
Final Area Report
Group of Evaluation Experts for Area 13 (GEV13)
Table of contents

GLOSSARY ........................................................................................................................................... 4
LIST OF TABLES ..................................................................................................................................... 6

1 INTRODUCTION ................................................................................................................................. 13
  1.1 THE GROUP OF EVALUATION EXPERTS (GEV) ............................................................................. 14
  1.2 THE TIMEFRAME ........................................................................................................................... 15
  1.3 DESCRIPTION OF THE AREA ......................................................................................................... 16

2 EVALUATION OF “RESEARCH PRODUCTS” .................................................................................... 18
  2.1 EVALUATION CRITERIA ................................................................................................................ 18
  2.2 JOURNAL CLASSIFICATION .......................................................................................................... 19
  2.3 THE EVALUATION PROCESS ......................................................................................................... 20
      2.3.1 Peer evaluation ...................................................................................................................... 20
      2.3.2 Bibliometric evaluation ........................................................................................................ 21
      2.3.3 Statistics ............................................................................................................................... 23
  2.4 EVALUATION RESULTS .................................................................................................................. 27

3 EVALUATION OF INSTITUTIONS ....................................................................................................... 30
  3.1 INDICATORS .................................................................................................................................... 31
      3.1.1 The first indicator .................................................................................................................. 31
      3.1.2 The second indicator ............................................................................................................ 31
      3.1.3 The third indicator ................................................................................................................ 31
      3.1.4 The IRAS1ij indicator of the VQR Call .................................................................................. 32
  3.2 RANKINGS OF INSTITUTIONS ....................................................................................................... 33
      3.2.1 Universities ............................................................................................................................ 33
      3.2.2 Research Institutions ............................................................................................................ 35
  3.3 COMMENTS ON THE MEANING OF THE INDICATORS FOR THE INSTITUTIONS IN THE AREA ...... 37

4 EVALUATION OF DEPARTMENTS ...................................................................................................... 38
  4.1 INDICATORS .................................................................................................................................... 39
      4.1.1 The first indicator .................................................................................................................. 39
      4.1.2 The second indicator ............................................................................................................ 39
      4.1.3 The third indicator ................................................................................................................ 40
      4.1.4 The IRD1ij indicator of the VQR Call .................................................................................. 40
  4.2 RANKINGS OF DEPARTMENTS ..................................................................................................... 41

5 EVALUATION OF RESEARCH STAFF MEMBERS ............................................................................ 43
6 SUMMARY OF RESULTS........................................................................................................................................45
7 FINAL CONSIDERATIONS ....................................................................................................................................48

APPENDIX A EVALUATION CRITERIA FOR THE AREA

APPENDIX B IMPUTATION OF MISSING BIBLIOMETRIC INDICATORS AND JOURNAL CLASSIFICATION

APPENDIX C ASSESSMENT FORM AND GUIDELINES FOR PEER REVIEWERS

APPENDIX D ANALYSIS OF INDIVIDUAL CITATIONS
GLOSSARY

ANVUR. National Agency for the Evaluation of Universities and Research Institutes.

AREAS. The sixteen scientific Areas into which the evaluation process has been divided.

CALL. The VQR 2011-2014 Call for participation.

CINECA. Inter-University Consortium for Computing. It managed the software interfaces and the administrative and accounting procedures related to the evaluation process.


EXPECTED PRODUCTS. The number of products that each Institution was expected to submit for evaluation, obtained by multiplying each research staff member by the number of products specified by the Call and adding up the results of the multiplication.

GEV. Groups of Evaluation Experts. The sixteen panels of experts in the disciplines of the scientific Areas that handled the evaluation of the research products submitted by the Institutions.


INSTITUTIONS. The Institutions subject to the VQR evaluation. They include: public and private Universities (with the obligation to undergo the evaluation), Research Institutions controlled by MIUR (with the obligation to undergo the evaluation), “assimilated” Research Institutions that asked to undergo the evaluation under the same rules of the Research Institutions controlled by MIUR, inter-University Consortia that asked to undergo the evaluation using a subset of the indicators applied to Universities and Research Institutions controlled by MIUR and, lastly, other Institutions that asked to undergo the evaluation under different rules agreed upon with ANVUR.

IRAS1-IRAS5. The indicators of research quality by Area and Institution defined by the Call, computed as a fraction of the aggregate value of an Area.

IRFS1. The final indicator of research quality of an Institution, that integrates the Area indicators IRAS1, …, IRAS5 with the weights attributed to the sixteen Areas.
**IRD1-IRD3.** The indicators of research quality by Area and Department defined by the Call, computed as a fraction of the aggregate value of an Area.

**IRD.** The final indicator of research quality of a Department, that integrates the indicators IRD1-IRD3 with the weights attributed to the sixteen Areas.

**LEGGE 240.** Law n. 240 of 30 December, 2010 “Rules on the organization of Universities, academic staff and recruitment, and delegation to the Government for the incentivization of quality and efficiency in the University system”.

**MACRO SC.** The *macrosettori concorsuali* (Academic Recruitment Macrosectors), consisting of groups of SSDs.

**MIUR.** Ministry of University and Research.

**RESEARCH PRODUCTS or PRODUCTS.** Contributions defined in Section 2.3 of the Call (articles, monographs, book chapters, etc.) obtained as a result of research activity and submitted to ANVUR for evaluation.

**RESEARCH STAFF MEMBERS.** The staff of the Institutions who authored or co-authored the research products subject to evaluation. The corresponding acronym is RSM.

**SSD.** The 370 *settori scientifico-disciplinari* (Scientific-Disciplinary Sectors) into which the sixteen Areas are organized.

**SUBGEV.** Homogeneous subsets of the GEVs defined on the basis of the characteristics of the scientific Area.

**VQR.** Evaluation of Research Quality.

**VQR1.** Evaluation of Research Quality 2004-2010.

LIST OF TABLES

Table 1.1: *Settori scientifico-disciplinari* (SSD) of the Area.

Table 1.2: Composition of the Group of Evaluation Experts.

Table 1.3: Organization of the experts by subGEV, corresponding SSDs, and distribution of the research products evaluated by the GEV.

Table 1.4: List of GEV Meetings.

Table 1.5: Number of research staff members (RSM) of the Area by SSD.

Table 2.1: Number of reviewers by subGEV and nationality (Italian and not Italian); reviewers repeated for each SSD of expertise.

Table 2.2: Number of reviewers and reviews by SSD and nationality (Italian and not Italian); reviewers repeated for each SSD of expertise.

Table 2.3: Number of assigned, completed, not completed, and refused reviews by reviewer nationality (Italian and not Italian).

Table 2.4: Expected and submitted products for Area 13 and number of distinct products.

Table 2.5: Products submitted to the Area by type and year of publication.

Table 2.6: Distribution of submitted research products by publication type and SSD of research staff members.

Table 2.7: Distribution of submitted research products by language of publication and SSD of research staff members.

Table 2.8: Distribution of submitted research products by publication type, year of publication, and SSD of research staff members.

Table 2.9: Number of submitted and expected products by SSD of research staff members.

Table 2.10: Number of research staff members by corresponding number of expected products and by SSD of research staff members.
Table 2.11:  Number and percentage of research products submitted by research staff members of Area 13 by GEV that evaluated them.

Table 2.12:  Number and percentage of research products evaluated by GEV13 by Area of research staff members.

Table 2.13:  Number of research products submitted to GEV13 and classified as not assessable by SSD of research staff members.

Table 2.14:  Percentage of submitted products evaluated in peer review, bibliometric analysis, and informed peer review, by SSD of research staff members.

Table 2.15:  Scores and distribution of submitted products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E) and by evaluation method.

Table 2.16:  Scores and distribution of submitted products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F) and by publication type.

Table 2.17:  Scores and distribution of products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F) in the Area, by SSD and subGEV of research staff members.

Table 2.18:  Scores and distribution of submitted products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F), by SSD of research staff members, and by year of publication.

Table 2.19:  Scores and distribution of submitted products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F) and by type and language of publication.

Table 2.20:  Scores and distribution of submitted products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F), by SSD of research staff members, and by type of publication.

Table 3.1:  List of Universities in alphabetical order.

Table 3.2:  Ranking of small Universities (in terms of the number of expected products of the research staff members of the Institution) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).
Table 3.3: Ranking of medium-size Universities (in terms of the number of expected products of the research staff members of the Institution) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.4: Ranking of large Universities (in terms of the number of expected products of the research staff members of the Institution) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.5: List of Universities in alphabetical order for all subGEVs of the Area.

Table 3.6: Ranking of small Universities by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.7: Ranking of medium-size Universities by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.8: Ranking of large Universities by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.9: List of Universities in alphabetical order for all SSDs in the Area.

Table 3.10: Ranking of small Universities by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.11: Ranking of medium-size Universities by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.12: Ranking of large Universities by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.13: List of Universities in alphabetical order for all MACRO SCs in the Area.
Table 3.14: Ranking of small Universities by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.15: Ranking of medium-size Universities by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.16: Ranking of large Universities by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.17: List of controlled and assimilated Research Institutions in alphabetical order.

Table 3.18: Ranking of controlled and assimilated Research Institutions on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.19: Ranking of controlled and assimilated Research Institutions by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.20: Ranking of controlled and assimilated Research Institutions by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.21: Ranking of controlled and assimilated Research Institutions by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.22: List of other Research Institutions that submitted products to the VQR on a voluntary basis in alphabetical order.

Table 3.23: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis in term of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).
Table 3.24: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.25: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.26: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.27: Number of active, not active, and partially active research staff members by Institution.

Table 4.1: List of University Departments, in alphabetical order, first by University, then by Department.

Table 4.2: Ranking of small Departments (in terms of the number of expected products of the research staff members of the Department) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.3: Ranking of medium-size Departments (in terms of the number of expected products of the research staff members of the Department) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.4: Ranking of large Departments (in terms of the number of expected products of the research staff members of the Department) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.5: List of University Departments in alphabetical order, first by University, then by Department, for all subGEVs of the Area.
Table 4.6: Ranking of small Departments by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.7: Ranking of medium-size Departments by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.8: Ranking of large Departments by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.9: List of University Departments in alphabetical order, first by University, then by Department, for all SSDs of the Area.

Table 4.10: Ranking of small Departments by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.11: Ranking of medium-size Departments by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.12: Ranking of large Departments by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.13: List of University Departments in alphabetical order, first by University, then by Department, for all MACRO SCs of the Area.

Table 4.14: Ranking of small Departments by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.15: Ranking of medium-size Departments by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).
Table 4.16: Ranking of large Departments by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.17: List of the substructures of the controlled and assimilated Research Institutions in alphabetical order.

Table 4.18: Ranking of the substructures of the controlled and assimilated Research Institutions on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.19: Ranking of the substructures of the controlled and assimilated Research Institutions by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.20: Ranking of the substructures of the controlled and assimilated Research Institutions by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.21: Ranking of the substructures of the controlled and assimilated Research Institutions by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 5.1: Ranking of Institutions by average score (VM) of research staff members and distribution by class of the average scores.

Table 5.2: Ranking of substructures of Institutions by average score (VM) of research staff members and distribution by class of the average scores.

Table 5.3: Average score (VM) of research staff members and distribution of the average score by University position (researcher -RU, associate professor -PA, full professor -PO) and by subGEV of research staff members.

Table 5.4: Average score (VM) of research staff members and distribution of the average score by year of birth and by subGEV of research staff members.
1 Introduction

From a variety of interrelated perspectives, such as research topics, analytic methodologies, and publication outlets, Area 13 (Economics and Statistics) lies in between the “hard” sciences, on the one hand, and the humanities and most of the other social sciences, on the other. In terms of publication outlets, the former are characterized by a prevalence of publications in journals indexed by international databases, while the latter more frequently rely on monographs and contributions to volume. Likewise, English is the language of choice for the former, while this is not necessarily the case for the latter. As a consequence of its intermediate position, research output in Area 13 is characterized by a marked heterogeneity. Together with a considerable share of research products published in international, indexed journals, we also observe a sizeable, albeit declining role for monographs, contributions to volume, and articles in non-indexed journals, sometimes published in Italian.

While these characteristics of Area 13 are persistent, and were already remarked in the Area Final Report for VQR 2004-2010, during the period covered by VQR 2011-2014 there has been evidence of an evolution, which has influenced the evaluation criteria adopted by the Group of Evaluation Experts currently in charge for Area 13 (GEV13). Within the present evaluation exercise, consistently with Decreto Ministeriale 458/2015 (DM) and the VQR Call, GEV13 has confirmed and extended the evaluation criteria introduced within the previous one. In particular, it has applied the peer-review evaluation methodology to publications in the form of monographs, contributions to volume, and articles in journals not included in the GEV journal list, while a simple bibliometric algorithm has been employed for articles in journals included in the list. To account for the growing presence of articles covered by Scopus beside ISI Web of Science (hereafter, WoS), GEV13 has considerably expanded the journal list compiled by the previous GEV by adding Scopus-indexed journals, formerly disregarded because of the limited coverage by Scopus at the time.

In line with VQR 2004-2010, and to overcome the persisting problem of the incomplete coverage of relevant journals in either database, GEV13 has applied an imputation methodology to include in the GEV journal list also journals not covered by bibliometric databases. GEV13 has also refined the criteria for the use of individual citations for bibliometric evaluation. Finally, a random sample of “bibliometric” products, that is, articles published in journals that belong to the GEV journal list and evaluated by bibliometric analysis, was also sent to peer review in order to implement a statistical comparison between the two evaluation methodologies.
The rest of Section 1 briefly describes GEV13 and the Area. Section 2 summarizes the evaluation criteria and process. Sections 3, 4 and 5 present evaluation results for Institutions, Departments, and research staff members within the Area. Section 6 contains a summary of the evaluation results. Section 7 concludes with a critical discussion of the evaluation experience of GEV13, with the aim to provide suggestions that could be useful for future VQR exercises. All tables are available in attachment.

Appendix A collects extensive documentation on the evaluation criteria. Appendix B describes the construction of the GEV journal list, the imputation methodology, and the ranking of journals. Appendix C contains the assessment form and the guidelines for reviewers. Appendix D analyzes the individual citations of the products submitted to GEV13 that were published in indexed journals belonging to the GEV journal list.

GEV13 thanks the members of the VQR group at ANVUR for their assistance, and especially Giulia Rossi, the GEV’s Assistant, for her competence, passion, and efficiency. The GEV is also grateful to the external reviewers for their crucial contribution to the evaluation process. A special thank goes to Sergio Benedetto, the VQR Coordinator, for his constant advice and relentless optimism.

1.1 The Group of Evaluation Experts (GEV)

GEV13 covers 19 settori scientifico-disciplinari (SSD) (Table 1.1) and is composed of 31 evaluation experts (Table 1.2), organized in 3 subGEVs (Table 1.3): Economics and Economic History (hereafter, Economics), Business Administration and Management (hereafter, Business), and Statistics and Mathematical Methods for Decisions (hereafter, Statistics). The GEV Coordinator is Graziella Bertocchi, the subGEV Coordinators are Agar Brugiavini (Economics), Salvatore Torrisi (Business), and Francesco Bartolucci (Statistics). The Assistant to the GEV is Giulia Rossi.

Within the GEV 7 members are affiliated with foreign universities, 2 of which are located in the United States, while the rest are located across 4 different European countries. The remaining 24 GEV members come from 21 Italian universities, of which 8 are from the South, 6 from the Center, and 7 from the North. Table 1.3 also shows that the number of products handled by each GEV member varies across subGEVs, with Business showing a larger than average workload.
Table 1.1: Settori scientifico-disciplinari (SSD) of the Area.

Table 1.2: Composition of the Group of Evaluation Experts.

Table 1.3: Organization of the experts by subGEV, corresponding SSDs, and distribution of the research products evaluated by the GEV.

GEV13 held two plenary meetings (Table 1.4). During the second one each subGEV also held a separate meeting. Exchanges constantly took place by electronic means throughout the evaluation process. Frequent Skype meetings involved the GEV Coordinator and the subGEV Coordinators. GEV members were informed by the GEV Coordinator, on a regular basis, about the progress of the evaluation. The Assistant attended all meetings and followed closely the entire process.

Table 1.4: List of GEV Meetings.

In terms of size, Area 13 comprises 4,700 research staff members (hereafter, staff members, or RSM). The largest SSDs in terms of staff members are SECS-P/01 and SECS-P/07, with over 700 members. At the other extreme, the Area also includes much smaller aggregates such as SECS-S/02 and SECS-P/04, with 20 to 30 members (Table 1.5). The number of the members of each subGEV is roughly proportional to the number of the staff members belonging to each SSD.

Table 1.5: Number of research staff members (RSM) of the Area by SSD.

1.2 The timeframe

GEV13 was appointed in September 2015 and was active until February 2017, for a total of 18 months. The first meeting was held in October 2015. In the subsequent weeks the GEV approved the evaluation criteria (Appendix A) with 30 members in favor and one against. The document was published on the ANVUR website on November 20, 2015, together with a provisional journal list. With a public announcement published on the same day, the GEV solicited from the academic community suggestions regarding the journal list, to be provided by December 4, 2015. The final version of the journal list, taking into account over 200 suggestions concerning over 1,000 journals, was published on December 14, 2015 together with an announcement detailing the revisions applied to the list. On January 14, 2016 the GEV published the journal classification reflecting the five VQR merit classes, together with an announcement explaining
the imputation and classification methodology. The journal classification used along the evaluation process has been available on the ANVUR website since January 22, 2016, following marginal revisions announced on January 22, 2016. Further clarifications and corrections were published on January 29, 2016, February 16, 2016, and January 13, 2017.

In the meantime, in November 2015 the GEV started to select reviewers to be employed for peer evaluation. The GEV also prepared guidelines for reviewers and the evaluation form.

Because of a delay in the availability of the interfaces needed to manage assignments and evaluations, the assignment of products to GEV members and their distribution to external reviewers started only in late May 2016, and continued throughout the Summer. During the Summer, GEV members also validated the bibliometric evaluations of products published in journals belonging to the GEV journal list. GEV13 completed and unanimously approved all evaluations in October 2016. The peer-review evaluations of a random sample of bibliometric products were completed in January 2017. The Area Final Report was prepared and then unanimously approved between January and February 2017.

1.3 Description of the Area

Area 13 is composed of about 4,700 research staff members distributed across 90 Universities and a small number (8) of Research Institutions. Among the latter, only Consiglio Nazionale delle Ricerche (CNR) counts a sizeable number of researchers. Among the former, only 82 count a significant number of research staff members (corresponding to at least 5 expected products to be submitted to the VQR). In terms of their presence in the Area, there are 51 small, 33 medium-size, and 6 large Universities, associated respectively with up to 100, 101 to 250, and more than 250 expected products. Because of the role of the disciplines taught by Area 13 members in University curricula, the Area is well represented, in a steady fashion, in the vast majority of the Institutions in the country. Within Universities, there is a significant presence of staff members of the Area in 186 Departments, while 14 additional Departments are associated with fewer than 5 expected products in the Area. On average, therefore, the Area is represented in two Departments within each University, but deviations from the average are substantial, since Department size varies considerably. The degree of fragmentation across Departments within the same University is, however, declining, especially after the reform introduced with Legge 240. CNR includes 8 Institutes with a significant presence of researchers from the Area.

The Area includes 4 macrosettori concorsuali (hereafter, MACRO SC): Business Administration and Management (representing about 40% of the research staff members of the
Area and 6 SSDs), Economics (30% and 5, respectively), Economic History (5% and 2, respectively), and Statistics and Mathematical Methods for Decisions (25% and 6, respectively).

In terms of research themes, methodologies, and publication outlets, the Area displays a high degree of heterogeneity, which often cuts across individual SSDs and even MACRO SCs. For instance, research staff members belonging to Business, Economics, and Mathematical Methods share research interests in the field of finance. Likewise, sophisticated empirical methodologies are applied not only by statisticians but also by applied economists. Some sub-fields such as game theory are studied both by mathematicians and economic theorists. A final example, among many others, is industrial economics, which witnesses the contributions of economists and business economists as well.

A significant number of researchers in the Area is also active in a multidisciplinary context: experts of experimental statistics, for instance, often collaborate with medical scientists and engineers, while theoreticians do so with mathematicians. Economists and business economists interact with sociologists, political scientists, psychologists, and jurists, commodity scientists do so with experts in chemical and environmental sciences.

In terms of publication style, the vast majority of the econometricians, statisticians, and mathematicians belonging to the Area have an established tradition of publication in international journals. At the opposite extreme, for economic historians (including specialists in economic thought) research monographs play an important role. Monographs and contributions to volume represent a significant portion of research output also for some research areas within Business. It is worth noting, however, that differences in publication style are also present within most SSDs, because of persistent different traditions, schools of thought, and analytical methods.

As a result of the coexistence of research products published in international indexed journals with monographs and contributions to volume, in terms of the availability of evaluation methodologies Area 13 is in intermediate position, between the “hard” sciences, on the one hand, and the humanities and most of the other social sciences, on the other. While most of the international journals which are relevant to the Area are indexed in the WoS and Scopus databases, other journals, and especially Italian journals, are not indexed. This is the reason why evaluation of journal articles for Area 13 within the VQR has been based, starting with VQR 2004-2010 (hereafter, VQR1), on a journal list compiled by GEV13 that includes non-indexed journals.
As mentioned before, during the period covered by VQR 2011-2014 there have been substantial changes, both external and internal to the Area. Starting with the former, the growing relevance of the Scopus database in terms of coverage has allowed GEV13 to expand the fraction of indexed journals in the GEV list. At the same time, as the analysis below will document, there has been a marked evolution in the characteristics of the products submitted to the VQR, along several interrelated dimensions. In terms of publication types, the share of journal articles has increased significantly, with a parallel reduction of the share of book chapters and monographs. In terms of language of publication, the share of products in English also increased considerably. As a result, the fraction of products that received a bibliometric evaluation also increased. It is worth noting that, while these trends have affected different fields at a different pace, they are visible also in the SSDs that traditionally rely on monographs and contributions to volume.

Taken together, the recent and fast developments described above witness a tendency toward internationalization for the Area as a whole, which is confirmed also in other dimensions. For instance, in recent years some Universities and Research Institutions have started to recruit on the international job market. Likewise, Italian Ph.D. students increasingly have found placement in foreign Institutions.

2 Evaluation of “research products”

2.1 Evaluation criteria

The evaluation of products by GEV13 has followed the methodology described in the document entitled “Criteria for the Evaluation of Research Outputs” (Appendix A).

In particular, GEV13 has applied the peer-review evaluation methodology, either external or internal to the GEV, to monographs, contributions to volume, and articles in journals not included in GEV13 journal list. On the other hand, bibliometric analysis, in combination with the expert judgment of the GEV, has been employed for articles in journals included in the GEV journal list. The final merit class of each bibliometric product was determined by (i) the merit class of the corresponding journal, as determined by the GEV journal classification based on four different bibliometric indicators, and (ii) the individual citations received by the article. No a priori differentiation was applied to the evaluation of alternative types of journal article (e.g., scientific articles, review essays, letters, etc.). A random sample of bibliometric products has also been sent to external reviewers for comparison and statistical purposes (for an analysis we refer to the ANVUR VQR Final Report).
The abovementioned evaluation criteria have been integrated with several public announcements whose content is summarized in the following. They are also included in Appendix A together with the evaluation criteria, of which they form an integral part.

2.2 Journal classification

The journal list for Area 13 has been compiled starting from the journal list available from VQR1, which has been expanded with selected journals from three sources: (i) the list provided by CINECA containing the publication outlets of all Italian researchers in Area 13 for the period 2011-2014, (ii) additions to the WoS database provided by Thomson Reuters, and (iii) the Scopus database provided by Elsevier. To be noted that the Scopus database was not considered for VQR1, because of its limited coverage at the time.

The GEV published a revised, preliminary journal list on the ANVUR website on November 20, 2015. Each included journal was equipped with a vast array of bibliometric information. The list was divided into five non-overlapping sub-lists: Business Administration and Management (hereafter, Business), Economics, Economic History (hereafter, History), General, and Statistics and Mathematical Methods for Decisions (hereafter, Statistics). Authors could submit articles published in journals in any sub-list regardless of their SSD or the SSD assigned by them to the articles. After collecting suggestions from the academic community, a final list was published on December 14, 2015. For the final classification the GEV selected the following four bibliometric indicators, leaving to the authors the choice of their preferred one: for WoS (https://www.webofknowledge.com), the five-year Impact Factor (IF5) and the Article Influence Score (AIS) and, for Scopus (http://www.journalmetrics.com), the Impact per Publication (IPP) and the SCImago Journal Rank (SJR).

While IF5 and IPP can be interpreted as measures of popularity, AIS and SJR can be interpreted as measures of prestige. The GEV's decision to rely on these complementary bibliometric indicators jointly is consistent with the use of bibliometric analysis by the international scientific community.

For non-indexed and partially-indexed journals, missing bibliometric indicators were imputed on the basis of the h-index from Google Scholar. Finally, separately for each sub-list (except General), the GEV produced a classification into the five merit classes reflecting the VQR percentiles (10/20/20/30/20) in terms of each of the four indicators IF5, AIS, IPP, and SJR. The 3 journals in category General were all assigned to the Excellent merit class due to their widely recognized standard.
As specified in the evaluation criteria, the GEV upgraded by one class, in all cases from Acceptable to Fair, 9 Italian journals, of which 4 from Business, 2 from Economics, 1 from History, and 2 from Statistics.

The journal classification used in the evaluation process was made available on the ANVUR website on January 22, 2016 and can be consulted at the following url: http://www.anvur.it/attachments/article/856/22_01_2016_riviste.xls. More detail on the GEV journal classification is available in Appendix B.

2.3 The evaluation process

The evaluation process was carried out as follows. Preliminarily, ANVUR took care of duplicate products submitted by co-authors affiliated with different institutions, in order to assure they would receive a uniform evaluation. Next, the GEV Coordinator and the subGEV Coordinators, jointly with the GEV Coordinators of other Areas, processed the requests from authors belonging to Area 13 to be evaluated by other GEVs and those from authors belonging to other Areas to be evaluated by GEV13. Finally, the subGEV Coordinators allocated products to GEV members. Products were assigned to subGEVs according to the SSD indicated by the author for the product. Each product was assigned to two GEV members according to their expertise, following closely the VQR policy regarding conflicts of interests. Each GEV member in charge did not know the identity of the other one and selected reviewers independently. The only exceptions to the rule requiring two GEV members to be in charge for each product were products authored by the GEV members themselves, which were handled by the GEV Coordinator, and products authored by the GEV Coordinator, which were handled by the VQR Coordinator.

2.3.1 Peer evaluation

Products to be peer-reviewed (i.e., monographs, contribution to volumes, articles published in journals outside the GEV list, plus a few other kinds of less-represented types such as conference proceedings, encyclopedia entries, patents, etc.) were preliminarily screened in order to identify products not eligible for evaluation assessable products, that is, not assessable (such as working papers, textbooks with no research content, curatorships not accompanied by an introduction or essay, unpublished mimeographs, abstracts, unsigned policy reports, slides, products not published within the VQR timeframe, duplicates submitted by the same institutions, etc.). All the products that were reported as potentially not assessable by GEV members were double-checked by the GEV Coordinator and the appropriate subGEV Coordinator to assure uniformity of
judgment. In case the pdf-document was missing or inadequate, institutions were contacted and requested to provide an adequate one.

The reviewer database available from VQR1 was expanded with other reviewers proposed by GEV members and screened by the GEV Coordinator and the subGEV Coordinators. Proposals included information on the reviewer’s expertise (including titles of recent publications) and scientific quality (proxied by the $h$-index). Throughout the evaluation process reviewers were added as needed, following the same procedure. It is worth noting that each GEV member had the option to select reviewers within the entire reviewer database assembled for the VQR upon proposals by all GEVs. The anonymity of reviewers was carefully preserved during the entire process. The evaluation form prepared by GEV13 according to the VQR Call and the accompanying reviewer guidelines are reported in Appendix C.

For every product sent to peer review, each GEV member was expected to obtain a single review, but to expedite the evaluation process GEV members had the option to indicate multiple reviewers, to be contacted sequentially in case of refusal or delay. GEV members also had the option to provide an internal review themselves. Once two reports were received, each GEV member had several options: accept the merit class proposed by the interface and computed averaging out the scores submitted by the reviewers; suggest a class other than the proposed one, with a motivation; propose another external or internal review; propose a consensus group including a third GEV member, most often the subGEV Coordinator.

Considering all products evaluated by GEV13 (including those sent by other GEVs upon requests of research staff members belonging to other Areas and excluding those by research staff members of Area 13 sent to other GEVs), reviewers agreed on the merit class in 34% of the cases, while GEV members agreed in 84% of the cases, even though in both cases the degree of disagreement could be limited to only one merit class (we refer to the ANVUR VQR Final Report for more detail). The level of agreement was similar across subGEVs. Additional external reviews were requested by GEV members in 2 cases, additional internal reviews were provided in 22 cases, and consensus groups were activated in 47 cases. In case of disagreement between GEV members, the decision about the final merit class was made by the subGEV Coordinator (or the GEV Coordinator if the subGEV Coordinator was one of the two GEV members in charge).

2.3.2 Bibliometric evaluation

Bibliometric products were also screened to verify their admissibility (publication within the VQR timeframe, absence of duplicates submitted by the same institutions, presence of an
adequate pdf-document, etc.), by the two GEV members in charge, as well as by the GEV Coordinator and the subGEV Coordinator of the relevant subGEV. In case the pdf-document was missing or inadequate, institutions were contacted and requested to provide one.

Prior to the assignment to GEV members, the GEV Coordinator and the subGEV Coordinators examined suggestions by authors to evaluate a bibliometric product using peer review, an option offered by the VQR Call subject to a set of reasons (emerging areas, areas of high specialization, interdisciplinary character). Peer reviews were requested for only 304 (6%) bibliometric products. All requests were individually screened and rejected, with a motivation, because authors provided reasons incoherent with the Call.

Consistently with the evaluation criteria, bibliometric products were initially assigned the merit class corresponding to the indicator chosen by the authors. In 307 cases the authors did not choose and, as previously announced, the products were assigned the highest class or, if one of the two databases was selected, the higher one out of the two indicators corresponding to the selected database. In 190 cases the authors failed to select the indicator associated with the highest class. Once a merit class was assigned, the GEV still had the option to evaluate a product through peer review, provided that both GEV members in charge would agree. Only 9 bibliometric products were evaluated through (internal) peer review. In all cases, they belonged to sub-types of the journal article category (such as rejoinders, editorials, comments, etc.).

Out of the products that were evaluated bibliometrically, on the basis of an analysis of individual citations the GEV upgraded by one class 50 journal articles, that is, 1% of the submitted articles published in journals included in the GEV journal list. Since the selected articles were in certain cases submitted by more than one co-author from different institutions, a total of 59 authors received an upgrade. Due to the general delay in the evaluation process individual citations were collected in July 2016 rather than in February 2016 (as previously indicated in a public announcement published on January 29, 2016).

Articles were ranked according to the ratio between their average annual citations (since publication) over the corresponding value of IF5 or IPP in 2014, as reported in the GEV journal list, whereas the choice of the database was left to the author (coherently with the author's choice of the indicator to be used for bibliometric analysis). To compute the number of average annual citations, the number of individual citations was divided by 5, 4, 3, and 2 for articles published in 2011, 2012, 2013, and 2014, respectively. The ranking included only articles published in journals with a strictly positive value of IF5 or IPP, with an evaluation below Excellent, and with
a number of citations per year strictly larger than the journal impact. This set of prerequisites was satisfied by 470 articles.

The 50 higher-ranking articles were selected separately from submissions associated with the WoS and Scopus databases, proportionally to all submissions in indexed journals (excluding articles classified as Excellent and those in journals with not strictly positive indicators, and including duplicate submissions). Since 38% and 62% of the resulting submissions were respectively associated with WoS and Scopus, 19 of the upgraded articles were associated with WoS and 31 with Scopus. On average (including duplicate submissions and products classified as Excellent) the ratio between article citations per year and journal impact was 1.34 for Scopus and 0.65 for WoS. This is reflected in the different thresholds for upgrades. For WoS upgrades the ratio was between 11.4 and 2.8, while for Scopus it was between 23.8 and 6.3. Out of the 59 co-authors obtaining the upgrade, the corresponding journals belonged to the Business, Economics, History, and Statistics sub-lists in 9, 31, 2, and 17 cases, respectively, and the upgrade occurred up from the Good, Fair, Acceptable, and Limited merit class in 38, 10, 10, and 1 cases, respectively.

No downgrades were applied on the basis of individual citations and self-citations were not excluded from the number of individual citations.

Appendix D contains a broad analysis of the individual citations of the products published in indexed journals that were submitted to GEV13.

### 2.3.3 Statistics

The first three tables of this section document the peer-review activity of the GEV. Information is reported only for reviewers and reviews that contributed to the actual evaluation of the Area, thus excluding information about reviewers and reviews used for the sample of bibliometric products sent to peer review only for comparison purposes. All GEV members acted as reviewers, providing 242 internal reviews, but GEV members acting as reviewers and their reviews are also excluded from the tables. The number of external reviewers contacted for Area 13 was 920. However, in the three tables below reviewers are counted more than once if they contributed reviews for products associated with to more than one SSD. It is also worth stressing that in this initial set of tables SSDs are those assigned by authors to products (which may differ from the SSD of the authors).
Table 2.1 reports the number of reviewers contacted by the GEV, by subGEV and by nationality of the reviewer affiliation. The table shows that 76% of reviewers were affiliated with Italian institutions, with a higher percentage (81%) for the Statistics subGEV.

Table 2.1: Number of reviewers by subGEV and nationality (Italian and not Italian); reviewers repeated for each SSD of expertise.

The first panel of Table 2.2 reports again information on reviewer nationality but at the finer SSD level. There is evidence of some variation which should, however, be taken with care since in some SSDs, such as SECS-P/05, there were very few products evaluated through peer review. The second panel presents data on reviews, again by SSD and nationality of the reviewers, and shows that 77% of the reviews was assigned to reviewers affiliated with Italian institutions, again with some variation across SSDs.

Table 2.2: Number of reviewers and reviews by SSD and nationality (Italian and not Italian); reviewers repeated for each SSD of expertise.

Table 2.3 shows, in the first panel, that out of 6,400 assigned reviews, 4,907 (77%) were completed, while 572 were not completed (9%) and 921 (14%) were explicitly refused. The percentage of completed reviews was slightly higher for Italian reviewers (77%) if compared to foreign ones (75%). The same table, in the second panel, reports the motivations for refusals to provide a review. In the majority of cases (54%) reviews were refused because the reviewer did not feel sufficiently competent.

Table 2.3: Number of assigned, completed, not completed, and refused reviews by reviewer nationality (Italian and not Italian).

The next set of tables describes the products under evaluation. Table 2.4 shows that out of 9,039 expected products to be submitted by staff members of the Area, GEV13 received 8,180 products, while 205 were submitted to other GEVs, for a total of 8,385. Subtracting multiple submissions by authors affiliated with different institutions, the GEV received 7,708 distinct products. The fraction of distinct products shown in the table (91.93%) is computed by dividing 7,708 by 8,385. GEV13 also received 120 products from staff members belonging to other Areas. To be noticed is that a monograph is always counted twice if so requested by the staff member that submitted it, according to a provision of the VQR Call. This occurred in 83 cases.

Table 2.4: Expected and submitted products for Area 13 and number of distinct products.

Table 2.5 shows that, for the 8,385 submitted products, the prevalent type of publication was represented by journal articles (73%), followed by contributions to volume (15%), research
monographs (9%), and other less represented types. The distribution by type was quite stable along the period covered by the current VQR, while it has changed considerably if compared with VQR1: for the previous VQR journal articles represented only 62% of the products, while contributions to volume and monographs represented 20% and 13%, respectively. Table 2.5 also shows that for the current VQR products were well distributed across the relevant four years of publication, 2011 to 2014, with a slight prevalence of products published in 2014 (29%).

Table 2.5: Products submitted to the Area by type and year of publication.

Table 2.6 describes how the main publication types were distributed by SSD of staff members (SSD_RSM). The share of journal articles was higher than average (above 84%) for SECS-P/05, SECS-S/01, SECS-P/01, and SECS-S/06. At the other extreme the fraction of journal articles was below average in SECS-P/12 (31%, despite an increase of 4 percentage points with respect to VQR1), where contributions to volume and monographs remain quite common because of the intrinsic characteristics of the discipline. The last column of the table provides information about the percentage of distinct products, which was lowest at around 90% for SECS-S/02 and SECS-S/06, implying for these sectors a higher percentage of products submitted by co-authors from different institutions.

Table 2.6: Distribution of submitted research products by publication type and SSD of research staff members.

Table 2.7 reports information on the language of publication. Overall, 77% of the submitted products were published in English (up from 57% for VQR1) and 23% in Italian, while other languages played a negligible role. Significant differences emerge among SSDs, with SECS-P/05, SECS-S/01, and SECS-S/02 showing English as the language of choice for over 95% of the submissions, and SECS-P/12 once again at the other extreme with a prevalence of publications in Italian (64%). These patterns point to a correlation between publication type and language, with an association between journal articles and English on the one hand, and monographs and Italian on the other.

Table 2.7: Distribution of submitted research products by language of publication and SSD of research staff members.

Table 2.8 combines information from previous tables by showing how submitted research products were distributed by publication type and SSD, and how this distribution varied by year of publication between 2011 and 2014. Even though, as previously remarked, the overall distribution of publication types was relatively stable along the four years covered by the VQR (see Table 2.5), some SSDs show a remarkable increase in the share of journal articles. SECS-P/04 went from 29% in 2011 to 41% in 2014, with a peak at 75% in 2013, even though the sector
was represented by a low number of products per year. A gradual shift toward journal articles is also noticeable for SECS-S/05 and SECS-S/04 within Statistics and SECS-P/10 and SECS-P/08 within Business, while within Economics SECS-P/12 showed a stable share around 32% and SECS-P/02 experienced a decline.

**Table 2.8: Distribution of submitted research products by publication type, year of publication, and SSD of research staff members.**

Table 2.9 documents that the fraction of submitted over expected products (93% on average) varies across SSDs, with SECS-P/12 showing the best performance (97%) and SECS-P/09 the worst (84%).

**Table 2.9: Number of submitted and expected products by SSD of research staff members.**

In Table 2.10 research staff members are divided according to the number of products they were expected to submit to the VQR. The vast majority was expected to submit two products, since most staff members in the Area are affiliated with Universities and young researchers represent a minority.

**Table 2.10: Number of research staff members by corresponding number of expected products and by SSD of research staff members.**

Table 2.11 shows that GEV13 evaluated the vast majority of the products submitted by staff members belonging to Area 13 (98%). The remaining products (142) are distributed across all other GEVs (except GEV10) with negligible percentages in each of them. GEV13 received from staff members belonging to the Area 276 requests to be evaluated by other GEVs. They were individually screened and only 51% of them were accepted.

**Table 2.11: Number and percentage of research products submitted by research staff members of Area 13 by GEV that evaluated them.**

Conversely, almost all the products evaluated by GEV13 (99%) were submitted by staff members of Area 13 (Table 2.12).

**Table 2.12: Number and percentage of research products evaluated by GEV13 by Area of research staff members.**

GEV13 judged as not eligible for evaluation, that is, as not assessable, 192 products, spread around all SSDs except SECS-P/13 (Table 2.13).

**Table 2.13: Number of research products submitted to GEV13 and classified as not assessable by SSD of research staff members.**
Table 2.14 shows how the resulting 8,193 assessable products (i.e., the 8,385 products submitted to GEV13 minus 192 not assessable ones) are distributed according to the evaluation method, that is, peer review, bibliometric analysis, and informed peer review. Informed peer review was only applied to 10 products evaluated, as requested by the authors, by other GEVs. According to the evaluation criteria of the bibliometric GEVs, a bibliometric product would have received also a peer-review evaluation had their algorithm, combining journal impact and individual citations, produced an undetermined merit class. In such case, the reviewer would have been informed about the bibliometric information on the product. The table shows that bibliometric analysis was applied to almost 68% of the products, while 32% were evaluated through peer review. The distribution of products by evaluation method differs sharply across subGEVs and, as shown in the table, even across SSDs within the same subGEV. For Economics, the share of bibliometric products ranges from 20% for SECS-P/12 to 82% for SECS-P/01, for Business from 39% for SECS-P/13 to 65% for SECS-P/08, for Statistics from 53% for SECS-S/05 to 95% for SECS-P/05. These differences once again can be mostly attributed to different publication styles. The relatively low share observed for SECS-P/13, however, is largely due to the fact that staff members in this SSD often publish in interdisciplinary and/or technical journals that are not included in the GEV journal list, so that the corresponding articles were evaluated through peer review. For VQR1, on the other hand, only 53% of the products were evaluated bibliometrically. The current increase can be attributed both to the expansion of the GEV journal list and to the evolution in publication habits in the Area.

It is also worth reporting that, out of the bibliometric products submitted to the VQR, only 5% were published in non-indexed journals.

Table 2.14: Percentage of submitted products evaluated in peer review, bibliometric analysis, and informed peer review, by SSD of research staff members.

2.4 Evaluation results

This section presents aggregate evaluation results for the Area. Each table reports aggregate and average scores as well as the distribution of the VQR final merit classes, that is, Excellent, Good, Fair, Acceptable, and Limited (with weights equal to 1, 0.7, 0.4, 0.1, and 0, respectively). Products classified as not eligible for evaluation and therefore not assessable are assigned a weight equal to 0. Results are presented by evaluation method, type and language of publication, and by SSD and subGEV of research staff members. It is worth noting that, in the next tables,
SSDs and the corresponding subGEVs are those of the research staff members (SSD_RSM), which may not coincide with those of the products.

Table 2.15 shows that the average score for the 8,193 submitted and assessable products of the Area is 0.51. Products classified as Excellent, Good, Fair, Acceptable, and Limited represent about 25%, 23%, 18%, 20%, and 13% of the products, respectively. If we distinguish by evaluation method, we find that for bibliometric products the average score is 0.65, while it is 0.21 for the peer-reviewed products. Turning to the distribution by merit class, 37% of the bibliometric products were classified as Excellent and 8% as Limited, against respectively 1% and 24% of the peer-reviewed ones. These aggregated data reflect the differences within subGEVs and SSDs, both in terms of publication outlets and consequent evaluation methods, previously documented.

Table 2.15: Scores and distribution of submitted products by merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E) and by evaluation method.

Table 2.16 reports evaluation results by publication type. It is worth noting that the products classified by the GEV as not assessable are now included, so that the average score is now only 0.5. Apart from patents, which show the highest average score (0.7) but are represented by only 2 products, the average score is higher than average (0.62) only for journal articles, which represent the largest share of products. The average score for the other types is well below average.

Table 2.16: Scores and distribution of submitted products by merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F) and by publication type.

Table 2.17 presents evaluation results by SSD and subGEV of research staff members, taking into account also the weight equal to 0 assigned to missing products. The resulting average score for the Area is 0.46. Significant differences appear across subGEVs. Economics and Statistics are both above average (0.53 and 0.52, respectively), while Business is below average (0.36). The distribution of the scores is also quite different, with Economics and Statistics showing, at the two extremes, both a larger share of products classified as Excellent and a lower share of products classified as Limited. Within each subGEV, again we can find marked differences across sectors. Within Economics, for instance, SECS-P/01 (0.58) scores much higher than SECS-P/04 (0.35) or SECS-P/12 (0.39), while the other SSDs are closer to the mean for the subGEV and still above the mean for the Area as a whole. It is within the Statistics subGEV that we find the SSD with the highest average score in the Area, SECS-P/05 (0.74), while other SSDs are below the mean for the Area (SECS-S/03, SECS-S/05, and SECS-S/04). Within Business, the
highest average score is achieved by SECS-P/08 (0.43), still below the mean for the average, while the lowest scores are reported for SECS-P/11 and SECS-P/09 (0.24 and 0.26, respectively). The table also reports information on missing products, with a particularly high fraction for SECS-P/09 (16%) and SECS-S/06 (13%), against a fraction equal to 7% for the Area as a whole.

Even though a comparison with VQR1 should be taken with extreme care, for the reasons detailed in Section 6, it is worth noting that the current data show for the Area a marked improvement. For instance, for VQR1 the fraction of Excellent products was only 18%, below the *ex ante* probability for a product to fall into the corresponding merit class (then 20%), with a particularly low performance for the Business subGEV (8%). For the current VQR, the fraction of Excellent products is 23%, above the corresponding *ex ante* probability (now 10%), with a fraction equal to 14% for Business. At the other extreme of the distribution by merit class, for VQR1 49% of the products fell below the median (with Business, Economics, and Statistics at 67, 37, and 38%, respectively), while now the equivalent sum of the Limited and Acceptable products is only 30% (with Business, Economics, and Statistics at 41, 23, and 21%, respectively).

Table 2.17: Scores and distribution of products by merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F) in the Area, by SSD and subGEV of research staff members.

Table 2.18 reports results by year of publication at the SSD level, which confirm the above described differences across SSDs. A larger average score for more recent products is noticeable for some SSDs, especially SECS-P/08 and SECS-P/10.

Table 2.18: Scores and distribution of submitted products by merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F), by SSD of research staff members, and by year of publication.

Table 2.19 provides information by type and language of publication. The highest average score is associated with journal articles published in English (0.67), while journal articles published in Italian obtain only 0.16. English is associated with higher average scores also for monographs and contributions to volume, even though the gap for Italian ones is now considerably reduced.

Table 2.19: Scores and distribution of submitted products by merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F) and by type and language of publication.
Table 2.20 disaggregates results by type of publication at the SSD level. Journal articles dominate other publication types, in terms of average score, for all SSDs except SECS-P/04, where monographs obtain a higher average score.

Table 2.20: Scores and distribution of submitted products by merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F), by SSD of research staff members, and by type of publication

3 Evaluation of Institutions

The GEV had the task to evaluate research products submitted by Institutions in order to collect information to compute the VQR indicators IRAS1, IRAS2, and IRAS5. In this section we focus in particular on the evaluation of the quality of the submitted products, introducing indicators computed from the same information to be used in order to determine IRAS1.

As previously recalled, on the basis of the VQR Call, individual products are assigned weights equal to 1, 0.7, 0.4, 0.1 and 0 for Excellent, Good, Fair, Acceptable, or Limited level, respectively; missing products are assigned a weight equal to 0; products which are not eligible for evaluation (i.e., not assessable) are assigned a weight equal to 0.

Defining respectively as \( n_{i,j,EC} \), \( n_{i,j,EL} \), \( n_{i,j,D} \), \( n_{i,j,A} \), \( n_{i,j,LIM} \), \( n_{i,j,MAN} \), and \( n_{i,j,NV} \) the number of Excellent, Good, Fair, Acceptable, Missing, and Not Eligible products of the \( i \)-th institution in the \( j \)-th scientific Area, the aggregate evaluation \( v_{i,j} \) of the \( i \)-th institution in the \( j \)-th Area can be obtained as:

\[
v_{i,j} = n_{i,j,EC} + 0.7 \cdot n_{i,j,EL} + 0.4 \cdot n_{i,j,D} + 0.1 \cdot n_{i,j,A} + 0 \cdot (n_{i,j,LIM} + n_{i,j,MAN} + n_{i,j,NV})
\]

In the next sections we propose three indicators of research quality which do not depend on the number of the research staff members in the institutions of the Area and the IRAS1\( _{i,j} \) indicator defined in the DM and in the Call, which takes into account both research quality and the number of the research staff members evaluated in the institution and belonging to the Area.

The value of \( v_{i,j} \) is the basis for the computation of the quality indicators for research output we propose in the following.

Since they do not take into account the size of an Institution, the first three indicators cannot be used by themselves for resource distribution. Nevertheless, they provide useful information on research quality for the Institutions in a given Area.
3.1 Indicators

3.1.1 The first indicator

Setting as \( n_{i,j} = n_{i,j,EC} + n_{i,j,EI} + n_{i,j,DI} + n_{i,j,A} + n_{i,j,LIM} + n_{i,j,MAN} + n_{i,j,NV} \) the number of expected products for VQR 2011-2014 for the \( i \)-th Institution in the \( j \)-th Area, the first indicator \( I_{i,j} \), defined between 0 and 1, is given by:

\[
I_{i,j} = \frac{v_{i,j}}{n_{i,j}} \tag{2}
\]

It represents the average score obtained by Institution \( i \) in Area \( j \).

3.1.2 The second indicator

Setting again as \( n_{i,j} \) the number of expected products for VQR 2011-2014 for the \( i \)-th Institution in the \( j \)-th Area, and as \( N_{IST} \) the number of Institutions, the second indicator is given by:

\[
R_{i,j} = \frac{v_{i,j}}{n_{i,j}} = \frac{I_{i,j}}{V_j/N_j} \tag{3}
\]

where \( V_j \) and \( N_j \) stand for the aggregate evaluation and the total number of expected products in the \( j \)-th Area, that is:

\[
V_j = \sum_{i=1}^{N_{IST}} v_{i,j}, \quad N_j = \sum_{i=1}^{N_{IST}} n_{i,j} \tag{4}
\]

Indicator \( R_{i,j} \) represents the ratio between the average score received by the products of the \( i \)-th Institution in the \( j \)-th Area and the average score received by all the products in the \( j \)-th Area. It provides a direct measure of relative research quality in a given Area as expressed by a specific Institution: values above 1 indicate a research output whose quality is above the average for the Area, values below 1 indicate below average quality.

3.1.3 The third indicator

The third indicator \( X_{i,j} \) is given by the ratio between the fraction of Excellent and Good products of the Institution in the Area and the fraction of Excellent and Good products in the
Area. Values of $X_{i,j}$ above 1 indicate that the Institution has a percentage of Excellent and Good products higher than the average in the Area. Formally:

$$X_{i,j} = \frac{n_{i,j,EC} + n_{i,j,El}}{\frac{\sum_{i=1}^{N_{IST}} (n_{i,j,EC} + n_{i,j,El})}{\sum_{i=1}^{N_{IST}} n_{i,j}}}$$

(5)

### 3.1.4 The IRAS1$_{i,j}$ indicator of the VQR Call

The IRAS1$_{i,j}$ indicator is defined in the VQR Call as the ratio between the total score achieved by an Institution in a given Area and the total evaluation of the Area itself:

$$IRAS1_{i,j} = \frac{v_{i,j}}{\sum_{i=1}^{N_{IST}} v_{i,j}} = \frac{v_{i,j}}{V_j}$$

(6)

It can be written as the product of the relative quality of the products submitted by a specific Institution in a given Area and an indicator of the size of the Institution in the same Area. The quality indicator is given by the ratio between the average score received by the products of the $i$-th Institution in the $j$-th Area and the average score received by all the products in the $j$-th Area, and coincides with the second indicator $R_{i,j}$ defined in (3), while the weight of the Institution ($P_{i,j}$) is simply given by the share of expected products for the $j$-th Area attributed to the $i$-th Institution:

$$IRAS1_{i,j} = \frac{v_{i,j}}{\sum_{i=1}^{N_{IST,j}} \frac{n_{i,j}}{N_j} \cdot \frac{I_{i,j}}{V_j} / \frac{n_{i,j}}{N_j} = R_{i,j} \cdot P_{i,j}}$$

(7)

The IRAS1$_{i,j}$ indicator re-defines the weight of an Institution in an Area, as measured by the share of expected products, on the basis of the relative quality of the expected products themselves. Thus, IRAS1 is a useful indicator especially for the allocation of funds across Institutions within the same Area, since it takes into account both the quality and the relative size of an Institution.
3.2 Rankings of Institutions

3.2.1 Universities

This section presents the rankings for the Universities in the Area. Table 3.1 lists the Universities involved in the VQR in alphabetical order and provides a set of information including, for each of them, the sum of the scores achieved in the evaluation, the number of expected products, the average score, the normalized average score, the position in the overall ranking, the overall number of Institutions, the size class, the ranking in the size class, the number of Institution in the size class, the percentage of the sum of Excellent and Good products, and the associated indicator X. For reasons of privacy, the table includes only Universities that submitted at least 5 products to the VQR, that is, the 82 Universities with more than 2 staff members. However, the Universities not shown in the table are still included in the computation of all rankings. The same rule is applied to all subsequent rankings. The average score \( I \) for the Universities of the Area is 0.46, which is well above the median and appears strongly correlated with the fraction of Excellent and Good products.

*Table 3.1: List of Universities in alphabetical order.*

It is preferable to comment in more detail on the rankings within the appropriate size class. The rankings by size class are presented in the next three tables (Tables 3.2-3.4), separately for small, medium-size, and large Universities. Size is defined in terms of expected products, and the small, medium-size, and large Universities shown in the tables are associated with a number of expected products between 5 and 100, between 101 and 250, and greater than 250, respectively. The rankings are obtained on the basis of the normalized average score \( R \). The tables report additional information, including the percentage of products in each merit class, the percentage of missing products, and the relative size in terms of expected products. Relative size allows the computation of the IRAS1 indicator, combining quality with quantity, which is inserted in the last column. Table 3.2 covers the 43 small Universities, Table 3.3 the 33 medium-size Universities, and Table 3.4 the 6 large Universities. The largest University in the Area is Bologna. Lucca - IMT, Padova, and Milano Bocconi rank first in the corresponding small, medium, and large size classes. Large Universities all show average scores above the mean for the Area as a whole. At the other extreme, only 11 out of the 43 small Universities achieve an average score above the mean of the Area, that is, a value of \( R \) greater than 1, which happens instead for about half of the medium-size Universities. Thus, the emerging correlation between size and quality appears to be driven, on the one hand, by the low scores of certain very small Universities and, on the other, by the good performance of the large ones. Among small
Universities, only 2 of the 11 reporting an above average score are located in the South. Among medium-size ones, this is true for 4 out of 16, while large Universities are all located in the North with the exception of Roma La Sapienza. This suggests a persistence of the gap between the South and the rest of the country in terms of research quality.

The next four tables offer a more detailed analysis at the finer subGEV level. Table 3.5 lists all Universities where a specific subGEV is represented, while Tables 3.6-3.8 present the corresponding rankings by size class. While for Universities, after inspection of the overall distribution of expected products, size classes were determined by referring to 100 and 250 expected products as thresholds, for subGEVs (as well as for other sub-entities such as SSDs, Departments, etc.) size classes are determined endogenously, by dividing by 3 the interval of the values of the number of expected products within the subGEV, without counting subGEVs with fewer than 5. For example, since for the first subGEV being considered, Business, the maximum number of expected products observed in Table 3.5 is 242 (for Milano Bocconi) Universities are defined as small, medium-size, and large, with respect to the Business subGEV, when they were expected to submit to the VQR 5-84, 85-163, and 164-242 products, respectively, by staff members belonging to the Business subGEV. For illustrative purposes, again we report the best performing Universities. Among small Universities in a given subGEV, Pisa S. Anna displays the highest score both for Business and Economics, Roma LUISS for Statistics. For the intermediate class, Bologna is first for Business, and Milano Bocconi both for Economics and Statistics. For the large class, which includes very few Universities, Milano Bocconi is first for Business, Bologna both for Economics and Statistics. This implies that Bologna and Milano Bocconi lead the rankings at the subGEV level for all subGEVs, once their appropriate size class is accounted for.
Table 3.7: Ranking of medium-size Universities by subGEV of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 3.8: Ranking of large Universities by subGEV of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Next, Tables 3.9-3.12 are organized at the SSD level. Again, size classes are determined endogenously according to the rule previously explained for subGEVs, so that thresholds necessarily differ across SSDs. The analysis at the SSD level uncovers significant heterogeneity. For instance, some Universities show a good performance in a particular SSD despite less satisfactory results at the Area or subGEV level. At the same time, some Universities tend to rank consistently in relatively high positions in several SSDs.

Table 3.9: List of Universities in alphabetical order for all SSDs in the Area.

Table 3.10: Ranking of small Universities by SSD of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 3.11: Ranking of medium-size Universities by SSD of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 3.12: Ranking of large Universities by SSD of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Finally, in Tables 3.13-3.16 rankings are organized by MACRO SC, leading to considerations in line with those applying to SSDs and subGEVs. It should be noticed, however, that SECS-P/05 is associated with the Statistics subGEV but it belongs to the Economics MACRO SC, while SECS-P/04 and SECS-P/12 are associated with the Economics subGEV but belong to the Economic History MACRO SC.

Table 3.13: List of Universities in alphabetical order for all MACRO SCs in the Area.

Table 3.14: Ranking of small Universities by MACRO SC of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 3.15: Ranking of medium-size Universities by MACRO SC of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 3.16: Ranking of large Universities by MACRO SC of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

3.2.2 Research Institutions

In this section we report results for two types of Research Institutions. The first set of tables (Tables 3.17-3.21) focuses on Research Institutions controlled by the Italian Ministry of Education, University and Research, and other assimilated Research Institutions, that is,
Institutions that submitted their research products to the VQR on a voluntary basis under the same rules applying to the MIUR-controlled Institutions. Institutions with fewer than 7 expected products are not reported, since for Research Institutions the number of expected products according to the VQR Call was larger than for Universities. This difference needs to be accounted for when comparing results with those concerning Universities. The tables report information, in alphabetical order, for only 3 Institutions: Consiglio Nazionale delle Ricerche (CNR), Consiglio per la ricerca in agricoltura e l’analisi dell’economia agraria (CREA), and Fondazione Bruno Kessler (FBK). Only CNR presents a sizeable number of expected products (156) and is represented in all the subGEVs of the Area. Similarly, only 5 SSDs are represented among this type of Research Institutions, all of them only within CNR. It is worth noting that for CNR the association between research staff members and SSDs was established for the purposes of the VQR. The average score \( I \) for CNR is 0.37, the lowest among the other Research Institutions of this kind, and below the average score for Universities (0.46). It should be stressed that the normalized average score \( R \) used for the rankings is computed separately for Universities and Research Institutions, that is, the denominator is the average score received by all the products in the Area by Universities and Research Institutions, respectively. Therefore, the values of \( R \) reported for Research Institutions cannot be compared with those of Universities. However, keeping in mind institutional differences, it is legitimate to compare the average score \( I \), i.e., the numerator of \( R \). For Research Institutions no differentiation based on size is presented.

<table>
<thead>
<tr>
<th>Table 3.17: List of controlled and assimilated Research Institutions in alphabetical order.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.18: Ranking of controlled and assimilated Research Institutions on the basis of the normalized average score ( R ) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).</td>
</tr>
<tr>
<td>Table 3.19: Ranking of controlled and assimilated Research Institutions by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).</td>
</tr>
<tr>
<td>Table 3.20: Ranking of controlled and assimilated Research Institutions by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).</td>
</tr>
<tr>
<td>Table 3.21: Ranking of controlled and assimilated Research Institutions by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).</td>
</tr>
</tbody>
</table>

Tables 3.22-3.26 refer to other Research Institutions that submitted their research products to the VQR on a voluntary basis. Again only 3 Institutions are listed (in alphabetical order), the largest one being Collegio Carlo Alberto, for which an average score of 0.84 is reported. The
other two are Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) and Consorzio Universitario di Economia Industriale e Manageriale (CUEIM).

Table 3.22: List of other Research Institutions that submitted products to the VQR on a voluntary basis in alphabetical order.

Table 3.23: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis in term of the normalized average score ($R$) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.24: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.25: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 3.26: Ranking of other Research Institutions that submitted products to the VQR on a voluntary basis by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

The last table of this section, Table 3.32, includes all Institutions previously considered, that is, both Universities and Research Institutions, and reports for each of them the total number of research staff members and the number of active, inactive, and partially active ones. Institutions with fewer than 3 members in the Area are not included. In percentage points, the sum of inactive and partially active members ranges widely between 0 and over 66%, even though the latter figure refers to a very small Institution, in terms of staff members.

Table 3.27: Number of active, not active, and partially active research staff members by Institution.

3.3 Comments on the meaning of the indicators for the Institutions in the Area

It is useful to recall how the indicators employed in the above analysis should be interpreted. The first indicator, $I_{IJ}$, is a measure of the quality of research output and takes value 1 if the Institution has submitted all expected products and all products have been evaluated as Excellent.

The second indicator, $R_{IJ}$, provides information on the position of an Institution with respect to the average for the Area. If its value is above 1, it means that the Institution’s quality is above the average for the Area, if it is below 1, its quality is below average. Once appropriately normalized, it could be employed for a distribution of resources aimed at rewarding quality only, independently of the size of Institutions.
The third indicator, $X_{i,j}$, provides information on how an Institution fares, with respect to the average for the Area, in terms of the fraction of Excellent and Good products. Values above 1 indicate that the Institution has a percentage of Excellent and Good products higher than average.

Finally, the $IRAS_{i,j}$ indicator, defined by the DM and the Call, combines a merely qualitative evaluation with the size of the Institution, and can be employed for a distribution of resources that can be viewed as an adaptation, based on quality, of a merely proportional distribution. In fact, if in all Institutions all products were to obtain the same average evaluation, then the indicator would reflect only the relative number of submitted products and, as a consequence, the relative weight of the Institution within a specific Area.

4 Evaluation of Departments

One of the tasks of the VQR is to provide to Institutions a ranking of Departments (or equivalent substructures) that can be employed as information by the decision-making bodies of the Institutions in order to distribute resources internally.

The university statutes approved following Legge 240 involve different types of Departments. The most common are given by:

- Departments that fully embed smaller pre-existing Departments;
- Departments that include parts of pre-existing Departments, with a strongly diversified structure which hardly fits one (or two) of the VQR Areas.

In both cases, it is necessary to build indicators at the Department level starting from the evaluation of the products associated with the research staff of each Department.

Defining respectively as $n_{i,j,k,EC}$, $n_{i,j,k,EL}$, $n_{i,j,k,D}$, $n_{i,j,k,LIM}$, $n_{i,j,k,MAN}$, $n_{i,j,k,NV}$ the number of Excellent, Good, Fair, Acceptable, Limited, Missing, and Non Eligible products of the $k$-th Department of the $i$-th Institution in the $j$-th scientific Area, the overall evaluation $v_{i,j,k}$ of the $k$-th Department of the $i$-th Institution in the $j$-th scientific Area can be obtained as:

$$v_{i,j,k} = n_{i,j,k,EC} + 0.7 \cdot n_{i,j,k,EL} + 0.4 \cdot n_{i,j,k,D} + 0.1 \cdot n_{i,j,k,A} + 0 \cdot (n_{i,j,k,LIM} + n_{i,j,k,MAN} + n_{i,j,k,NV})$$

In this section, as for Institutions, we introduce three quality indicators of the products submitted by Departments, which are independent of the number of the staff members being
evaluated for the Area within the Departments themselves. Since they do not take into account the size of Departments, they cannot be employed by themselves for the distribution of resources, but they need to be integrated with (or altogether replaced by) the IRD1 indicator, which takes into account both research quality and the size of the Department within the Area. Nevertheless, the first three indicators provide useful information about research quality of a Department within a specific scientific Area.

4.1 Indicators

4.1.1 The first indicator

Setting as \( n_{i,j,k} \) the number of expected products for the VQR for the \( k \)-th Department in the \( i \)-th Institution in the \( j \)-th Area, the first indicator, \( I_{i,j,k} \), defined between 0 and 1, is given by:

\[
I_{i,j,k} = \frac{v_{i,j,k}}{n_{i,j,k}} \quad (9)
\]

It represents the average score of Department \( k \) of Institution \( i \) in Area \( j \).

4.1.2 The second indicator

The second indicator is given by

\[
R_{i,j,k} = \frac{v_{i,j,k}}{n_{i,j,k}} = \frac{I_{i,j,k}}{\frac{\sum_{i=1}^{N_{IST}} v_{i,j}}{N_j}} \quad (10)
\]

where \( V_j \) and \( N_j \) stand for the aggregate evaluation and the total number of expected products in the \( j \)-th Area.

Indicator \( R_{i,j,k} \) represents the ratio between the average score received by the products of the \( k \)-th Department in the \( i \)-th Institution in the \( j \)-th Area and the average score received by all the products in the \( j \)-th Area. It provides a direct measure of relative research quality in a given Area as expressed by a specific Department: values above 1 indicate a research output whose quality is above the average for the Area, values below 1 indicate below average quality.
4.1.3 The third indicator

The third indicator $X_{i,j,k}$ is given by the ratio between the fraction of Excellent and Good products of the Department in the Area and the fraction of Excellent and Good products in the Area. Values of $X_{i,j}$ above 1 indicate that the Department has a percentage of Excellent and Good products higher than the average in the Area. It can be formally defined analogously to (5).

4.1.4 The IRD1_{i,j,k} indicator of the VQR Call

The IRD1_{i,j,k} indicator is defined in the VQR Call as the ratio between the aggregate score achieved by a Department in a given Area and the aggregate evaluation of the Area itself:

\[ IRD1_{i,j,k} = \frac{v_{i,j,k}}{\sum_{i=1}^{N_{IST}} v_{i,j}} \]  \hspace{1cm} (11)

It can be written as the product between an indicator of relative quality for the products submitted by a specific Department in a given Area and an indicator of the size of the Department within the same Area. The quality indicator is given by the ratio between the average score received by the products submitted by the $k$-th Department in the $i$-th Institution in the $j$-th Area and the average score received by all the products in the $j$-th Area, and corresponds to the third indicator $R_{i,j,k}$ defined in (10), while the size of the Department ($P_{i,j,k}$) is simply given by the share of products of the $j$-th Area due by the $k$-th Department in the $i$-th Institution:

\[ IRD1_{i,j,k} = \frac{v_{i,j,k}}{\frac{n_{i,j,k}}{\sum_{i=1}^{N_{IST}} v_{i,j}} \times \frac{n_{i,j,k}}{N_j}} = R_{i,j,k} \times P_{i,j,k} \]  \hspace{1cm} (12)

Thus, the IRD1_{i,j,k} indicator re-defines the weight of a specific Department within a specific Institution in a specific Area, as measured by the share of expected products, on the basis of the relative quality of the products themselves. As such, IRD1 is a useful indicator especially for the allocation of funds across Departments within the same Institution in the same Area, since it takes into account both research quality and the relative weight of the Department.

The rankings of the Departments in the Area have been obtained using the indicators $I_{i,j,k}$ and $R_{i,j,k}$:
4.2 Rankings of Departments

This section presents rankings for Departments, obtained with the same criteria applied to Institutions (where indicator IRD1 replaces IRAS1). It is worth noting that the indicators exclusively refer to research staff members of Area 13. Therefore, if a Department includes members of other Areas, it will also appear in the rankings of the other Areas. Once again, information is reported only for University Departments with at least 5 expected products. Table 4.1 presents, in alphabetical order, a list of 186 University Departments in the Area, with information about the relevant indicators. The table confirms the heterogeneity of the way Departments are organized within each University. Only relative large Universities show separate Departments for the research areas corresponding to subGEVs. In many cases staff members of Area 13 are affiliated with Departments where they represent a minority if compared to other Areas. Even for Departments belonging to the same University, the positions in the ranking exhibit significant variations. For example, the first University with multiple Departments listed in the table is Bari, which includes 4 Departments in the Area, with average scores ranging from 0.18 to 0.54.

Table 4.1: List of University Departments, in alphabetical order, first by University, then by Department.

It is more appropriate to comment in detail on the rankings within the appropriate size class. Tables 4.2-4.4 show the rankings, separately by size class, where size is determined by dividing by 3 the interval of the values of the number of expected products, without considering Departments with fewer than 5. Therefore, since the maximum number of expected products observed in Table 4.1 is 186 (for Milano Cattolica, Facoltà di Economia), for Area 13 Departments are defined as small, medium-size and large when they were expected to submit to the VQR 5-65, 66-125, and 126-186 products, respectively. Lucca - IMT (which is organized in a single Department), one of the 7 Departments of Milano Bocconi (Analisi delle politiche e management pubblico), and one of the 2 Departments of Roma Tor Vergata (Economia e Finanza) rank first in the corresponding small, medium, and large size classes. To be noticed is that the majority of large Departments achieve an above-average score.

Table 4.2: Ranking of small Departments (in terms of the number of expected products of the research staff members of the Department) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.3: Ranking of medium-size Departments (in terms of the number of expected products of the research staff members of the Department) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).
Table 4.4: Ranking of large Departments (in terms of the number of expected products of the research staff members of the Department) on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

The following tables present results at the subGEV level (Tables 4.5-4.8), again by size class. Since size classes are determined according to the rule previously explained for Departments, thresholds necessarily differ across subGEVs.

Table 4.5: List of University Departments in alphabetical order, first by University, then by Department, for all subGEVs of the Area.

Table 4.6: Ranking of small Departments by subGEV of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.7: Ranking of medium-size Departments by subGEV of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.8: Ranking of large Departments by subGEV of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

The next set of tables replicates the above analysis at the SSD (Tables 4.9-4.12) and at the MACRO SC level (Tables 4.13-4.16), always by size class, where thresholds are determined according to the same rule. The analysis of the rankings reveals that some Universities can excel in a single SSD or MACRO SC even when they perform in the average for the Area as a whole.

Table 4.9: List of University Departments in alphabetical order, first by University, then by Department, for all SSDs of the Area.

Table 4.10: Ranking of small Departments by SSD of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.11: Ranking of medium-size Departments by SSD of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.12: Ranking of large Departments by SSD of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.13: List of University Departments in alphabetical order, first by University, then by Department, for all MACRO SCs of the Area.

Table 4.14: Ranking of small Departments by MACRO SC of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).

Table 4.15: Ranking of medium-size Departments by MACRO SC of research staff members and distribution of products by final merit class (Excellent - A; Good - B; Fair - C; Acceptable - D; Limited - E; Not assessable - F).
Table 4.16: Ranking of large Departments by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

The remaining part of this section (Tables 4.17-4.21) analogously reports evaluation results for substructures within Research Institutions. Only CNR includes multiple (6) separate Institutes with more than 7 expected products, and is represented in multiple subGEVs (all 3), SSDs (3), and MACRO SC (3). Results are presented by Area, subGEV, SSD, and MACRO SC, without any differentiation based on size.

Table 4.17: List of the substructures of the controlled and assimilated Research Institutions in alphabetical order.

Table 4.18: Ranking of the substructures of the controlled and assimilated Research Institutions on the basis of the normalized average score (R) and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.19: Ranking of the substructures of the controlled and assimilated Research Institutions by subGEV of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.20: Ranking of the substructures of the controlled and assimilated Research Institutions by SSD of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

Table 4.21: Ranking of the substructures of the controlled and assimilated Research Institutions by MACRO SC of research staff members and distribution of products by final merit class (Excellent -A; Good -B; Fair -C; Acceptable -D; Limited -E; Not assessable -F).

5 Evaluation of research staff members

The rankings so far examined are based on the average score of the research products submitted to the VQR. Therefore, they illustrate how the distribution of the products by merit class affects the performance of Institutions and Departments. However, the information so far presented refers to individual products and not to individual authors. Assume, for example, that a Department including 20 research staff members submits to the VQR 40 products (to simplify, we assume that each member is expected to submit 2 products). Assume, next, that 20 of these products are classified as Excellent. It could be the case that each member of the Department has submitted one of them, which implies a uniform distribution of the Excellent products among all Department members. At the other extreme, it could be the case that all 20 Excellent products are submitted by 10 members, whose individual average score would then be equal to 1, with a corresponding score below 1 for the other 10 members. In other words, the data presented in the
previous sections do not allow an analysis of the distribution of research quality within each Department, or Institution. Therefore, in this section we present data about the average score of each research staff member, computed as the average of the scores of the products submitted by each of them (without counting co-authored products submitted by co-authors). The highest average score each member can achieve, by submitting two products that are both classified as Excellent, is 1, while the lowest is of course 0. The score would be, for instance, equal to 0.5 if a member submits an Excellent and a Limited product (or an Excellent product alone, when a member is partially active). In interpreting results, we shall abstract from any consideration concerning co-authorships, in line with the VQR rules that do not allow Institutions (with the exception of substructures of CNR, INAF, INFN, and INGV) to submit the same product associated with two or more different co-authors.

Table 5.1 reports, separately for Universities and the two types of Research Institutions previously defined, the distribution of the average scores of the research staff members of the Area. Average scores are grouped into 7 classes (0, 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-0.99, and 1). The first panel of the table shows that, out of the 4,554 University staff members, almost 14% receive an average score of 0 while, at the other extreme of the distribution, 15% receive the highest possible score of 1. The fraction of staff members scoring 1 is correlated with the rank of the Institution. For instance, the three leading Universities in terms of the overall rank also show the highest percentages of staff members scoring 1. The fraction scoring 0 varies widely from 0 to 83%. It is worth noting that even high-ranking Universities, such as Milano Bocconi, display a relatively high fraction of staff with 0 average score, which is compensated by a large fraction of members scoring 1. Moreover, members scoring 1 are present in 59 of the 82 Universities shown in the table, which suggests a large geographic dispersion of top-quality researchers. A possible drawback of this fact is that it may prevent the accumulation of critical mass and, given the characteristics of post-graduate training in the Area, hinder the sustainability of effective doctoral programs. The other two panels of the table show analogous data for Research Institutions, while Table 5.2 presents data for substructures, starting with University Departments. Within CNR, there is a large variation across Institutes in the percentage of researchers scoring 1.

*Table 5.1: Ranking of Institutions by average score (VM) of research staff members and distribution by class of the average scores.*

*Table 5.2: Ranking of substructures of Institutions by average score (VM) of research staff members and distribution by class of the average scores.*
In Table 5.3 the average score of the research staff members and its distribution are presented separately for different levels of University position, that is, researcher (roughly equivalent to assistant professor), associate professor, and full professor, and also by subGEV. In addition to differences in the level of the average score across sugGEVs, that confirm previous findings, this table shows that the average score of associate professors is higher than that of full professors and researchers. At the same time, the distribution by class of the average score reveals that the share of members scoring 1 declines with the level of the positions, for all subGEVs. Interestingly, however, differences across subGEVs emerge at the opposite tail of the distribution, with Business showing the largest share of staff scoring 0 among full professors, while for Statistics and Economics this occurs for researchers. This table, as well as the next, only provides data for Universities.

Table 5.4 offers a complementary perspective by associating research quality to age, again by subGEV. Years of birth are aggregated into classes, each including 5 years (except for the first and the last). The average score tends to increase with age for all subGEVs. For Economics we observe a jump in average quality for the generation born between 1946 and 1955. For Business the increase is more gradual and starts only with those born in the 1960s. Statistics shows higher scores for the older generations, if compared to the other two subGEVs, and very high scores also for the younger ones. Inspection of the share of staff members scoring 1 shows a very rapid increase with age for all subGEVs. To reconcile results by position and by age, it is fair to conjecture that the most productive staff members are probably relatively young associate professors and researchers. In particular, the fact that research quality increases with age but at the same time it is highest for associate professors is due to the presence of relatively old researchers who are not productive or are not producing high-quality research. An implication of these findings is that one of the determinants of the overall ranking of Universities is the different composition of the research staff along the age dimension.

Table 5.3: Average score (VM) of research staff members and distribution of the average score by University position (researcher -RU, associate professor -PA, full professor -PO) and by subGEV of research staff members.

Table 5.4: Average score (VM) of research staff members and distribution of the average score by year of birth and by subGEV of research staff members.

6 Summary of results

This section briefly summarizes the main results emerging from the evaluation of Area 13 in the VQR 2011-2014 (for short, VQR2). In terms of products being submitted, out of the over 8,000
products submitted to GEV13, 74% were journal articles, up from 62% in VQR1. The share of articles written in English increased from 57% to 77%. The share of products evaluated with bibliometric analysis increased from 53% to 68%. In all these interrelated dimensions, despite deep differentiations both at the subGEV and the SSD level, and even within SSDs, there is evidence of a generalized evolution of the Area toward increased internationalization and, at the same time, a choice of publication outlets gradually getting closer to that of the “hard” sciences.

Turning the attention to the evaluation of products, average scores were higher for journal articles, for articles in English, and for bibliometric products. Reflecting the higher evaluations of these products, products submitted by research staff members belonging to the Economics and Statistics subGEVs obtained results above the average for the Area (0.46), while the Business subGEV was below the average, although for Business there is evidence of a marked improvement relative to VQR1, with a consequent gradual convergence across subGEVs. However, marked differences still remain across SSDs, even within the same subGEV. For instance, SECS-P/05 (in the Statistics subGEV) obtained the overall best average score, while relatively low evaluations were obtained by SECS-P/04 and SECS-P/12 (in the Economics subGEV) and especially SECS-P/11 and SECS-P/09 (in the Business subGEV). These differentiations largely reflect the choice of publications submitted to the VQR, as well as the well-known specific position of Economic History and History of Economic Thought within the Area.

In order to interpret evaluation results for Institutions, it is instructive to compute, as a term of comparison, the score that an Institution would have achieved had it submitted all the expected products, all of them assessable, and assuming a distribution of its submitted products equal to the international ex ante distribution of products. Accordingly, multiplying the VQR score for each merit class (1, 0.7, 0.4, 0.1, 0) times the corresponding percentiles (10, 20, 20, 30, 20), the average score for this representative Institution would be equal to:

\[ 1\times 0.1 + 0.7\times 0.2 + 0.4\times 0.2 + 0.1\times 0.3 + 0\times 0.2 = 0.35 \]

If we compare the above result with the average score for the Universities in the Area, which is equal to 0.46, even discounting for the fact that the products submitted to the VQR are a selected sample of the entire research output of the Area in a given period, we can cautiously assert that, overall, Area 13 is in relatively good standing if compared with the international context. Once again, however, it is necessary to stress the presence of considerable dispersion around average quality. There is indeed pervasive heterogeneity, as previously remarked, not only across subGEVs and SSDs, but also along the size and the geographic dimensions, with
relatively better achievements for larger Universities and for Universities located in the North and in the Center, despite the presence of several exceptions. Evaluation results concerning University Departments confirm the same variegated picture.

Another conclusion emerging from the analysis is that research in Area 13 is mostly concentrated in Universities, rather than Research Institutions, and that the largest entity among the latter, CNR, is showing an average score lower than that of Universities. Despite the emergence of high-performing smaller entities among Research Institutions, this implies that heterogeneity arises also when we compare the institutional and organizational framework within which research is conducted.

By aggregating evaluation results at the level of individual research staff members, in order to understand how research quality is distributed within Institutions or Departments, we showed that, despite the fact that a large share of members scoring at the top is associated with the overall ranking, top researchers are largely dispersed, and that demographic dynamics are an important determinant of institutional performance in terms of research quality, with younger generations moving toward higher standards.

To conclude, it should be stressed that a direct comparison with the evaluation results of VQR1 needs to be taken with caution, for (at least) the following reasons: (i) the differences in the details of the Call regarding the scores and percentiles associated with the merit classes; (ii) the change in the number of years (from 7 to 4) and the number of products (from a maximum of 3 to a maximum of 2) subject to evaluation; (iii) the refinement in the evaluation criteria introduced by the GEV and, in particular, the expansion of the coverage of the GEV journal list; (iv) the presence of a learning-by-doing effect, active for Institutions and individual authors alike, both in the selection of the products to be submitted for evaluation and in the general awareness of the implications of the process; (v) lastly and perhaps most important, the transformation in the set of research staff members subject to the VQR: namely, because of retirements and new hires, the population under evaluation is no longer the same if compared with VQR1. Only a deeper investigation could reveal the relative weight of the above factors, but it is fair to say that they are not likely to pull into the same direction, so that some may compensate others.

With the above warnings in mind, we can nevertheless conclude that VQR2 shows evidence of an improvement in the performance of Area 13. In particular, using once again as a benchmark the average score that can be computed for a representative Institution, we can observe that, for the Institutions in the Area, the actual score resulting from VQR1 (0.32) was
lower than the benchmark (then, 0.41), while it is higher for VQR2. This relative improvement can be attributed to parallel and interrelated circumstances, which can be summarized as follows. First, for all the disciplines in the Area, there is a clear tendency to higher and more internationalized research standards, as reflected in the kind of products submitted to the VQR, namely, articles in indexed journals that tend to achieve a higher evaluation. Second, a determinant which was already in place during the period covered by VQR1 and is now even more noticeable is the demographic dynamics, which imply a growing role for younger and more productive researchers, thus reinforcing the tendency toward internationalization and higher standards. This factor is likely to play a decisive role in the future, but only if appropriately sustained with adequate recruitment policies. Last but not least, we like to think that the introduction at the national level of a consistent practice of merit-based research evaluation policy has already delivered visible and welcomed results.

7 Final considerations

This final section is devoted to a critical discussion of the experience of GEV13 along the evaluation process, with the aim to provide suggestions which could be useful for future VQR exercises. Most of the points have been raised during the Final GEV Meeting that took place on October 24, 2016.

There was general agreement about the fact that the interface provided by CINECA was not user-friendly, and that the constant delays in its accessibility caused serious disruption. Problems regarding the actual availability of reviewers through the interface were particularly acute, both because it was difficult for GEV members to assign products to the desired reviewers, and because reviewers themselves reported impediments of various sorts with the use of the interface. As a result, GEV members were imposed undesirably tight deadlines in order to complete their task on time.

Even though the GEV met the VQR deadlines for the publication of the evaluation criteria and the journal classification, there was also general agreement on the need to publish earlier, and with more detail, the evaluation material. In particular, the time allotted to the GEV should be such to allow the publication of the journal classification well in advance with respect to the submission of products to the VQR. It would also be useful to provide a more exhaustive description of not assessable products and to develop a separate bibliometric classification for journal articles that belong to minor sub-types (such as review articles, rejoinders, etc.).
On other issues, opinions among GEV members were more varied. Some GEV members suggested that, even though VQR2 already restricted the top merit class if compared to the previous exercise, an even sharper ranking, aimed at ensuring higher recognition to high-quality journals and articles, would be more effective in terms of incentives to improve performance, especially for young researchers at the start of their career. Some GEV members thought that, given the evolution of the Area in the direction of growing internationalization, in the future no special treatment should be granted to Italian journals, while other GEV members instead pointed out the signaling value of rewarding a small number of Italian journals that are investing in the internationalization process. Some also argued that the weight assigned to individual citations in the evaluation of bibliometric products should be increased; in this respect, the analysis of individual citations provided in Appendix D provides interesting cues which might be useful for the future. Moreover, concerning the imputation methodology applied for the journal list, it was noted that a certain “database effect” may be present: since Google Scholar tends to index a wider set of products if compared to WoS and Scopus, the $h$-index values may be overestimated, with a possible effect on imputed values. Further research needs to be done also to verify the potential impact of the number of papers appeared in a journal on either the corresponding $h$-index or the bibliometric indices values.

Several GEV members felt that, especially for foreign reviewers, it was difficult to take into account the merit classes established by the VQR Call and that more detailed guidelines should have been provided. Concerning the internal organization of the GEV, it was pointed out that the size of each subGEV was not strictly proportional to the share of products to be evaluated in peer review; since this share is not known in advance, a solution would be to take advantage of the possibility to recruit additional GEV members along the evaluation process, if needed.

Some GEV members suggested that ANVUR should try to coordinate the journal list prepared by the GEV with the one compiled for Abilitazioni Scientifiche Nazionali (ASN), even though the GEV is aware that the VQR and ASN pursue separate goals and operate under different rules established by the University Ministry. The issue of the potential presence of predatory journals in the GEV journal list, and more generally also in indexed databases, was addressed as well.

Within each subGEV, other more specific issues were also discussed.

For Business, problems were reported regarding the evaluation of products in commodity science, because of their highly technical content, despite the availability of reviewers selected by the other Areas. The sharp differences emerged in the evaluations across SSDs, with a poor performance in particular for Financial markets and institutions and Corporate finance, raised
some questions regarding the comparability of results across SSDs within Area 13 and even the subGEV. To some extent, however, the fact that the final report publishes rankings at the SSD level should respond to these concerns.

For Economics, the peculiarity of the Economic History products was highlighted: disagreements between reviewers were more common for this SSD and, given the large share of peer-reviewed products, this might have contributed to its weaker performance, despite evidence of an improvement relative to VQR1.

For Statistics, some GEV members remarked that Mathematics journals tend to be penalized by being in the same sub-list with Statistics journals, since they generally have lower values of the bibliometric indicators. However, many other GEV members were against restricting sub-lists and/or subGEVs to an individual SSD, since authors can and do publish in any journal in the list independently on how sub-lists are organized, reflecting the fact that research topics are not SSD-specific. In fact, different citations standards also apply to different sub-fields within Statistics, with applied and computational statistics journals showing higher values of the indicators if compared to methodological journals. However, while differences in citation standards across sub-fields are undeniable, the fact that the VQR aims at evaluating Institutions rather than individuals should attenuate this problem. This implies that Institutions, as remarked by several members from all subGEVs, should refrain from employing VQR results and in particular the GEV journal list for inappropriate purposes.