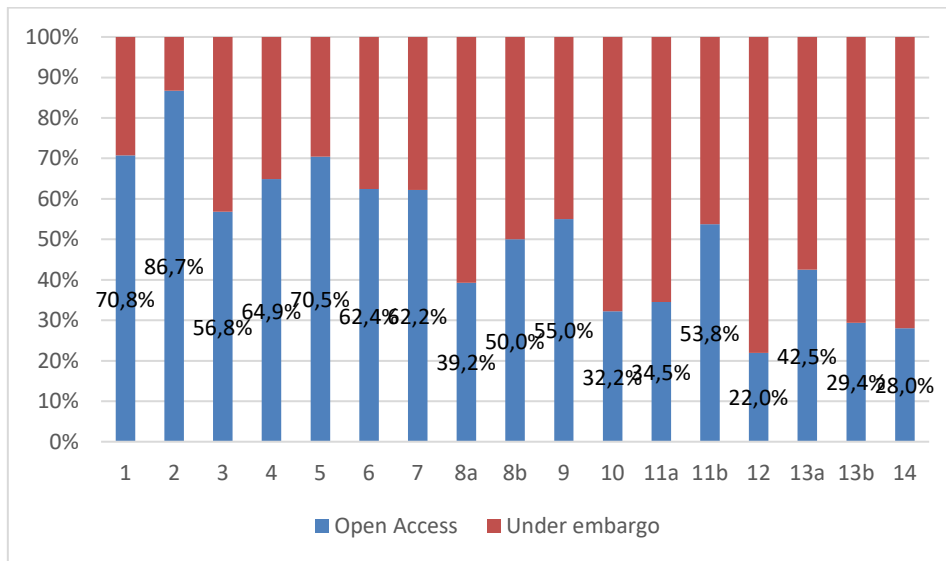
Differences among areas in the share of open access products are probably also due to the different types of publication prevailing in the areas themselves: indeed, it is likely that in the areas where the majority or almost all of the products submitted are journal articles, the share of open access products is higher than in the areas where a significant share of publications is represented by book chapters and monographs. Figure 2.3.8 confirms that the share of OA products is particularly high for journal articles and considerably lower for the other types of publication.

Figure 2.3.8 - Availability of open access products by type of publication.
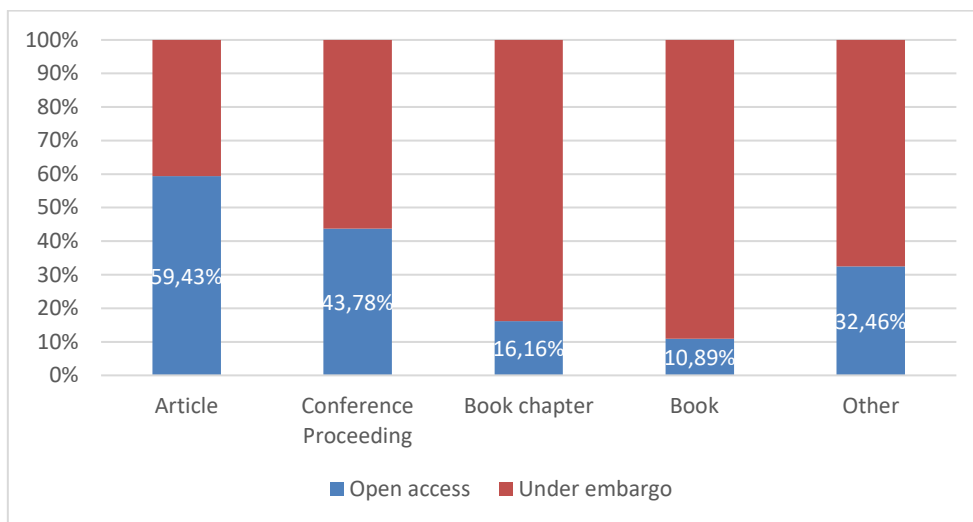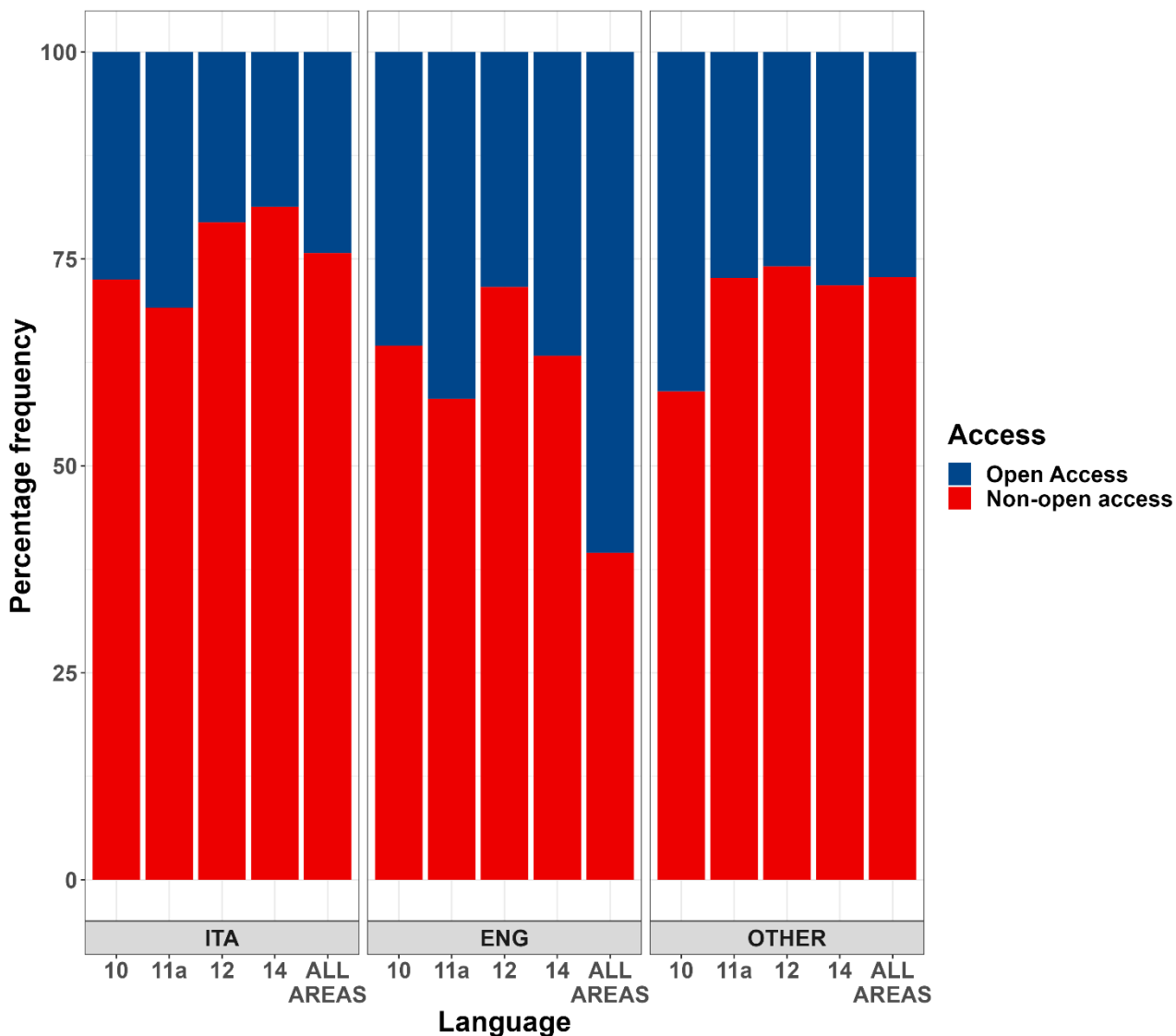

Figure 2.3.9 focuses on the areas where the share of OA products is lowest, showing the presence of open access products depending on the language of the publication.

It can be noted that the share of open access products is systematically higher for publications in English than in Italian and other languages. This suggests a low uptake of open access publication methods by national publishers.

Figure 2.3.10 illustrates the availability of open access products in relation to the type of institution, distinguishing between universities, research institutes, and institutions that participated in the VQR on a voluntary basis.

Figure 2.3.10 - Availability open access products by type of institution.

Figure 2.3.10 - Availability open access products by type of institution.
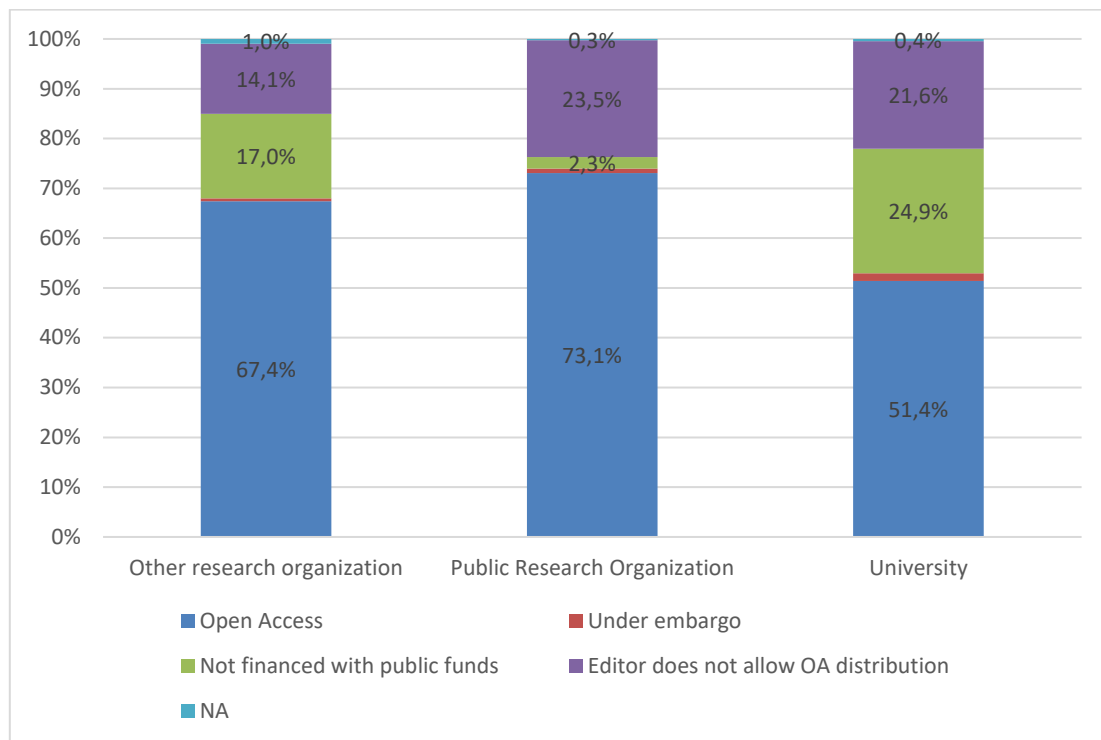
On average, the diffusion of open access products is higher in Public Research Institutes (EPR, which stands for "Enti Pubblici di Ricerca") than in universities. Among the causes of lack of open access products, one may note that in universities research that is not at least 50% publicly funded prevails. By contrast, this factor is of little relevance to research institutes, where the main reason for not having open access to publications is the persistence of publication embargoes. Even in the institutions that participated in the VQR on a voluntary basis – which, however, make up only 1,1% of the evaluated products – open access to publications is more widespread than in universities, reaching 67,4% of the evaluated products.

Figure 2.3.11 presents the evolution of the availability of open access products over time.

**Figure 2.3.11 - Availability of open access products by year.**

Figure 2.3.11 - Availability of open access products by year.
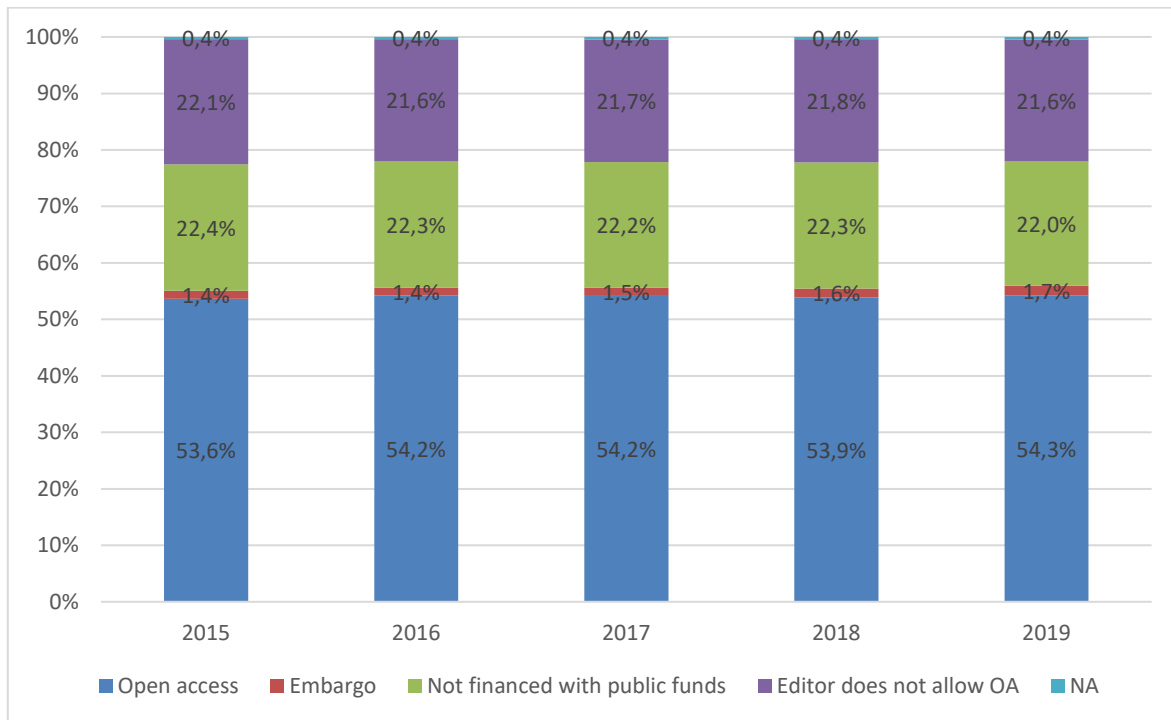
No significant differences emerge across years, with the share of open access products remaining fairly constant around 54% of the products submitted to the VQR, with small fluctuations from one year to the other.

The criteria defined in the VQR Call do not assign any role for evaluation to the availability of open access products. Nonetheless, it is possible to investigate with hindsight whether the products' availability in open access has any relationship with the evaluations obtained. However, any correlation should in no way be interpreted so as to argue for any causal link between the open access of scientific work and the evaluation results. These are first pieces of evidence that should be further investigated in the future (see also the section containing the multivariate analysis of VQR data). This partial analysis of the data shows that the presence in the excellence categories (A and B) is slightly higher for open access products than the average (29,2 against 25,9 for category A and 40,1 vs. 39,2 for category B, cf. Figure 2.3.12).

Figure 2.3.12 - Open access and evaluation results.
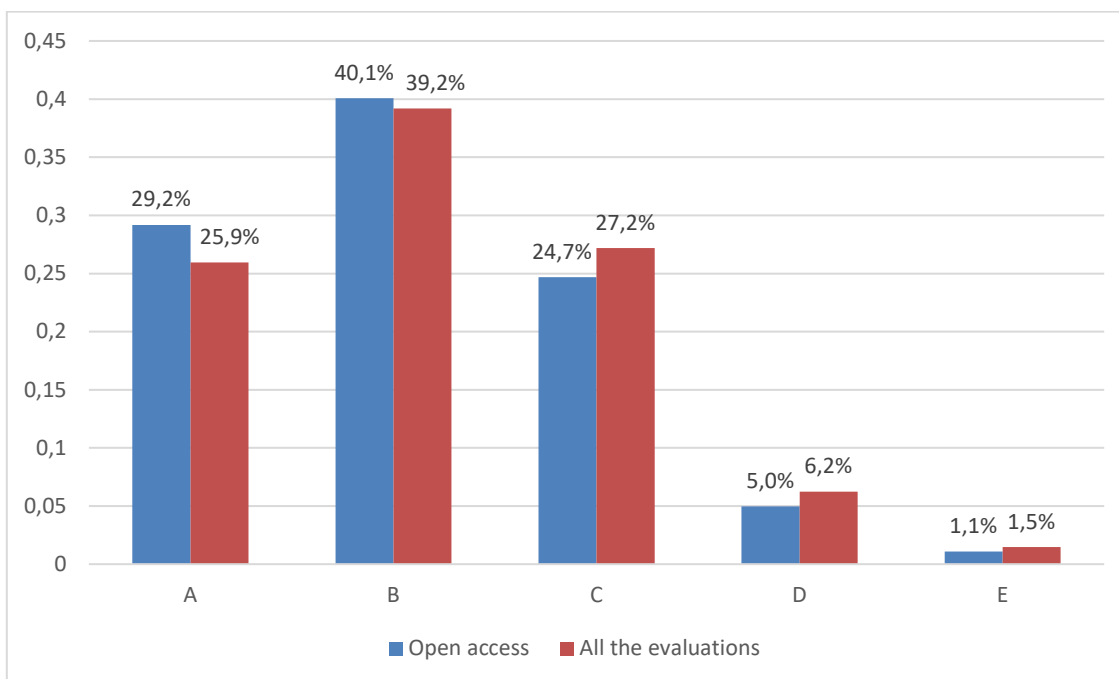
Below some of the main findings from this section of the report:

- Over 50% of the products submitted to the VQR are available in OA.
- There are significant differences between areas and types of publication.
- Open access is more widespread among research institutes, than among universities.
- There are no hints of a temporal evolution of the OA diffusion over the VQR period.
- On average, OA products are assessed slightly more favourably than the general average.
- The definitions provided in the Call as regards the methods of products' dissemination in OA and the version to be submitted may be reviewed in the light of the evolving international debate on the subject.

2.3.7 Some consideration for the future

Let us conclude this section with some comments on the VQR 2015-2019. The VQR 2015-2019 was certainly characterised by several innovations compared to previous evaluation exercises. Informed peer review represented an important and substantial difference with respect to the previous VQRs, reducing the central role played by bibliometric evaluation in the VQR 2015-2019. This new evaluation method required the involvement of a considerable number of experts to carry out the peer review. In this context, it would be advisable for the future to develop a database of experts to cover all subjects treated by the products evaluated for the VQR and to increase the number of international experts, also in cooperation with other international institutions. An important aspect, which deserves further investigation, concerns the different evaluation between the STEM+LS areas (based on bibliometrically informed peer-review) and SSH areas (based on peer review).

Another consideration concerns the possibility for the institutions to submit for each researcher a flexible number of products, varying between zero and four. Such innovation may have had important consequences on the evaluation results themselves. Finally, the requirement for open access products has emphasised the importance of publicly funded research products being accessible to the entire scientific community and, more broadly, to the entire population, while of course respecting the rules to protect publishers. More generally, open science activities should be enhanced, with open sharing of results to increase collaboration and innovation in scientific research.

2.4 Analysis of the VQR 2015-2019 results regarding the evaluation of Third Mission case studies

2.4.1 The Third Mission and Impact Assessment in VQR 2015-2019

The Quality Assessment of Research (VQR) 2015-2019 is the first of the three VQRs which envisaged the evaluation of Third Mission (hereafter, TM) activities (art. 2 paragraph 3 of the VQR Call). Already in the previous VQR 2011-2014, the Third Mission was included in the evaluation exercise, but only in an exploratory form and through the survey of all TM activities carried out by universities and research institutes. The results did not affect the distribution of the ministerial funding.

In the VQR 2015-2019, on the other hand, the impact of TM activities is mainly evaluated through the analysis of case studies (art. 9 of the VQR Call). Universities and research institutes were invited to submit for evaluation a number of case studies related to Third Mission activities, which was proportional to the number of departments of the institution, and whose impact has been verified in the period 2015-2019.

The use of the case study methodology makes it possible to bring out the institutional and territorial specificities of universities and research institutes, to enhance best practices and to spread new knowledge and approaches to TM activities. In this perspective, the definition of impact is deliberately broad and aims to give each institution the opportunity to highlight the most relevant TM initiatives from an economic, social and cultural point of view. Such change in methodological approach, compared to the previous VQR exercise, was strongly demanded by the academic community and sanctioned by the Ministry of Universities and Research in the decrees launching the third VQR exercise (Ministerial Decree no. 1110/2019 and subsequent integration contained in Ministerial Decree no. 444/2020, 11 August 2020).

The case studies concern activities whose impact has been verified in the following areas:

- *a) Intellectual and industrial property valorisation (patents, plant varieties and any other product referred to in article 2, paragraph 1, of Legislative Decree no. 30/2005);*
- *b) Academic entrepreneurship (es. Spin-offs, start-ups);*
- *c) Technology transfer structures and other Third mission intermediaries (e.g. technology transfer offices, incubators, science and technology parks, consortia and associations for the Third Mission);*
- *d) Production and management of artistic and cultural heritage (e.g. museum hubs, archaeological excavations, musical activities, historical buildings and archives, historical libraries and newspaper libraries, theatres and sports facilities);*

- *e) Clinical experimentations and health protection (e.g. clinical trials, studies on medical devices, non-interventional studies, biobanks, patients' empowerment initiatives, veterinary clinics, information and prevention days, screening and awareness-raising campaigns);*
- *f) Lifelong learning and open education (e.g. continuous education courses, Continuing Medical Education courses, MOOCs);*
- *g) Public engagement activities;*
- *h) Production of public goods and policy instruments for inclusion (e.g. development of public interest programmes, participation in urban development and land-use projects or in participatory democracy initiatives, consensus conferences, citizen panels);*
- *i) Innovative tools in support of Open Science;*
- *l) Activities related to the 2030 UN Agenda for Sustainable Development Goals (SDGs).*

This types of field of action partly refer to the Guidelines for filling in the Single Annual Report on Research and Third Mission/Social Impact of 7 November 2018 (SUA-TM 2018), which follow the approach adopted by the TM Evaluation Handbook (ANVUR 2015) and in the previous VQR. In particular, in the VQR 2011-2014, on the basis of the guidelines contained in the ANVUR 2015 Handbook, two broad areas of TM activities were identified, "Research valorisation" and "Production of public goods and policy instruments for inclusion", each of which was divided into four areas of intervention.

Table 2.4.1 – Structure of SUA-TM/IS (2018).

I.0 – Third Mission/Social Impact strategic objectives	
RESEARCH ENHANCEMENT	PRODUCTION OF PUBLIC GOODS
I.1 Intellectual property management (patents and plant varieties)	I.5 Cultural heritage and activities management (archaeological excavations, museum hubs, musical activities, historical buildings and archives, historical libraries and newspaper libraries, theatres and sports facilities)
I.2 Spin-off companies	I.6 Public health activities (clinical trials, non-interventional studies and patients' empowerment, support facilities)
I.3 Third-party activities	I.7 Lifelong education, lifelong learning and open education (lifelong learning, Continuing Medical Education, Skills Certification, School-Work Alternance, MOOC)
I.4 Third mission intermediaries (technology transfer offices, placement offices, incubators, science parks, consortia and associations for the Third Mission)	I.8 Public Engagement

In the VQR 2015-2019, this broad notion of Third Mission is therefore maintained. Indeed, 7 out of the 8 areas envisaged in the ANVUR Guidelines are taken up and the scope is further widened with the introduction of the fields of action "Production of public goods and policy instruments for inclusion", "Innovative Tools in support of Open Science" and "Activities related to the 2030 UN Agenda for Sustainable Development Goals".[2]

---

[2] The VQR 2015-2019 does not include the activity carried out on behalf of third parties because it is considered, rather than a specific TM activity of, a cross-cutting tool for the various fields.

In the VQR 2011-2014 the evaluation of the Third Mission performance of the evaluated institutions was carried out through informed peer review, i.e. using experts' assessment in combination with a set of indicators and metrics built according to the guidelines of the ANVUR Manual and making use of the data collected through SUA-RD Part Three (SUA-TM), see Table 2.4.1. The evaluation was therefore based on a survey approach, aimed at monitoring all Third Mission activities carried out by the institutions.[3]

The VQR 2015-2019 marks a discontinuity with the previous approach, not as regards the definition of Third Mission, whose broad meaning is confirmed and is consistent with the previous approach, but especially as far as evaluation methodology is concerned, which is completely new. Indeed, it does not consider the whole set of activities carried out, but a selection of case studies chosen by the evaluated institutions, which concern actions and activities that generated an impact during at least part of the 2015-2019 period.

2.4.2 The case studies

The VQR Call, under art. 9 paragraph 1 states that, for the purposes of evaluating Third Mission activities, universities are required to submit for evaluation a number of case studies equal to half of the number of their Departments; research institutes and institutions participating in the VQR on a voluntary basis may submit a number of case studies equal to the number of their own departments or similar units. Moreover, the case studies can be referred to the Departments or similar units, up to a maximum of two per department, or to the entire institution.

The case studies are related to TM activities, carried out in one or more of the fields of action, whose impact can be verified in the period 2015-2019. In particular, they must refer to actions and activities that can have taken place both before and during the 2015-2019 period, but that must have generated an impact in part or in the entire 2015-2019 period. Case studies that envisage a future or potential impact, or one which falls outside the evaluation period, are not considered eligible for evaluation.

The case study consists of a descriptive report of a TM activity carried out in one of the above-mentioned fields of action. The description covers not only the activities, but also the conditions under which they were carried out, with special regard to the external and internal context, the role played by the institution, the development over time, the subjects involved and their role, the resources employed and, more generally, all the elements useful to qualify the actions undertaken[4].

By impact is meant the transformation or improvement that (possibly in connection with the results of scientific research produced by the institution) has been generated for the economy, society, culture, health, the environment or, more generally, for the prevention of economic, social and territorial inequalities to increase the quality of life in a certain area (local, regional, national, European or international). It also involves the reduction or prevention of damage, risks, or other negative external circumstances. Priority in evaluation is given to the impact generated externally (including any spillovers within the institutions). Moreover, besides a detailed description of the activities and the conditions under which they took place, the case study should include a description of the impact generated, with special attention to the difference between the situation before the activity was carried out and the situation after. Such difference must be

---

[3]Further details on the Third Mission evaluation in the VQR 2011-2014 can be found at the following link: https://www.anvur.it/download/rapporto-2018/ANVUR_Rapporto_Biennale_2018_Sezione_9.pdf.

[4] The format for the submission of case studies can be found in Annex 2 of the VQR Call 2015-2019.

verifiable through the proposal, on the part of the institutions, of relevant impact indicators, which may be self-certified, derived from internal or external monitoring activities or from consultation with recipients and stakeholders. Besides selecting the case studies, the institutions were therefore asked to choose and propose indicators relevant for their evaluation.

2.4.3   The evaluation process

As required by art. 3, paragraph 4, of Ministerial Decree no. 1110/2019, the evaluation of the case studies was carried out by a Group of 30 Third Mission Experts (TM Interdisciplinary GEV), which complements the 17 disciplinary GEVs. The experts were chosen by ANVUR's Governing Board[5] from among over 300 candidates who responded to two public calls[6], one reserved for researchers and the other open to stakeholders from outside the evaluated institutions. This arrangement was designed to have a balanced interaction between highly qualified expert professors and figures representing public management, the productive and financial world, cultural institutions and territories.

The TM GEV coordinator was selected by the ANVUR Governing Board from among the GEV members. The TM GEV's evaluation was expressed through a quality assessment of each case study submitted by the institutions, taking into account the four criteria provided for by the VQR Call (art. 9, paragraph 5):

a) *Social, economic and cultural dimensions of the impact;*
b) *Relevance to the context of reference;*
c) *Added value for recipients;*
d) *Contribution to the submitting institution, emphasising the scientific aspect where relevant.*

Having completed the phase of submission of the case studies by the evaluated institutions, the TM GEV assigned the case studies to be evaluated to the individual members of the group and then proceeded to their collegial evaluation. According to the quality assessment, each case study was classified by the GEV in one of the following five categories: *A - Excellent and extremely relevant, B - Excellent, C - Standard, D - Sufficient relevance, E - Low relevance or not eligible.*

The results of the evaluation, expressed through the above-mentioned categories, provided a quality profile of the Third Mission activities of the institution, although not differentiated by field of action or by department, as stated in the Call (art. 10, paragraph 1d). This choice was recommended by the TM GEV itself, which opted to provide the evaluation only at the institutional level.

---

[5] For further details on the definitions and evaluation criteria, please refer to the "Document on how to evaluate case studies Expert Group of the Interdisciplinary Impact/Third Mission Assessment" available at the following https link://www.anvur.it/wp-content/uploads/2021/02/Document-GEV-TM.pdf.

[6] GEV - ANVUR - National Agency for the Evaluation of the University and Research System.

2.4.4 The evaluation results

The assessment categories expressed by TM GEV were later converted into scores based on the provisions of the Ministerial decree concerning the General Guidelines for Universities Programming 2021-2023 and Indicators for the periodic evaluation of results (Ministerial Decree no. 289/2021).

The decree establishes the criteria for the allocation of the bonus share of the Regular Fund for universities (Art. 6) and, starting from the year 2022, states that the 60% of the bonus share will be allocated on the basis of the results of the VQR 2015-2019, assigning a weight to each of the quality profiles of the institutions set out in the ANVUR Call (paragraph 2):

a) quality profile of permanent staff and recruitment policies - weight 90 %
b) quality profile of training in research - weight 5%
c) quality profile of research enhancement activities (Third Mission) - weight 5%.

The decree then sets out the use of an indicator for the distribution constructed as a weighted average of the indicators derived from profiles a), b) and c), as well as the criteria for converting these profiles into indicators, by defining the scores to be attributed to the five assessment categories, as shown in Table 2.2.2.

2.4.5 Analysis of the VQR 2015-2019 results related to the evaluation of case studies

This section presents an in-depth analysis of the characteristics of the case studies submitted and of the institutions which submitted them, as well as their relationship with the evaluation results; such analysis may provide useful guidance for the future of the evaluation process.

In particular, the analysis will focus on the fields of action of the case studies and the characteristics of the evaluated institutions (type and, in the case of universities, size and geographical area). The number of case studies submitted by the institutions, according to the provisions of the VQR Call, will then be examined in relation to the number of departments and the size of the institution.

Overall, the number of case studies submitted for evaluation by the universities and research institutes participating in the VQR was 676. Article 9, paragraph 3 of the VQR Call establishes that the case studies refer to actions whose impact can be verified in the fields of action set out in the same article and shown in Figure 2.4.1. Assuming an equal distribution of the fields of action, we would have about 68 case studies per field of action, i.e. an average percentage of 10% of the fields of action. On the contrary, as Figure 2.4.1 shows, just over a third of the case studies referred to the field of action *g) Public Engagement activities* (33%), while only 1 % of the cases refers to the field of action *i) Innovative tools in support of Open Science.* These two fields of action, placed at the poles of the distribution, represent outliers, deviating significantly from the average. The other fields of action are more evenly distributed, ranging from 11% of field *c)* to 6% of field *f)*.

Figure 2.4.1 - Distribution of the fields of action of the case studies (total number 676).
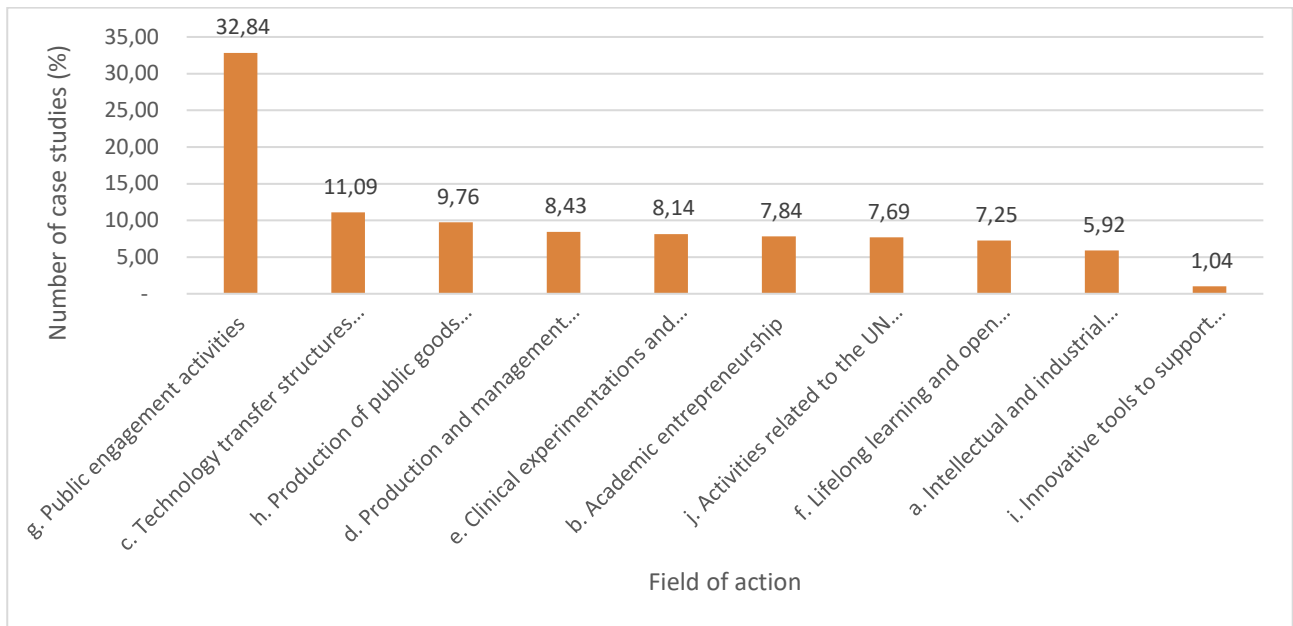
Institutions could also combine the main field of action with a maximum of two other fields of action. Considering all the attributions of the fields of action (both the main and the associated one) to the case studies, we obtain 1.040 choices (see Figure 2.4.2). Again, the field of action *g) Public engagement activities* is mentioned in one third of the cases (32%) while field *i) Innovative Tools in support of Open Science* features the lowest percentage ever (2%). On the other hand, a higher percentage is recorded for the field *h) Production of public goods and policy instruments for inclusion*, the field *j) Activities related to the UN Agenda 2030 for Sustainable Development Goals (SDGs)* and the field *f) Life-long learning and open education.*

Figure 2.4.2 - Distribution of the choices considering all the attributions of the fields of action (total number 1.040).

Figure 2.4.2 - Distribution of the choices considering all the attributions of the fields of action (total number 1.040).
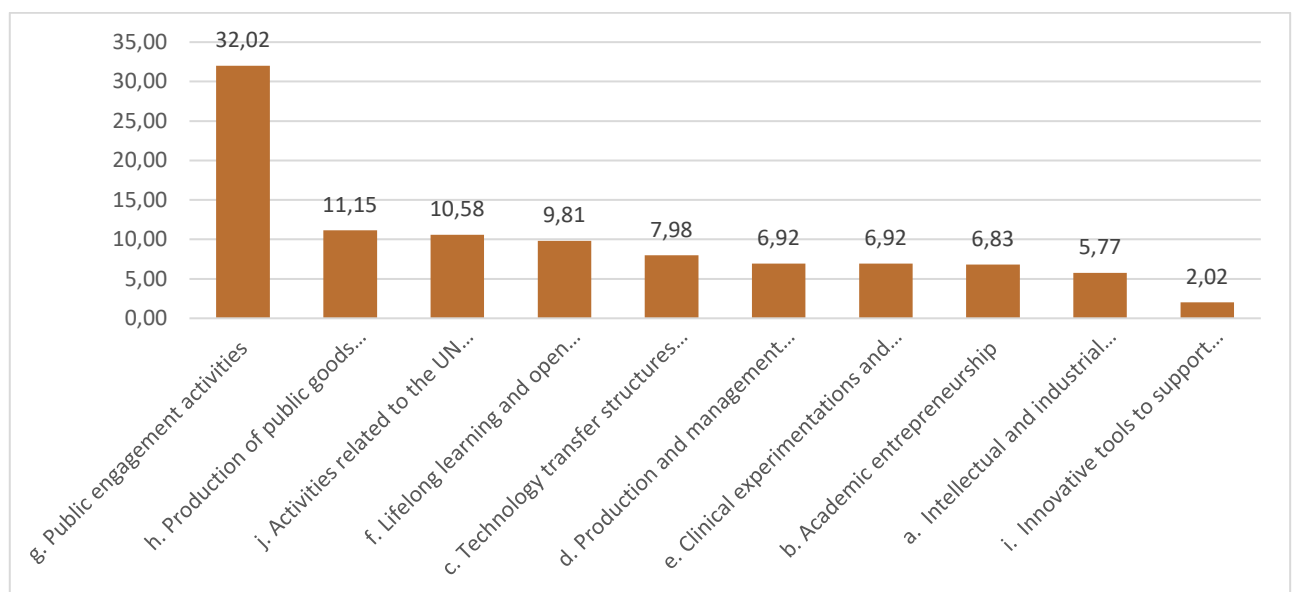
Table 2.4.2 shows how, in percentage, the fields of action are associated with each other in the case studies (co-occurrence). On the diagonal we find the percentage of the main fields of action, with no additional associated fields.

Fields *i) Innovative tools in support of Open Science*, e*) Clinical experimentations and health protection* and *b) Academic Entrepreneurship*, were presented with no associated fields more than others. Conversely, the fields *d) Production and management of artistic and cultural heritage*, *j) Activities related to the UN 2030 Agenda for Sustainable Development Goals (SDGs)* and *h) Production of public goods and policy instruments for inclusion* are those that most have been associated with other fields of action, especially the field *g) Public engagement activities*.
It should be noted that field *g) Public engagement activities* co-occurs with almost all fields (column g) and significantly with field *d) Production and management of artistic and cultural heritage* and with field f) *Life-long learning and open education.* With field *f)* there is a mutual co-occurrence, as the case studies that have the field *g) Public engagement activities* as their main field of action, co-occur with the field *f)* in 10% of the cases.

Table 2.4.2 - Co-occurrence of choices made on fields of action; coloured cells indicate percentage of case studies in those fields of action equal to or greater than 10%.

Main field of action	to.	b.	c.	d.	e.	f.	g.	h.	i.	j.	Total
a. Intellectual and industrial property valorisation	**64,5**	14,5	6,5	0,0	4,8	0,0	6,5	0,0	1,6	1,6	**100,0**
b. Academic entrepreneurship	10,8	**71,6**	4,1	0,0	5,5	0,0	4,1	0,0	0,0	4,1	**100,0**
c. Technology transfer structures and other Third mission intermediaries	7,7	6,0	**64,1**	0,9	0,9	1,7	12,8	0,0	1,7	4,3	**100,0**
d. Production and management of artistic and cultural heritage	0,0	2,0	0,0	**55,9**	0,0	1,0	30,4	4,9	0,0	5,9	**100,0**
e. Clinical experimentations and health protection	2,7	0,0	0,0	0,0	**73,3**	4,0	9,3	4,0	0,0	6,7	**100,0**
f. Lifelong learning and open education	0	0	0	1,4	2,7	**67,1**	20,5	5,5	0,0	2,7	**100,0**
g. Public engagement activities	0,0	0,0	0,0	3,0	0,9	10,2	**66,7**	8,1	2,1	9,0	**100,0**
h. Production of public goods and policy instruments for inclusion	0,0	0,0	0,0	1,9	1,9	7,5	17,8	**61,7**	2,8	6,5	**100,0**
i. Innovative tools in support of Open Science	0	0	0	0	0	0	12,5	0	**87,5**	0	**100,0**
j. Activities related to the 2030 UN Agenda for Sustainable Development Goals (SDGs)	1,1	0,0	1,1	1,1	2,2	5,6	18,0	12,4	0,0	**58,4**	**100,0**

It is now possible to analyse the evaluation results within the fields of action, expressed both through the score obtained and through the corresponding assessment category. Figure 2.4.3 shows the distribution of the average scores in the various fields of action. The score is determined by the ratio between the total score obtained by the case studies in that field of action and the number of case studies submitted for each field of action. Overall, the average score assigned to the case studies was 0,64 (the score ranges from 0 to 1). Compared to this value, fields *b), e), d), a)* and *c)* have an above-average score. It should be noted that all the fields related to research valorisation are present (for the definition, see the SUA-TM 2018 Guidelines). Conversely, new fields *h), i),* and *j),* and the remaining *f)* and *g)* are below the average score.

Figure 2.4.3 - Average score distribution by field of action and total[7].
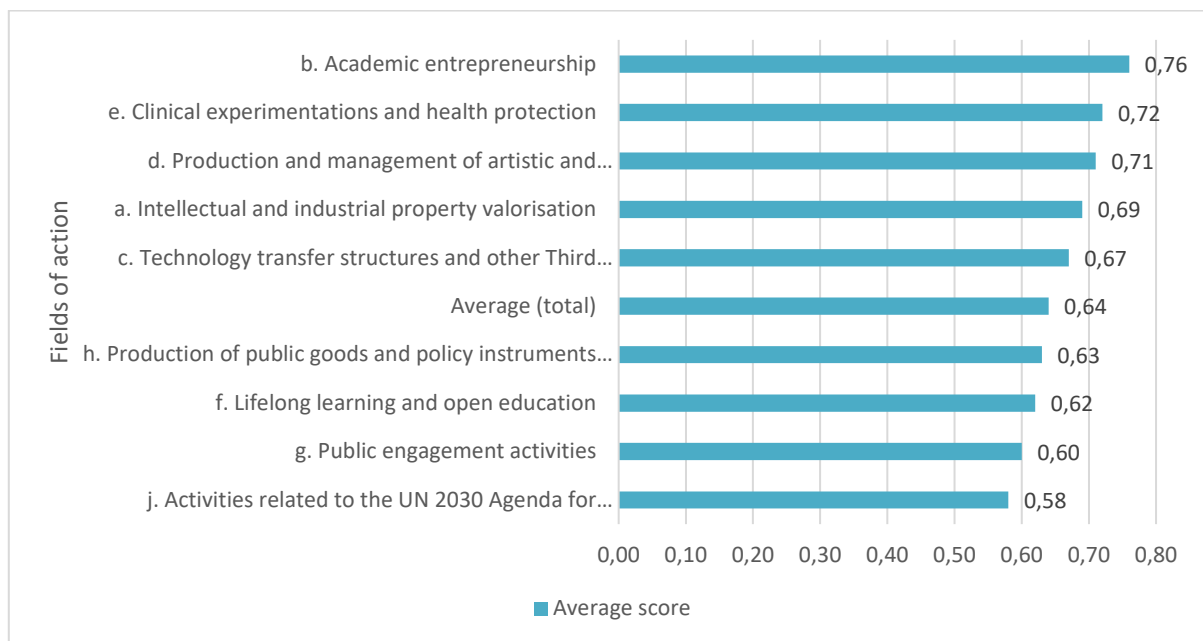
Let us now consider the evaluation results in relation to the assessment categories attributed to the case studies, by field of action. As Figure 2.4.4 shows, overall, the evaluators tend to place in category B (Excellent) a little more than 49% of the submitted cases studies. By adding category A (Excellent and extremely relevant) to category B, so as to consider excellence as a whole, the joint percentages of these first two categories (A + B) result in about 61%. The field which received the highest number of excellent evaluations (A+B) is the Field b) *Academic entrepreneurship*, with about 80% of cases falling into these two assessment categories. It also features the highest percentage of case studies assessed as A (32%). Fields *d), e),* and *a)* follow; the first two show the highest percentages of B, while the field *a)* shows the highest percentage of A, after field *b)*. The fields that feature lower percentages of A+B evaluations are *i), j), f)* and *g).* It should be noted that field *j) Activities related to the UN Agenda 2030 for the Sustainable Development Goals (SDGs)* shows a very high percentage of case studies evaluated as A (17%), but also the highest percentage of case studies evaluated as E (19% compared to an average of 6% for this assessment category).

---

[7] The field of action i) Open science is not reported because fewer than 10 case studies were presented.

Figure 2.4.4 - Distribution of assessment categories by field of action and total[8].
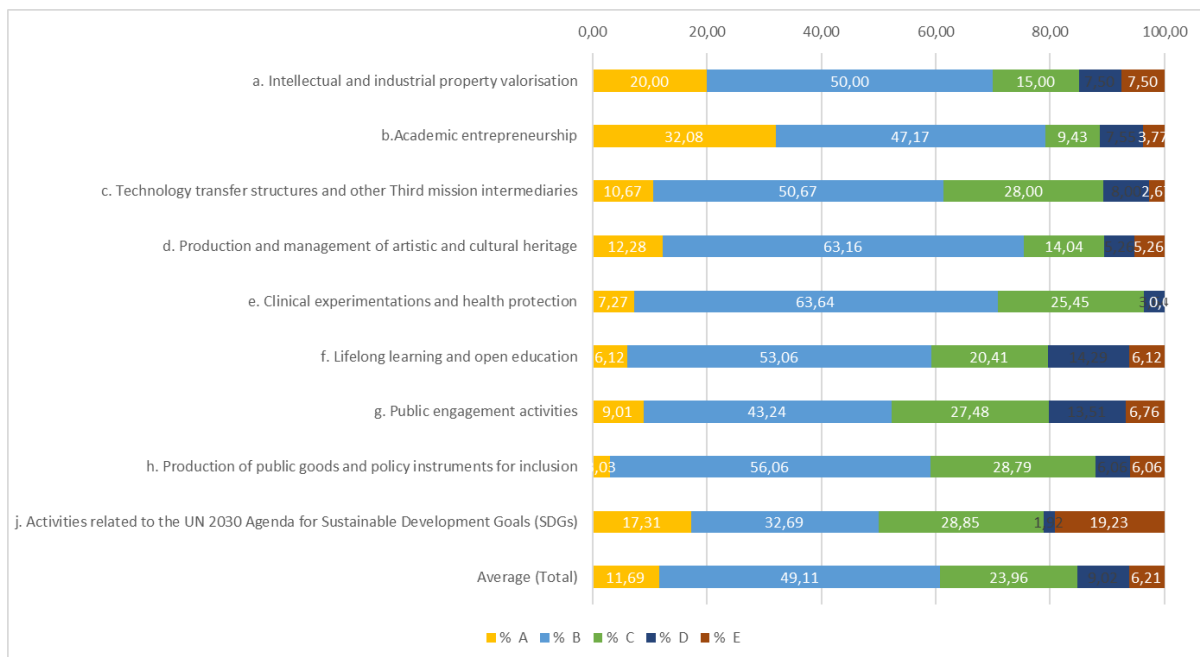
2.4.6 In-depth analysis of the outcomes with a focus on the type of institution evaluated and the geographical area

This section presents an in-depth analysis of the type of institution (university/research institute; institution/department), the geographical location (of the universities only) and the relationship of these factors with the evaluation results.

Overall, the following institutions participated in the VQR 2015-2019: 98 universities, 14 public research institutes (EPR) supervised by the Ministry of Education, University and Research, 22 institutions which participated on a voluntary basis. Universities are divided into State universities (61), non-State universities (20), Schools of advanced Studies with special regulations (6) and telematic universities (11). Out of the 676 case studies submitted for evaluation, 68,5% were submitted by universities, 23,5% by the research institutes and the remaining 8% by other institutions (see Figure 2.4.5).

---

[8] The field of action i) Open science is not reported because less than 10 case studies were presented.

Figure 2.4.5 - Percentage of case studies submitted by type of institution.
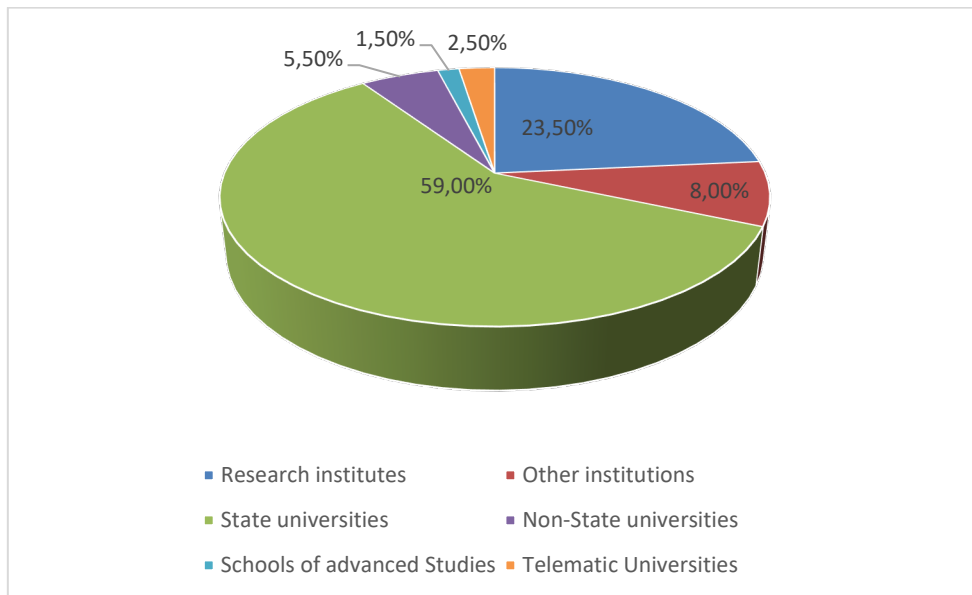
As Figure 2.4.6 shows, universities and research institutes (which include here both public research institutes and other institutions that participated in the VQR) make different choices on the fields of action of the case studies. Both types of institutions prioritise *g) Public engagement activities* as their main field of action*, but the research institutes in a significantly higher percentage (43%) than universities (28%). Moreover, the research institutes show significantly higher percentages than universities in the field *a) Intellectual and industrial property valorisation.* For research institutes, higher percentages are for fields *c) Technology transfer structures and other Third mission intermediaries* (with the same percentage as universities), *b) Academic entrepreneurship* and *f) Lifelong learning and open education* follow*. Probably due to a different institutional focus, the universities diversified their fields of action more, although confirming a distribution of fields *g)* and *i)* far from the average. After field *g)*, indeed, the most chosen fields are *h)*, *d)* and *c)*; not far behind, fields *e)*, *j)* and *b)* follow. Moreover, unlike the research institutes, a marked low weight is given to field *a)*.
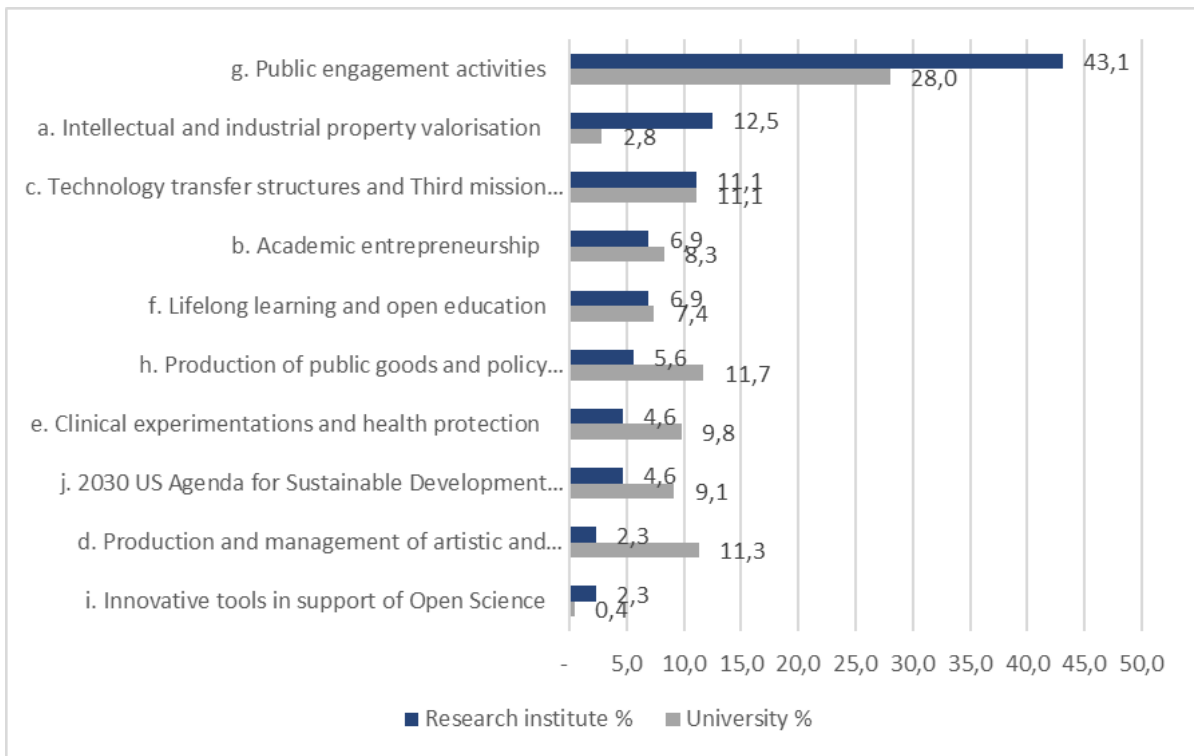
Concerning the evaluation results, both the average score obtained and the distribution of the assessment categories for each type are considered. The average score is determined by the ratio between the overall score obtained by the case studies and the number of case studies expected[9] for each type of institution. As Figure 2.4.7 shows, State universities and Schools of advanced Studies with special regulations obtain the highest average score. Research institutes show a slightly lower score than the overall average score. They are followed by non-State universities, institutions which participated in the VQR on a voluntary basis and, finally, telematic universities.

---

[9] The number of expected case studies corresponds to the number of submitted case studies, with only four exceptions concerning 3 universities and a research institute that each submitted 1 case study less.

Figure 2.4.7 - Average score by type of institution.

Figure 2.4.7 - Average score by type of institution.
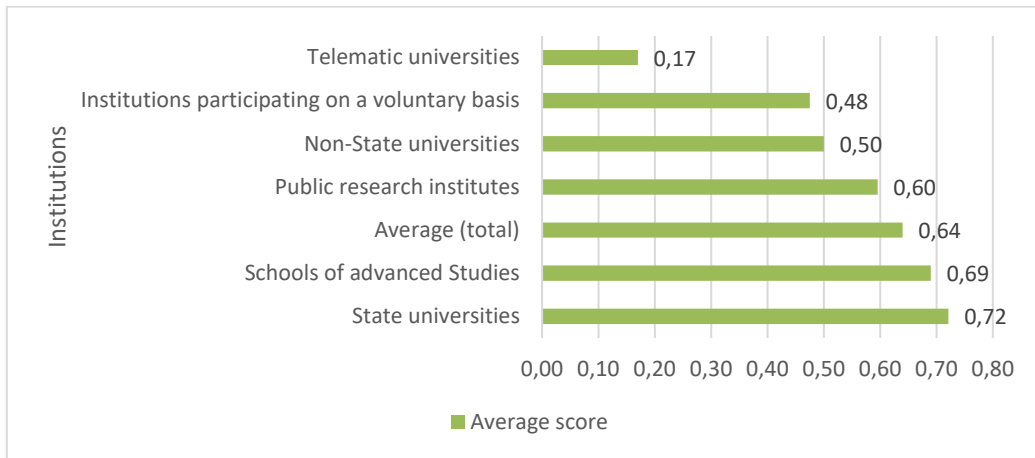
If we consider the assessment categories attributed to the Institutions (Figure 2.4.8), the evaluations falling in category A (Excellent and extremely relevant) are quantitatively significant for the Schools of advanced Studies and State universities. State universities also feature the highest percentage of case studies evaluated as B; thus, by taking categories A and B together, State universities receive excellent evaluations in 73% of the cases against an average of 61%. Schools of advanced Studies feature near-average evaluations in A+B categories, while all other institutions perform worst.

Figure 2.4.8 - Distribution of assessment categories by type of institution.
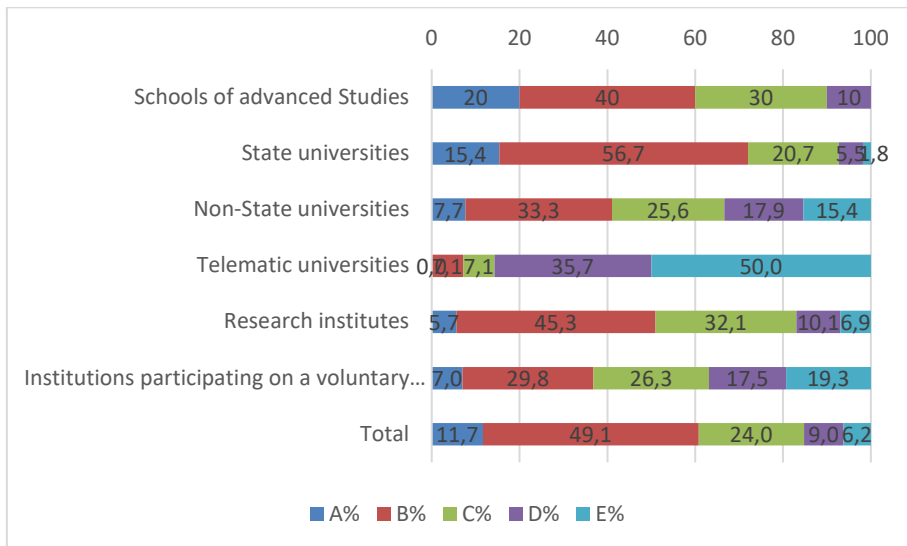
From these results we can conclude that the institutions have shown different Third Mission strategies and a different degree of maturity in the presentation of experiences and their impact. Within the universities, the picture is more diversified, both in terms of fields of action and in terms of results, with a significantly better performance of State universities than of the others. At the institutional level, the different attention paid to Third Mission issues could be further investigated by comparing these results with the institutions' programme documents, which set the Third Mission strategy and give an account of the peculiarities of each of them.

The VQR Call, art. 9, paragraph 2, states that the case studies are generally referred to departments or similar units or, should the action involve the entire institution, the case study may be referred to the latter. A case study can also be submitted by more than one department, while respecting the constraint of maximum two case studies per department. A total of 42,6% of the case studies were submitted by the entire institution, while the 48,7% were submitted by department or similar units. Only 8,7% of case studies were submitted by several departments (mainly 2).

Figure 2.4.9 - Percentage of case studies submitted by the entire institution and departments.
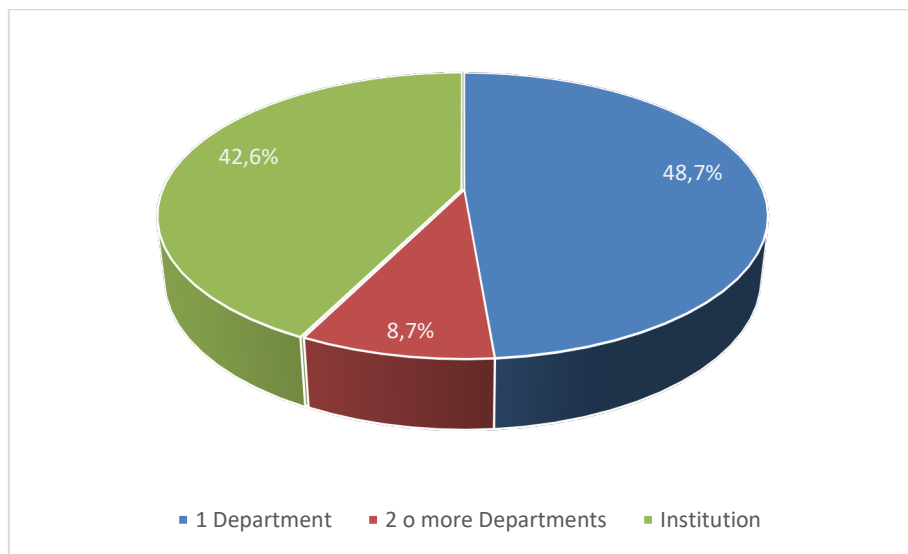
The fact that the case study refers to the institution, rather than to the department, does not seem to have affected the evaluation. Indeed, in the distribution of assessment categories the case studies of both types of institution show values close to the average in the five categories (Figure 2.4.10).

Figure 2.4.10 - Distribution of the assessment categories of the case studies submitted by the entire institution and departments.
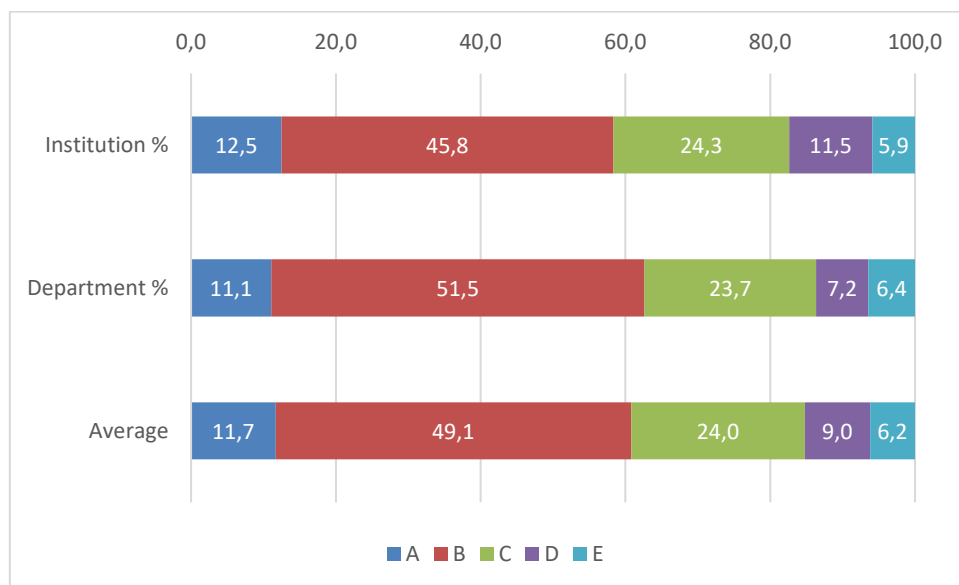
From a geographical and territorial point of view, some peculiarities emerge concerning universities only, as shown in Figure 2.4.11, which shows the fields of action by geographical area. Field *g) Public engagement activities* is the most widespread throughout the country, although to a lesser extent in the North-West and the South.

In the South the percentage of case studies relating to the production of artistic and cultural goods (field *d*) is more relevant, a result that had already emerged in the VQR 2011-2014[10]. Other important fields in the universities of the South are the Production of public goods and policy instruments for inclusion (field *h*), and academic entrepreneurship (field *b*).

The universities of Central Italy, besides featuring a high share of case studies related to Public engagement activities (field *g*, 31%), show a good percentage of case studies related to the Production of public goods and policy instruments for inclusion (field *h*, 12%) and, compared to the other universities, show the highest percentage of case studies concerning activities related to the 2030 UN Agenda for Sustainable Development Goals (field *j*, 11%).

In the universities of the North-West, in addition to public engagement activities (field *g*, 28%), a good number of case studies related to technology transfer (field *c*, 11%), clinical trials (field *e*, 11%) and the production of public goods and policy instruments for inclusion (field *h*, 11%). Moreover, a good percentage of case studies concern activities related to 2030 UN Agenda (field *i*, 10*%).*

In the North-East the highest percentage of case studies related to public engagement activities is to be found (field *g*, 34%), but also the highest percentage of case studies related to technology transfer (field *c*, 16%). The production of artistic and cultural goods is the third field of action for this geographical part (field *d*, 10*%*), while the intellectual property valorisation (field *a*), which presents in general a low number of cases, shows the highest percentage compared to the other parts (5%).

---

[10] Further details can be found in the 2018 ANVUR Biennial Report, Section 9

Figure 2.4.11 - Distribution of fields of action by geographical area of universities.

Legend:
- a. Intellectual and industrial property valorisation
- b. Academic entrepreneurship
- c. Technology transfer structures and other Third mission intermediaries
- d. Production and management of artistic and cultural heritage
- e. Clinical experimentations and health protection
- f. Lifelong learning and open education
- g. Public engagement activities
- h. Production of public goods and policy instruments for inclusion
- i. Innovative tools in support of Open Science
- j. Activities related to the 2030 UN Agenda for Sustainable Development Goals (SDGs)

Universities therefore show a widespread commitment over the territory, represented here by the different choice in the selection of fields of action. The South confirms a greater focus in the field of cultural heritage and the North a vocation for technology transfer. The Centre has also chosen more cross-cutting fields, such as the 2030 US Agenda and inclusion policies.

As we have seen, universities, except for non-State universities, were received the highest evaluations. As displayed in Figure 2.4.12, the universities' evaluations show a share of case studies in category B (Excellent), accounting for almost 52% of the case studies evaluated by GEV TM. If category A (Excellent and extremely relevant) is also added, so as to consider excellence as a whole, the resulting percentage values of these first two categories (A + B) is equal to 67%.

This percentage rises to almost 87% in the Northeastern universities. Universities of the North-West, on the other hand, are close to the average score, with evaluations in categories A + B equal to 67,5%. They are followed by the universities of the Centre and the South, whose evaluations in categories A + B are below the average (61,7% and 60,6%, respectively).

Figure 2.4.12 - Distribution of assessment categories by geographical location of universities.
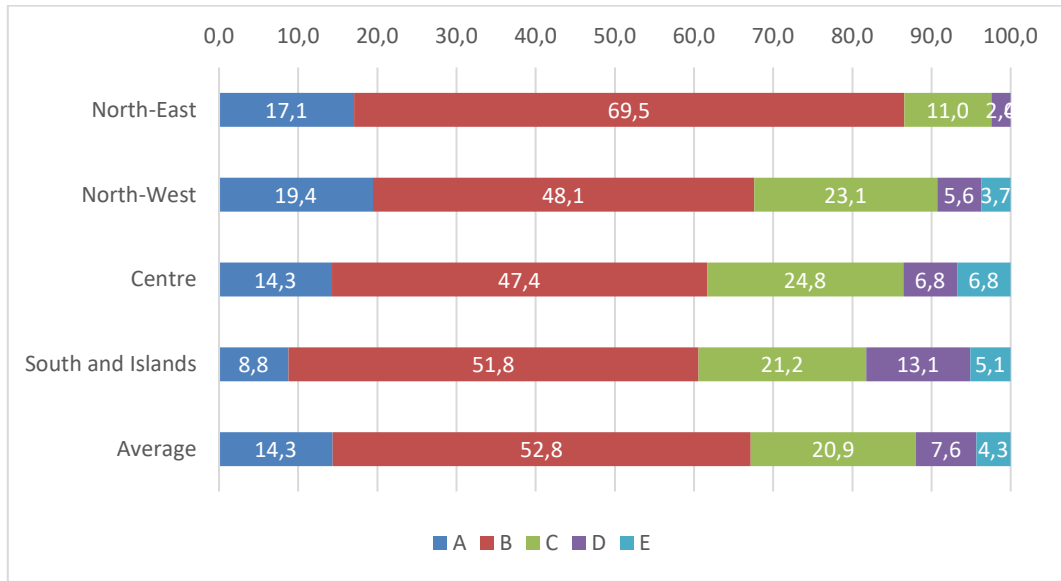
2.4.7   Degree of agreement between the two GEV members' assessments

Figure 2.4.13 shows the degree of agreement of the TM GEV's evaluators in the first phase of the evaluation, in which they made an independent assessment of the case study. In the second phase they worked together to find common ground for the final judgment.[11]

In 36% of the cases, the evaluators agreed on the assessment category to be assigned to the case studies, while in 45% of the cases the assessment of the two evaluators diverged only by one category.

---

[11] In the first phase, the evaluators assigned a score to each criterion; scores were then summed up. At this stage only a total score was assigned, not an assessment class for each evaluator. In order to analyse the degree of agreement, the scores were then converted into categories, following the transcoding Figure used by the GEV. The restrictions indicated by the GEV and taken over in the second phase of the evaluation have not been not taken into account.

Figure 2.4.13 - Degree of agreement of the evaluators
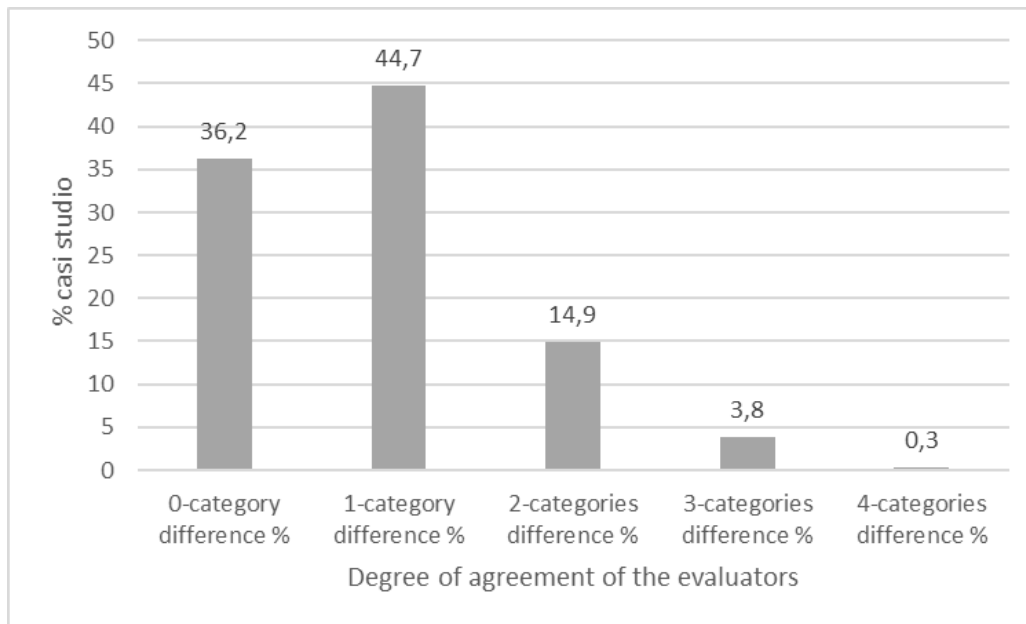
2.4.8 Number of case studies to be submitted in relation to the number of departments and the size of the institution

As far as universities are concerned, the number of departments can also vary among universities of similar size and, consequently, the number of expected case studies also differs. As shown in Table 2.4.3, within the size classes represented by the number of enrolled students,[12] there is a great variety in the number of expected case studies and a deviation from the average in some cases very high, especially in the size class 1, with a standard deviation of 5,7.[13]

---

[12] All universities are considered except for Schools of advanced Studies, because the data on enrolments is missing.
[13] The size classes were defined in a joint work by MUR, ANVUR and ISTAT in 2020. More details at http://ustat.miur.it/opendata/. On the same occasion, it was decided to use numerical classes without dimensional labels.

Table 2.4.3 - Size class by number of enrolled students, universities, and case studies; minimum and maximum, average and standard deviation.

Size class	No. of enrolled students	No. of universities	No. of case studies	Min. case studies	Max case studies	Average	Standard deviation
1	60.000 and over	6	106	13	30	17,7	5,7
2	from 30.000 to 59.999	13	102	1	12	7,8	2,9
3	from 20.000 to 29.999	16	99	1	9	6,2	2,0
4	from 10.000 to 19.999	20	82	1	8	4,1	1,5
5	less than 10.000	37	64	1	3	1,7	0,8

Universities of comparable size have submitted for evaluation very different amounts of case studies.

As the TM GEV pointed out, the number of case studies to be submitted shall be proportional to the size of the institution, just as it is the case for research products, which are not submitted according to the number of departments, but to the number of researchers. We notice that the final outcome of the evaluation, converted into the IRAS4 indicator, refers to the size of the institution, rather than to the number of case studies submitted. Indeed, the IRAS4 indicator was calculated in relation to the number of submitted research products that have been placed in an assessment category at least equal to "sufficient relevance" (Ministerial Decree no. 289 of 25/03/2021). With a view to case studies proportional to the size of the institution, the IRAS4 coefficient itself could be calculated on the basis of the submitted case studies only, compared to those expected. Instead, the choice of the departmental criterion has led some institutions of comparable size (but not with the same number of departments) to submit the best case studies, taking advantage of the multiplication coefficient itself.

It should be specified that these aspects do not concern ANVUR choices, but originate from provisions contained in ministerial decrees, in particular Ministerial Decree no. 1110/2019 and subsequent integration contained in Ministerial Decree no. 444/2020 for what concerns the number of case studies, and the Ministerial Decree no. 289/2021 (General Guidelines for Universities Programming 2021-2023 and indicators for the periodic evaluation of results) for what concerns the methods for calculating IRAS4.

To provide an initial guidance in view of a possible redefinition of the rules of the assessment exercise, in the following we put forward some hypotheses regarding the number of case studies to be submitted in the event that it is decided to relate this number to the number of personnel, similar to what is done in the Research Excellent Framework (REF 2021) and also in the VQR for research products.

In particular, three hypotheses were considered (see Table 2.4.4), progressively increasing the number of expected case studies according to the size class of personnel. In the most "cautious" case, the total number of case studies to be evaluated would be 539, thus lower than in the VQR 2015-19; the second case is in line with the current total number of submissions, while in the third case the number of case studies to be evaluated increases by about 60 compared to the last VQR. As mentioned before, this is an initial hypothesis, which may still serve as a guide for future choices.

Table 2.4.4 - Size class by number of personnel, number of universities by class, and number of expected case studies.

Size class	No. of universities	No. of expected case studies (hypothesis 1)	No. of expected case studies (hypothesis 2)	No. of expected case studies (hypothesis 3)
Up to 40	11	1	2	3
from 41 to 70	10	2	3	4
from 71 to 120	10	3	4	5
from 121 to 210	9	4	5	6
from 211 to 300	9	5	6	7
from 301 to 530	10	6	7	8
from 531 to 700	10	7	8	9
from 701 to 950	9	8	9	10
from 951 to 1.000	5	9	10	11
over 1.000*	15	10	11	12
Total	**98**	**539**	**673**	**735**

2.4.9 Concluding remarks

In the context of the VQR 2015-2019, the Third Mission was a major new element. The new evaluation methodology, based on the impact of case studies, has brought to light a growing attention on the part of the institutions towards TM, often considered marginal compared to teaching and research. In this sense, as it has been emphasised on several occasions, it seems appropriate to question the name Third Mission, which refers precisely to a "thirdness" with respect to the other missions and which does not take into account the reciprocity of the knowledge exchange processes in which Third Mission is carried out.

The evaluation based on case studies has made it possible to enhance the specificities of certain contexts and of the many activities that the institutions address to the territories, thus contributing to the social, economic, and cultural development of the country. The results confirm different positioning of the TM within the institutional strategies adopted, which depend on the specificity of the institution, its scientific-cultural background and local context.

The broad and complex definition of impact, applied to the various fields of action, allowed the institutions to highlight the experiences they considered most important, ranging from the valorisation of research and technology transfer in the narrow sense, to initiatives of social, economic, and institutional commitment in the broader sense. Furthermore, the institutions were free to choose the indicators they considered most appropriate to illustrate the impact of the case study.

The wide range of fields of action and the possibility of indicating other associated fields also made it possible to involve all disciplines, including the humanities and social sciences, and enhanced multidisciplinarity. However, such breadth of the fields of action has not always led to a correct attribution, on the part of the institutions, of the main field of action to the case study: a deep analysis is needed, first and foremost on public engagement, but also on the newer fields, in order to re-define or re-classify them better, or rather to re-frame TM by re-defining the fields of action.

In the future it might be useful to analyse and select indicators for each criterion and field of action reported in the case studies, so as to verify the weight of each criterion assigned by the institutions and, at the same time, provide them with guidance or good practices on indicators.

3 The participatory evaluation: the results of a surveys on the quality of the procedures adopted by ANVUR

3.1 The strategic framework and regulatory references

3.1.1 The legislative decree no. 150/2009

Articles 7 and 19bis of Legislative Decree no. 150/2009 require public administrations to carry out an annual assessment of organizational and individual performance, which takes place through the adoption (and periodic updating, subject to the binding opinion of the Independent Evaluation Board - OIV) of a Performance Measurement and Evaluation System (SMVP) that, with the introduction of art. 19bis, entails the participation of citizens and other end-users in the evaluation itself. More specifically, the participation of citizens (also in associated form) and end users can take place both through direct communication with the OIVs and through satisfaction survey systems concerning the activities and services provided by public administrations.

3.1.2 ANVUR Performance Programme 2022-24

Among the main changes in the Performance Programme 2022-24, compared to the previous 2021-23 edition, there is the qualification of some strategic objectives in terms of public value produced by the Agency, and their implementation with a view to participatory evaluation that involves stakeholders.
With the Resolution of the Governing Board no. 72 of 13/04/2022, which approved the 2022-24 Performance Programme, ANVUR has planned a series of actions aimed at surveying the degree of satisfaction of the end users of its services. As far as the VQR is concerned, these actions have resulted in the production and delivery of two satisfaction questionnaires for the activities and services carried out during the VQR 2015-2019. The questionnaires were addressed, respectively, to the contact persons of the institutions participating in the VQR and to the members of the disciplinary and TM GEVs, who were directly involved during the VQR in the assessment of research products and case studies.

3.2 The opinion survey of the institutions participating in the VQR 2015-2019

3.2.1 The questionnaire submitted to the representatives of the participating institutions

Within the framework outlined so far, the opinion survey submitted to the institutions participating in the VQR 2015-2019 is divided into four thematic sections, plus an open question aimed at collecting suggestions and observations from the representatives of the institutions and a concluding session aimed at profiling the responding institution. The four thematic sections concern respectively:
1. The quality of the assistance service on the contents of the VQR Call carried out by ANVUR staff (by phone and/or via institutional e-mail) during the phase of verification/indication of the departments and researchers under evaluation, as well as of submission of research products and Third Mission case studies.

2. The quality of the IT platform provided by ANVUR in collaboration with CINECA for the verification/ indication of departments.

3. The quality of the IT platform provided by ANVUR in collaboration with CINECA for the verification/ indication of researchers.

4. The quality of the IT platform provided by ANVUR in collaboration with CINECA for the submission of research products and Third Mission case studies.

For each of these sections the institutional representatives were asked to express a rating from 1 (semantically corresponding to "*not at all satisfied*") to 10 ("*completely satisfied*"); intermediate scores did not correspond to semantic labels. For each section the following items were included:

1. First section, on the quality of the assistance of ANVUR staff:
   a. Availability.
   b. Helpfulness and kindness.
   c. Promptness of responses.
   d. Clarity and completeness.
2. Second, third and fourth section, on the quality of the various platforms provided in collaboration with CINECA:
   a. User-friendliness: intuitiveness of use and clarity of captions and/or explanatory popups.
   b. Functionality: completeness and comprehensiveness of functions.

Finally, the questionnaire ended with an open box aimed at collecting comments or suggestions.

3.2.2 Responses from representatives of the participating institutions

A total of 112 institutional representatives responded to the questionnaire, amounting to 84% of the 134 institutions participating in the VQR 2015-2019. In particular, all the representatives of the Schools of advanced Studies, 71% of the representatives of research institutes, 84% of the representatives of the State universities, 71% of the representatives of the non-State universities and 82% of the representatives of the institutions that participated in the VQR on a voluntary basis responded.

3.2.2.1 Overview

The analysis of the average scores associated with the various items of the four thematic sections of the questionnaire shows overall a fairly high degree of satisfaction on the part of the representatives of the participating institutions, almost always above the average of 7,5, with peaks even higher than 8,5 for the items of the first section (Table 3.2.1): the degree of satisfaction with the assistance provided by ANVUR staff.[14]

---

[14] We are aware of the limits that this kind of survey implies by starting from response sets of Likert-type scales that do not include the do not answer/do not know mode.

Table 3.2.1 - Overview of the average scores associated with the responses of the first section: the quality of the assistance service on the contents of the VQR Call carried out by ANVUR staff (by phone and/or via institutional box) during the phase of verification/indication of the departments and researchers under evaluation, as well as of the submission of research products and Third Mission case studies; by type of institution and size class of the university.

Type of institution	# institutions	Availability	Helpfulness	Promptness	Clarity
Other institution participating in the VQR on a voluntary basis	18	8,6	9,2	8,1	8,3
Public research institute (supervised by the MUR)	10	8,4	9,3	8,5	7,4
School of advanced Studies with special regulations	6	8,8	9,0	8,8	8,8
Non-State university	22	7,6	8,1	7,5	7,6
State university	56	7,9	8,4	7,8	7,8
Size class of the university	**# institutions**	**Availability**	**Helpfulness**	**Promptness**	**Clarity**
Less than 10.000	37	8,5	8,8	8,4	8,4
from 10.000 to 19.999	18	7,7	7,8	7,4	7,4
from 20.000 to 29.999	11	6,7	7,5	6,6	6,4
from 30.000 to 59.999	14	8,0	8,5	8,1	8,4
60.000 and over	4	6,5	8,5	5,8	6,8
Total	**112**	**8,0**	**8,6**	**7,9**	**7,9**

In general, looking at the average scores, the representatives of the institutions that participated in the VQR on a voluntary basis and those of non-State universities are slightly more critical. Among the representatives of the universities, the worst results in terms of average scores in all thematic sections concern the representatives of the medium-small universities.

Table 3.2.2 - Overview of the average scores associated to the responses of the second section: quality of the IT platform provided by ANVUR in collaboration with CINECA for the verification/indication of the departments; by type of institution and size class of the university.

Type of institution	# institutions	User-friendliness	Functionality
Other institution participating in the VQR on a voluntary basis	18	7,3	7,2
Public research institute (supervised by the MUR)	10	8,0	8,2
School of advanced Studies with special regulations	6	8,7	8,7
Non-State university	22	7,2	7,2
State university	56	8,2	8,1
Size class of the university	**# institutions**	**User-friendliness**	**Functionality**
Less than 10.000	37	8,3	8,2
from 10.000 to 19.999	18	7,9	7,6
from 20.000 to 29.999	11	6,5	6,6
from 30.000 to 59.999	14	8,4	8,5
60.000 and over	4	7,8	8,0
Total	**112**	**7,9**	**7,8**

Overall, the average degree of appreciation of the IT platforms for the verification of institutions and researchers is close to an average score of 8, both for user-friendliness and functionality (Table 3.2.2 and Table 3.2.3). The results for these two platforms are entirely uniform among the various types of institution and, among the universities, for the different size classes.

Type of institution	# institutions	User-friendliness	Functionality
Other institution participating in the VQR on a voluntary basis	18	7,3	7,3
Public research institute (supervised by the MUR)	10	8,6	8,3
School of advanced Studies with special regulations	6	8,5	8,5
Non-State university	22	7,1	7,2
State university	56	8,2	8,1
Size class of the university	**# institutions**	**User-friendliness**	**Functionality**
Less than 10.000	37	8,1	8,1
from 10.000 to 19.999	18	8,2	8,0
from 20.000 to 29.999	11	6,5	6,6
from 30.000 to 59.999	14	8,1	8,2
60.000 and over	4	7,8	7,0
Total	**112**	**7,9**	**7,8**

The platform for the submission of research products and TM case studies is the least appreciated of all, although with average scores always close to 7,5 (Table 3.2.4). More critical of this platform are non-State universities and, among universities, the medium-sized ones, which are probably more affected by several problems in the management of research product archives, as well as in the internal organization for the submission procedures that, in this VQR, were the responsibility of the individual departments.

Table 3.2.4 - Overview of the average scores associated with the responses of the fourth section: the quality of the IT platform provided by ANVUR in collaboration with CINECA for the submission of research products and Third Mission case studies; by type of institution and size class of the university.

Type of institution	# institutions	User-friendliness	Functionality
Other institution participating in the VQR on a voluntary basis	18	7,0	7,0
Public research institute (supervised by the MUR)	10	8,4	8,2
School of advanced Studies with special regulations	6	7,5	7,5
Non-State university	22	6,6	6,7
State university	56	7,8	7,9
Size class of the university	**# institutions**	**User-friendliness**	**Functionality**
Less than 10.000	37	7,7	7,7
from 10.000 to 19.999	18	7,6	7,5
from 20.000 to 29.999	11	6,1	6,4
from 30.000 to 59.999	14	7,9	8,0
60.000 and over	4	7,3	7,5
Total	**112**	**7,5**	**7,5**

3.2.3 Comments and suggestions from the representatives of the participating institutions

At the end of the questionnaire, the representatives of the institutions were asked to provide comments on the platforms used during the phases of verification of departments and researchers and submission of products, and/or suggestions for their improvement in the future (Table 3.2.5). A total of 24 representatives submitted comments, representing 21% of respondents.[15]

Table 3.2.5 - Respondents who submitted comments by type of institution and size class (in case of universities).

type/size class of institution	Less than 10,000	from 10,000 to 19,999	from 20,000 to 29,999	from 30,000 to 59,999	Not to indicate	Total
Other institution	0,0% (0)	0,0% (0)	0,0% (0)	0,0% (0)	66,7% (6)	25,0% (6)
Public research institute	0,0% (0)	0,0% (0)	0,0% (0)	0,0% (0)	33,3% (3)	12,5% (3)
School of advanced Studies	33,3% (2)	0,0% (0)	0,0% (0)	0,0% (0)	0,0% (0)	8,3% (2)
Non-State university	33,3% (2)	0,0% (0)	0,0% (0)	25,0% (1)	0,0% (0)	12,5% (3)
State university	33,3% (2)	100,0% (4)	100,0% (1)	75,0% (3)	0,0% (0)	41,7% (10)
Total	100,0% (6)	100,0% (4)	100,0% (1)	100,0% (4)	100,0% (9)	100,0% (24)

---

[15] Only "full" comments were counted, i.e., those with an actual content. Three comments were ignored because they did not contain any valid or meaningful remarks (e.g.: "/"; "-", "no particular remark").

Among the comments, we mention the need of a clearer documentation to illustrate the procedure, a better coordination between CINECA services to support both IRIS and VQR platforms, a stable VQR site, not amended while the products are conferred.

A final remark, at a more general level, concerns the need to enhance the involvement of the institutions also in the planning phase of the evaluation exercise itself.

3.3 The opinion survey of the GEV members participating in the VQR 2015-2019

3.3.1 The questionnaire submitted to the GEV members

The survey of the opinions of the GEV members participating in the VQR 2015-2019 is also divided into four thematic sections, plus an open question aimed at collecting suggestions and comments and an initial section aimed at profiling to respondent with respect to the role played within the GEV. The four thematic sections concern respectively:

1. Quality of support service during the evaluation exercise carried out by the Research Evaluation Area and the GEV assistants.
2. Quality of the IT platform provided by ANVUR in collaboration with CINECA as regards the tools related to the evaluation of research products [of the case studies for the Third Mission GEV].
3. Quality of the IT platform provided by ANVUR in collaboration with CINECA as regards the tools related to the selection of external reviewers and the assignment of the research products [of the case studies for the Third Mission GEV].
4. Quality of the IT platform provided by ANVUR in collaboration with CINECA as regards the tools related to citation and bibliometric information in support of the evaluation of products. This last section is only addressed to the GEV respondents who claimed to have made use for the evaluation of the citation indicators made available by the platforms or published on the ANVUR website.

For each of these sections, GEV members could assign a score from 1 to 4; each score corresponds semantically to: "*1 - not at all satisfied*"; "*2 - dissatisfied*"; "*3 - satisfied*"; "*4 - completely satisfied*". For each section the following items were included:

1. First section, on the quality of the support service provided by ANVUR staff and GEV assistants:
   a. Availability.
   b. Helpfulness and kindness.
   c. Promptness of responses.
   d. Clarity and completeness.
2. Second, third and fourth section, on the quality of the various platforms designed in collaboration with CINECA:
   a. User-friendliness: intuitiveness of use and clarity of captions and/or explanatory popups.
   b. Functionality: completeness and comprehensiveness of functions.

Finally, the questionnaire ended with an open question aimed at collecting comments or suggestions.

3.3.2 The responses from GEV members

A total of 466 GEV members responded to the questionnaire; two of them, however, did not agree to participate in the survey. Therefore, the coverage rate was 72%, namely 464 respondents out of the 645 GEV members serving at the time of the conclusion of the evaluation activities, as shown in Table 3.3.1.

Table 3.3.1 - GEV members and GEV assistants serving during the VQR 2015-2019.

Area	Name of disciplinary area	GEV type	Assistants	# GEV members according to the VQR call	# Supplementary GEV members	# GEV members with contract	# GEV members serving at the end of the VQR
1	Mathematics and computer sciences	STEM+LS	1	29	4	34	33
2	Physics	STEM+LS	2	43	4	48	47
3	Chemistry	STEM+LS	1	31	4	37	34
4	Earth sciences	STEM+LS	1	19	3	23	22
5	Biology	STEM+LS	2	49	4	53	53
6	Medicine	STEM+LS	2	80	1	84	80
7	Agricultural and veterinary sciences	STEM+LS	1	36	4	42	40
8a	Architecture	SSH	1	17	2	17	17
8b	Civil engineering	STEM+LS	1	15	4	18	17
9	Industrial and information engineering	STEM+LS	2	59	3	64	63
10	Classics, Philology, Literary Studies, Art History	SSH	2	56	2	60	59
11th	History, Philosophy and Pedagogy	SSH	1	29	2	31	31
11b	Psychology	STEM+LS	1	12	4	14	14
12	Law studies	SSH	1	39	2	46	43
13a	Economics and Statistics	SSH	1	22	3	25	23
13b	Business Studies	SSH	1	18	2	21	21
14	Political and Social sciences	SSH	1	16	0	18	18
TM	Third Mission/Interdisciplinary Impact	SSH	2	30	0	33	30
	Total		24	600	48	668	645

Respondents belonged in 31% of cases to a SSH GEV and in the remaining 69% of cases to a GEV from the STEM+LS area. Compared to the number GEV members serving at the end of the evaluation exercise, therefore, the GEV members belonging to STEM+LS GEVs are slightly overrepresented (69% vs. 62%).

GEV and Sub-GEV coordinators are also over-represented among the respondents: 55 out of 464, equal to about 12%, against 61 out of 645, equal to 9,5%.

3.3.2.1 Overview

Unlike the survey on the opinions of the institutional representatives, in the survey on the opinions of the GEV members the scale of scores ranges from 1 to 4: the average scores associated with the various items of the four thematic sections of the questionnaire show a rather high degree of satisfaction on the part of the GEV members, which is almost always higher, on average, than 2,5, with peaks even above 3,5 for the items in the first section, concerning the degree of satisfaction with the quality of the support service provided by ANVUR staff and GEV assistants. Among the four items of the first section, in particular, the one with the lowest average score is clarity (3,6); the one with the higher average score is helpfulness and kindness (3.8). There are no substantial differences between the average scores associated with these items by the GEV members belonging to STEM+LS and SSH area, nor with respect to the role played by the respondent within the GEV[16], see Table 3.3.2.

Table 3.3.2 - Overview of the average scores associated with the responses of the first section: the quality of the support service during the evaluation exercise carried out by the Research Evaluation Area and the GEV assistants; by type of GEV and role of the respondent within the GEV.

GEV type	# respondents	Availability	Helpfulness	Promptness	Clarity
STEM+LS	321	3,8	3,9	3,7	3,6
SSH	143	3,7	3,7	3,6	3,5
Role in the GEV	**# respondents**	**Availability**	**Helpfulness**	**Promptness**	**Clarity**
GEV member	409	3,7	3,8	3,7	3,6
GEV or sub-GEV Coordinator	55	3,8	3,9	3,8	3,6
Total	**464**	**3,7**	**3,8**	**3,7**	**3,6**

Regarding the platforms used for the evaluation of research products and case studies and the platform for the selection of external reviewers and assignment of products for evaluation, the average scores associated with the two items (user-friendliness and functionality) are almost one point lower than the previous section.

Unlike what was found in the survey addressed to the institutional representatives, who primarily appreciated the functionality of the platforms provided in collaboration with CINECA, in the case of the platforms used by GEV members, what is on average more appreciated is the ease of use at the expense of functionality, see Table 3.3.3.

---

[16] What was said in the previous section about the use of mean scores on Likert-type scales also applies here.

Table 3.3.3 - Overview of the average scores associated to the responses of the second section: the quality of the IT platform provided by ANVUR in collaboration with CINECA regarding the tools related to the evaluation of research products [of the case studies for the Third Mission GEV]; by type of GEV and role of the respondent within the GEV.

GEV type	# respondents	User-friendliness	Functionality
STEM+LS	321	2,8	2,6
SSH	143	2,6	2,3
Role in the GEV	**# respondents**	**User-friendliness**	**Functionality**
GEV member	409	2,7	2,5
GEV or sub-GEV Coordinator	55	2,7	2,6
Total	**464**	**2,7**	**2,5**

The GEV members belonging to non-bibliometric GEVs appear more critical, especially with regard to the platform for the selection of external reviewers and the assignment of products, which receives the lowest average score: 2.3 with respect to functionality, see Table 3.3.4.

Table 3.3.4 - Overview of the average scores associated to the responses of the third section s: the quality of the IT platform provided by ANVUR in collaboration with CINECA regarding to the tools related to the selection of external reviewers and the assignment of research products [of the case studies for the Third Mission GEV]; by type of GEV and role of the respondent within the GEV.

GEV type	# respondents	User-friendliness	Functionality
STEM+LS	321	2,7	2,5
SSH	143	2,5	2,3
Role in the GEV	**# respondents**	**User-friendliness**	**Functionality**
GEV member	409	2,6	2,4
GEV or sub-GEV Coordinator	55	2,6	2,5
Total	**464**	**2,6**	**2,4**

The last thematic section, as mentioned in the sub-section on the structure of the questionnaire, is addressed only to GEV members of the STEM+LS area who, in their evaluation activity, have resorted to the citation and bibliometric information made available through a specific tool on the evaluation platform and on the Agency's institutional website.

In this case, too, the average scores are above 3: 3,2 for user-friendliness; 3.1 for the functionality, see Table 3.3.5. Once again, there are no significant differences according to the role of the respondents within the GEV.

GEV type	# respondents	User-friendliness	Functionality
STEM+LS	321	3,2	3,1
Role in the GEV	**# respondents**	**User-friendliness**	**Functionality**
GEV member	283	3,2	3,1
GEV or sub-GEV Coordinator	38	3,2	3,2
Total	**321**	**3,2**	**3,1**

3.3.3 Comments and suggestions from the GEV members

At the end of the questionnaire, respondents were asked to provide comments on the evaluation exercise that had just been completed or suggestions for its improvement in the future. A total of 246 respondents gave comments, amounting to 53% of respondents, see Table 3.3.6.

Table 3.3.6 - Respondents who provided comments by type of GEV and role within the GEV.

GEV type /role	GEV member	GEV or sub-GEV Coordinator	Total
STEM+LS	60,1% (128)	51,5% (17)	58,9% (145)
SSH	39,9% (85)	48,5% (16)	41,1% (101)
Total	100,0% (213)	100,0% (33)	100,0% (246)

Respondents mainly commented on the following issues:

1. support received throughout the evaluation process, both from ANVUR and more frequently from the GEV assistants, almost always highly appreciated;
2. number of GEV members, considered insufficient both to evaluate the high number of research products submitted and to cover the variety of scientific topics;
3. possibility of a more extensive use of external reviewers, and the need to involve them in a training process, in view of a better selection by keywords and not only by disciplinary fields;
4. interaction with the IT platform which, although facilitating much of the GEVs' work, has often been perceived as cumbersome and unresponsive.

Overall, the comments – the vast majority of which are negative – are almost always provided in the form of suggestions aimed at improving the overall structure of the future process. There are very few cases of

unqualified negative opinions. More frequently, comprehensive comments have been offered that almost entirely follow the points of attention already presented in this document regarding the entire process.

4   The international debate on the reform of research assessment: where are we?

In January 2022, the European Commission launched the process of drafting an agreement for the reform of research evaluation (https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/process-towards-agreement-reforming-research-assessment-2022-01-18_en) which gradually reached a global dimension with the participation of non-European institutions. On 8 July 2022, the final version of the agreement was presented to a stakeholder assembly, which brought together more than 350 organisations from over 40 countries (https://research-and-innovation.ec.europa.eu/document/download/bd53d467-0220-4c9d-8b63-26eb56303ba1_en), which had expressed an interest in drafting the document. The organisations involved include public and private research funders, universities, research centres, institutes and infrastructures, associations and their alliances, national and regional authorities, accreditation and evaluation agencies, academic societies and associations of researchers and other relevant organisations, representing a wide variety of views and perspectives. The institutions involved provided feedback to the various drafts of the agreement, which were prepared by a team of representatives from the Association of European Universities (EUA), Science Europe and the European Commission. Therefore, a global coalition of research funding organisations, research organisations, national/regional evaluation authorities and agencies – as well as associations of these organisations, scientific societies and other relevant organisations – has been established, with the aim of working together towards a systemic reform on the basis of common principles and an agreed schedule, and of facilitating the information exchange and mutual learning among all those intending to improve research evaluation practices (https://coara.eu/).

The objective of the ongoing process is to share a direction for changes in research evaluation practices, with the overall objective of maximising the quality and impact of research. The agreement includes some commitments underlying the reform process, discusses the timeline of the reform and establishes the principles for a coalition of organisations willing to work together to implement the changes. On 6 October 2022, the final document was signed by ANVUR; the Agency's participation in the International Coalition is a further element that demonstrates the Agency's strong will to actively participate in international initiatives on a topic that is central to its activities. The decision to sign the document and join the international coalition, taken by the Governing Board at its meeting on 3 October 2022, represents a starting point for the activities that the Agency will have to develop for the definition of common rules at international level, starting from this policy document. The signing of such document comes at the end of a process of active participation and discussion carried out through the participation of Prof. Menico Rizzi, a member of the Governing Board with responsibility for research on evaluation and the Agency's representative in the International Coalition, which officially became operational on 1/12/2022, when a steering board (including Prof. Menico Rizzi from the ANVUR Governing Board), its president and two vice-presidents were elected.