# Italian Research Assessment VQR: Framework, Achievements and Controversies

XIAOWU, WANG*
Education Research Institute, China University of Geosciences, Wuhan

## ABSTRACT

Research assessment is an important basis for university funding and a solid pillar of the national innovation system. The Italian research assessment VQR has been reformed rapidly with outstanding achievements and representative controversies, thus showing its significant research value. In this paper, on the basis of studying the VQR's framework, we summarize its achievements and controversies. VQR has stimulated university competition, assisted higher education accountability, and improved research management mechanisms. However, it has also encountered controversies in methods, indicators, algorithms, and submission procedures, etc. Based on the detailed research and discussion on VQR, this paper can provide experience for the evolution and development trend of world research assessment.

## CCS CONCEPTS

• **Applied computing**; • **Education**;

## KEYWORDS

VQR, Italy, Research assessment

**ACM Reference Format:**
XIAOWU, WANG. 2022. Italian Research Assessment VQR: Framework, Achievements and Controversies. In *2022 the 7th International Conference on Distance Education and Learning (ICDEL) (ICDEL2022), May 20–23, 2022, Beijing, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3543321.3543364

## 1 INTRODUCTION

Research assessment is directly related to the development direction of national scientific research, thus becoming a research hotspot in academia. Previous research mainly focused on its system composition and methodological failures, accordingly paid less attention to the achievements and controversies of research assessment. Except for the United Kingdom, Italy is the only country in Europe that requires a frequent large-scale national research assessment exercise. Italian research assessment has undergone continuous reform and has established a regular national research assessment exercise, named Valutazione delle Qualità della Ricerca (VQR), with a high research value.

VQR is the successor of Valutazione Triennale della Ricerca (VTR). VQR was launched by Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR). According to the Law 286 (Legge 24 novembre 2006, n.286), ANVUR aimed to evaluate the research quality of Italian universities and research institutions [1]. During the implementation, VQR was established as a new national research assessment exercise in 2011. So far, VQR1 and VQR2 have been completed, and the VQR3 is in progress.

In general, after decades of reform, VQR has achieved remarkable achievements in universities' research quality, government accountability, and scientific research management. Simultaneously, it has also caused controversies on assessment methods, assessment indicators, assessment algorithms, etc. To this end, this comprehensive research of Italian research assessment's framework, achievements, and controversies can enrich research in this field, summarize the experience, and provide new perspectives for other countries in the world. The rest of this paper is organized as follows. Section 2 describes the whole framework of VQR. Section 3 summarizes VQR's achievements. The most prominent controversies around VQR are presented in Section 4. Section 5 concludes the paper.

## 2 VQR'S FRAMEWORK

### 2.1 Assessment subjects

VQR is evaluated by the GEV (Gruppo Esperti della Valutazione), a group of national and international experts nominated by ANVUR. The distribution of the VQR expert group is shown in the Table 1[1].

### 2.2 Assessment objects

ANVUR states that the results of the VQR cannot and should not be utilized to evaluate individual researchers. As a result, the assessment is intended for: 1) state universities (Università statali); 2) legally recognized non-state universities (Università non statali legalmente riconosciute); 3) research institutions administrated by the Italian Ministry of Education, Universities, and 4) Other public and commercial entities doing research upon specific request and agreement to incur associated costs. VQR accepts products of journal articles, book chapters, conference proceedings, critical reviews, commentaries, book translations, patents, samples, project plans, software, databases, exhibits, artworks, compositions, and monographs, etc.

*Xiaowu Wang received the B. Ed. degree from the Faculty of Education, Hubei University, in 2019. She is currently pursuing the master's degree in pedagogy with Education Research Institute, China University of Geosciences, Wuhan. Her major research interest are research assessment and foreign education.


*ICDEL2022, May 20–23, 2022, Beijing, China*

ACM ISBN 978-1-4503-9641-7/22/05…$15.00
https://doi.org/10.1145/3543321.3543364

[1]The table 1 contains information from files on ANVUR's official website: https://www.anvur.it

**Table 1: GEV Areas and Numbers**

| VQR1 | | VQR2 | | VQR3 | |
|---|---|---|---|---|---|
| GEV Areas | No. | GEV Areas | No. | GEV Areas | No. |
| 1 Mathematics and Computer Sciences | 25 | | 22 | | 29 |
| 2 Physics | 18 | | 33 | | 43 |
| 3 Chemistry | 23 | | 22 | | 31 |
| 4 Earth Sciences | 9 | | 15 | | 19 |
| 5 Biology | 38 | | 33 | | 49 |
| 6 Medicine | 79 | | 58 | | 80 |
| 7 Agricultural and Veterinary Sciences | 24 | | 20 | | 36 |
| 8 Civil Engineering and Architecture | 28 | 8a Architecture | 14 | | 17 |
| | | 8b Civil Engineering | 9 | | 15 |
| 9 Industrial and Information Engineering | 40 | | 33 | | 59 |
| 10 Ancient History, Philology, Literature and Art History | 42 | | 36 | | 56 |
| 11 History, Philosophy, Pedagogy and Psychology | 38 | 11a History, Philosophy, Pedagogy | 25 | | 29 |
| | | 11b Psychology | 6 | | 12 |
| 12 Law | 37 | | 32 | | 39 |
| 13 Economics and Statistics | 36 | | 31 | 13a Economics and Statistics | 22 |
| | | | | 13b Economics and Management | 18 |
| 14 Political and Social Sciences | 13 | | 11 | | 16 |
| Total | 450 | | 400 | | 570 |

**Table 2: Research-related Indicators and Weights**

| Period | Research-related Indicators | Weights | Description |
|---|---|---|---|
| VQR1 | Research Quality | 0.5 | Quality of participating research results |
| | Attract Resources | 0.1 | Successful in winning external competitive funding in this field |
| | Staff Mobility | 0.1 | Outcomes for those promoted during the assessment period |
| | Internationalization | 0.1 | Over three months of visits and visitors; including outstanding outcomes from overseas researchers |
| | Higher Education | 0.1 | Number of active researchers in the field (PhD students, postdoctoral fellows, and researchers) |
| | Self-owned resources | 0.05 | Own financial resources in this field |
| | Improvement | 0.05 | Progress in research quality compared to VTR |
| VQR2 | Research Quality | 0.75 | Quality of participating research results |
| | Recruitment | 0.2 | Results submitted by researchers recruited by the institution or joining at a higher level during the assessment period |
| | Attract Resources | 0.1 | Successful in winning external competitive funding in this field |
| | Higher Education | 0.1 | Number of active researchers in the field (PhD students, postdoctoral fellows, and researchers) |
| | Improvement | 0.3 | Progress in research quality compared to VQR1 |

## 2.3 Assessment indicators

VQR includes research-related and third mission-related indicators. Table 2 shows the research-related indicators and their weights[2]. In the early period, VQR1 concurrently evaluated the third mission and the research quality. VQR2 separated the third mission into a distinct area, which was evaluated by ANVUR. Furthermore, VQR3 established an interdisciplinary GEV to evaluate the third mission independently.

## 2.4 Assessment process and methodology

*2.4.1 Assessment process.* VQR1 assessment procedure involves three actors: ANVUR, GEV, and the participating universities (often called STRUCTURE). In the VQR2, four actors were involved: the

[2]The table 2 also contains information from files on ANVUR's official website: https://www.anvur.it

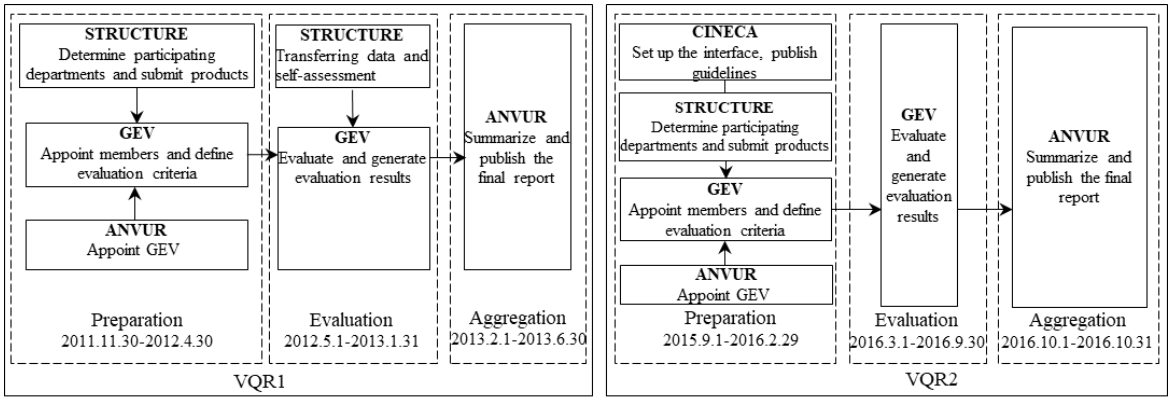


**Figure 1: VQR1 and VQR2 Assessment Process**

Inter-University Consortium for Computing (CINECA), ANVUR, GEV, and STRUCTURE. The process for VQR1 and VQR2 and the period are concluded in Figure 1.

*2.4.2 Assessment methodology.* VQR combined bibliometric and peer-review for evaluating the quality of participating research products. This allowed each GEV to decide the assessment methods. In general, GEVs in STEM (Science, Technology, Engineering, Mathematics) area prefer the bibliometric methods, whereas GEVs in HSS(Humanities and Social Sciences) area is more inclined to the peer review. However, some specific areas, such as Group IV Earth Science GEVs, integrate bibliometric and peer-review.

## 2.5 Use of assessment results

The results of VQR are used in the following ways: (1) It serves as a guide for allocating public funds to institutions. Italian universities are funded mainly by the Ordinary Financing Fund (Fondo di Finanziamento Ordinario, FFO). (2) It generates a national ranking of Italian universities and research institutions. It disseminates information to the outside world and influences universities' and research institutions' reputations. (3) It provides data for the Italian government to allocate additional funding, such as the University Department of Excellence (Dipartimenti Universitari di Eccellenza) funding.

## 2.6 Monitoring and correction

Due to the advancement of technology and requirement, VQR has always been revised and modified under the watchful eye of numerous stakeholders. In general, the monitoring parties comprise the government, companies, parents, students, and other stakeholders in higher education, and the revision focuses on the assessment indicators and algorithms. The main revisions come from the following two parts:

1. ***Assessment indicators pay more attention to socio-economic impact***. VQR's indicators have been adjusted to emphasize the third mission and the economic value of research. Especially in VQR3, great importance is attached to the impact of universities on society, economy, and culture [2].
2. ***Assessment algorithms consider marginal products***. In the algorithm of VQR1, a small change in an indicator will cause a huge change in the final result, which is unfair to the marginal products. In order to address this problem, VQR2 updated the algorithms by inserting a slope.

# 3 ACHIVEMENTS OF VQR

## 3.1 Stimulate competition to improve the quality of research in Italian universities

The purpose of assessment is to improve rather than to prove. Hence, the purpose of research assessment is to identify the best universities and provide selective funding. Thus, it could effectively improve the research quality and productivity and ultimately promote the development of national science and technology.

Firstly, VQR stimulates competition for scarce resources. Cattaneo et al. have shown that implementing a performance-based research financing scheme increases university-level research productivity [3]. Until 2009, Italian government funding to universities was input-oriented, thereby covering the demands of all universities equally. At the time, competition amongst Italian institutions was not very fierce. In contrast, the VQR allocates performance-based funding, which means that institutions that excel at research can earn a greater share of the funding. Thus, it has a positive incentive to raise the quality of research at Italian universities.

Besides, VQR determines a large amount of university funding. In the previous, VTR can only decide less than 2% of FFO. In contrast, VQR directly affects FFO, and the amount of FFO according to the results of VQR basically shows an upward trend: 7% in 2009, 20% in 2015, and 23% in 2016 [4]. Ancarani et al. pointed out that VQR can effectively influence funding and promote the research quality. Especially in some fields (earth sciences, biological and medical sciences) that need to be evaluated by bibliometric, the impact of funding supply on the quality of research is significant [5]. In this context, the implementation of VQR has a positive incentive for research in Italian universities, encouraging them to be active in research activities. Meanwhile, the more active universities are in research, the more research resources (e. g. grants, etc.) they attract. More research resources in turn will support the development of research more effectively, creating a virtuous circle.

### 3.2 Assist accountability to promote the reform of Italian scientific research governance

VQR builds a bridge between public funding and university performance, thus playing a pivotal role in accountability. The goal of accountability in higher education is to ensure that government departments, social institutions, and individuals monitor higher education institutions' use of resources. Under the supervision and assistance of higher education accountability, the way funds are used and the efficiency of research activities gradually become transparent.

1. ***Italian universities need to reorganize order***. Prior to the VQR, Italian universities had a period of chaotic governance. After the Bologna Declaration, the Italian university system underwent major reforms. However, the government exercised only a limited level of control throughout the reform. As a result, a national research assessment was implemented to regulate and constrain the research activities of Italian universities.
2. ***VQR assists accountability***. In 2006, ANVUR established VQR to provide an open, transparent, nationwide ranking of Italian universities. Stakeholders in higher education rely on the VQR-published ranking tables to hold universities accountable. Universities that do low-quality research and perform poorly in the VQR face some penalties (e.g., receiving less public funding than other high-performing universities as well as damage to their reputation and reduced student enrollment). Thus, VQR has developed into a governmental tool to assist accountability.

### 3.3 Improve mechanisms to improve the management of Italian scientific research

In addition to stimulating competition and assisting accountability, the VQR has improved the mechanisms for managing research at Italian universities. The advancement of research management at universities is primarily reflected in the steady establishment of research standards and the growth of external communication channels.

1. ***VQR helps to establish a set of research standards.*** Turri asserts that academic critiria are rapidly becoming more standardized in Italy [6]. In 2007, the Italian Academy of Business Administration and Management (AIDEA) released the first standard Italian journal ratings [7]. Additionally, VQR is contributing to the establishment of uniform research standards. The standardization of CRIS (Current Research Information Systems) is a byproduct of VQR. In VTR, the majority of universities lack a CRIS; however, in VQR1, over half of universities have a separate CRIS, and in VQR2, universities are required to utilize the CINECA-developed CRIS [4].
2. ***VQR expands external communication channels.*** The ranking table provided by VQR broadens external communication channels. VQR was quoted in 1406 articles in the two months following the publication of the results (July 17–September 20, 2013) [8]. Additionally, VQR's results help the screening of the University Department of Excellence. Italy enacted Law 232 (Legge 11 dicembre 2016, n. 232) in 2016, which established funding for universities' departments that are excellent in research [9]. The results of VQR2 serve as a reference for selecting these outstanding departments.

## 4 CONTROVERSY AROUND VQR

Within a few years, VQR had a significant impact on university funding, reputation, and internal resource distribution. Due to VQR's significance, its implementation has sparked lots of controversies. This section discusses the most prominent controversies such as assessment methods, assessment indicators, assessment algorithms, submission products, etc.

### 4.1 Assessment methods

Numerous conferences and related materials on research assessment advise that assessment methodologies should objectively reflect the distinctions between scientific domains. According to the San Francisco Declaration on Research Assessment, bibliometric indices such as the two-eight rule, which results in 20% of publications receiving 80% of citations, do not accurately reflect the genuine quality of research products. Rather than depending excessively on bibliometrics, the scientific community should develop more robust tools for assessing research [10]. The Leiden Manifesto (Leiden Manifesto) proposes ten principles for assessment, including the following: bibliometrics cannot replace peer review; quantitative assessment should complement qualitative expert assessment, and so on [11].

Firstly, many researchers such as Moed [12], Hicks [13], Archambault [14], and Nederhof [15] note that bibliometrics cannot be used in the HSS. Numerous researchers have evaluated the scientific validity of VQR assessment methods and found that the hybrid methods now utilized by VQR are feasible. Bertocchi et al. observed that the results of bibliometrics and peer review are mostly consistent, indicating that the two methodologies can be merged [16]. Traag et al discovered that the degree of agreement between bibliometrics and peer review was actually greater than the degree of agreement between two independent reviewers [17].

However, others have questioned the hybrid assessment method. Geuna notes that even when the bibliometric technique is used in conjunction with peer review, the disadvantages of peer review, such as its high cost and time commitment, remain [18]. Abramo argues that in hard science, the findings of peer review and bibliometric assessment approaches are significantly different [19]. Baccini & De Nicolao [20], Ancaiani [5], and others observed that peer review and bibliometric scores in the same subject varied. In this situation, integrating peer review and bibliometrics for assessment could lead to distorted results.

### 4.2 Assessment Indicators

Many researchers have conducted critical analyses of the journal impact factor and citation metrics used in VQR's bibliometric methodology.

Firstly, for articles with a short publication time, the journal impact factor can assist in the evaluation. Anfossi et al. contend that it is inadvisable to employ a single citation metric when evaluating "young" works [21]. Bonaccorsi notes that for recently works, the journal impact factor indicator should be used to supplement the

evaluation process [22]. Second, the use of journal impact factor measures can aid in the establishment of academic culture. According to Ancaiani et al., the use of journal indicators can communicate to young researchers that researchers must benchmark themselves against the most rigorous peer review process [5].

However, the journal impact factor used in research assessment only takes the mean of the citation distribution into account, which may obscure the genuine worth of the research products. Abramo et al. note that there is no empirical or theoretical evidence to support the scientific validity of the VQR's indicators during their development, publication, and adoption [23]. Additionally, they contended that the VQR's method of evaluating research products in a global rather than a national level would be unfair to those conducting catch-up research in areas of national strategic relevance [24]. Second, there is a lack of transparency about the journal impact factor. Mingers pointed out that the journal impact factors calculation method lacks transparency [25]. Similarly, over time, the correlation between the journal impact factor and the publication's average citation impact will weaken. Lozano et al. discovered a decreasing trend in the strength of the relationship between impact factors and citation counts since 1990 [26]. As a result, it is unlikely that the indicators utilized in VQR's bibliometric approach are sufficient to assess the quality of research outputs efficiently.

## 4.3 Assessment algorithms

When various indicators are employed for assessment, the algorithms used to aggregate the findings are critical. If the aggregation procedure is not scientific, it will not correctly and scientifically reflect the assessed subject's genuine scientific productivity.

Numerous researchers have conducted research on the VQR algorithm and found that it is scientifically valid. Checchi et al. used VQR's algorithm to evaluate papers submitted to the REF (Research Excellence Framework) in 2014 and discovered a correlation of greater than 0.80 with the country ranking obtained through peer review, indicating that VQR's algorithm can be backed up by other national research assessment [27]. Benedetto et al. investigated the VQR algorithm's mistake and they concluded that the impact could be as low as 0.05 percent of articles submitted by Italian authors [28].

Many experts have cast doubt on the VQR assessment process. Baccini et al. investigated each assessment domain in the VQR and found that the results of domain 13 (economics and statistics) are distorted to some extent by the VQR's evaluation algorithm [20]. Franceschini questioned VQR2's slope method, saying that any nonlinear adjustment of the two indicators distorts the underlying indicators' statistical features [29].

## 4.4 Submit procedure

Many researchers support the VQR submit procedure, citing the great difficulty and expense of reviewing all research results from all researchers. If all of the research products were to be peer-reviewed, the Italian government would face a huge expenditure of personnel and financial resources. Australia, the only country that submits all products for evaluation, has chosen to focus its efforts only on STEM fields. Bonaccorsi says that while the plan to submit all products for evaluation is a worthwhile idea in principle, it faces significant difficulties before becoming a reality [22].

Many academics note that VQR's process of selecting products is not always efficient. Abramo et al. note that the maximum feasible score degraded by 23 to 32% when compared to the score of the efficient choice [30]. Franceschini et al. observed that because each researcher submits a very small number of products, the VQR is rarely an accurate picture of the researcher's genuine productivity [29]. Two high-quality publications within four years does not seem to be a sufficient condition for proving the excellence of an average researcher, i.e. One swallow does not make a summer.

## 4.5 Journal ratings

VQR requires support from other work, such as the rating of A-level journals. Numerous researchers support expert-based journal ratings. Giménez-Toledo et al. note that expert-based journal ratings should be a factor in journal rating. This way, official or informal indicators of quality can be determined [31]. Hicks and Wang claim that journal ratings based on expert opinion can help broaden researchers' horizons and promote the movement of information and knowledge within academia [32].

However, Willmott [33], Alvesson [34], Mingers&Willmott [35], Mingers&Yang [36], and others contend that journal ratings contribute to conformism by stifling original and unorthodox research. Simultaneously, it has the potential to encourage monodisciplinary, which would be detrimental to the advancement of interdisciplinary study. Rafols et al. note that journal ratings discourage multidisciplinary research and may have a detrimental effect on the evaluation performance and financial resources of interdisciplinary research organizations [37].

## 4.6 Impact of assessment results

*4.6.1 Does it exacerbate geographic disparities?* Italy has a notable north-south divide in its higher education system: the southern part is economically less developed and has fewer institutions, whilst the northern part is more developed and has more universities. And the majority of universities in southern Italy rank towards the bottom of VQR's ranking tables.

The VQR, a performance-based research funding system, makes it more difficult for southern universities to compete and win against northern universities. This may aggravate the geographical divide between northern and southern Italy, so reducing their ability for research. Grisorio et al. compared the performance of universities in northern, central, and southern Italy, concluding that there is cross-geographical heterogeneity in the research performance of Italian universities with the north-south disparities being particularly pronounced [38].

However, Checchi et al argue that the VQR exacerbates geographical disparities in Italy is not experimentally supported. They assessed and contrasted the research quality of universities after VQR, observing a tendency toward the mean between universities in southern Italy and those in northern and central Italy [39]. The North-South divide in terms of research quality between Italian universities is gradually narrowing. This means that VQR has mitigated some of Italy's geographical disparities and altered inequities.

*4.6.2 Does it have an effect on academic staff's drive to conduct research?* Mingers and Leydesdorff suggest that it is critical to examine the impact on researchers' behavior when evaluating research using a bibliometric method [25]. Dahler-Larsen contends that performance indicators can have unforeseen consequences in public management [40]. Burrows contends that metrics can operate as a "baton," motivating some actions and discouraging others, so affecting researcher behavior [41]. Due to dwindling research funding, an increasing number of young scientists are relocating abroad. Italy currently spends barely 1.25 percent of GDP on research. 69 Italian researchers urged the European Union to exert pressure on the Italian government to increase research funding to "subsistence levels" [42].

## 5 CONCLUSION

The research summarizes the framework and achievements of VQR. Then it conducts a comprehensive analysis of the controversies over VQR among its researchers from many countries and areas. As can be seen, VQR has consistently enhanced and promoted its efficacy since its inception. It listens to the voices of disputes from all sides and continues to improve its algorithms, etc. in order to seek a more scientific evaluation.

While research evaluation has been a topic of debate in academic and policy circles for many years, undertaking research assessment activities requires a particular level of experience and ability. Italy, it has been noted, lacks a genuine evaluation culture, effective and efficient information systems, and reliable and complete data. There is no tradition of rigorous scientific research evaluation, and there are insufficient data management and analytic institutions to address these concerns [43]. Italy has progressively built a regular national system of research assessment, VQR, despite the absence of an assessment culture, through many changes. It takes time for Italy to create an evaluation culture and competence. The numerous achievements and controversies surrounding VQR may serve as a useful experience for new ideas in research assessment globally.

## REFERENCES


[1] Normattiva. LEGGE 24 novembre 2006, n.286[EB/OL]. (2020-9-25)[2021-3-15]. Available: https://www.normattiva.it/atto/caricaDettaglioAtto?atto.dataPubblicazioneGazzetta=2006-11-28&atto.codiceRedazionale=006G0307&atto.articolo.numero=0&atto.articolo.sottoArticolo=1&atto.articolo.sottoArticolo1=10&qId=&tabID=0.3706253273664925 6&title=lbl.dettaglioAtto

[2] ANVUR.BANDO. VQR 2015-2019[EB/OL].(2020-9-25)[2021-3-15]. Available: https://www.anvur.it/wp-content/uploads/2020/09/Bando-VQR-2015-19_25-settembre_2020_versione-accessibile.pdf

[3] M. Cattaneo, M. Meoli, and A. Signori. 2016. Performance-based funding and university research productivity: The moderating effect of university legitimacy. The Journal of Technology Transfer, vol. 41, no. 1, pp. 85-104.

[4] S. Biesenbender. 2019. The governance and standardisation of research information in different science systems: A comparative analysis of Germany and Italy. Higher Education Quarterly, vol. 73, no. 1, pp. 116-127.

[5] A. Ancaiani *et al.* 2015. Evaluating scientific research in Italy: The 2004–10 research evaluation exercise. Research Evaluation, vol. 24, no. 3, pp. 242-255.

[6] M. Turri. 2016. The difficult transition of the Italian university system: growth, underfunding and reforms. Journal of Further and Higher Education, vol. 40, no. 1, pp. 83-106.

[7] D. Gomes, E. Giovannoni, F. Gutiérrez-Hidalgo, and H. Zimnovitch. 2015. Moving from regional to international publishing in accounting history: Pressures, issues, strategies and implications. Accounting History, vol. 20, no. 2, pp. 183-205.

[8] B. Blasi, S. Romagnosi, and A. Bonaccorsi. 2017. Playing the ranking game: Media coverage of the evaluation of the quality of research in Italy. Higher Education, vol. 73, no. 5, pp. 741-757.

[9] G. Ufficiale. LEGGE 11 dicembre 2016,n.232 [EB/OL].(2016-12-21)[2021-2-4]. Available: https://www.gazzettaufficiale.it/eli/id/2016/12/21/16G00242/sg#

[10] M. Bladek. 2014. DORA: San Francisco declaration on research assessment (May 2013). College Research Libraries News, vol. 75, no. 4, pp. 191-196.

[11] D. Hicks, P. Wouters, L. Waltman, S. De Rijcke, and I. Rafols. 2015. Bibliometrics: the Leiden Manifesto for research metrics. Nature News, vol. 520, no. 7548, p. 429.

[12] H. F. Moed, M. Luwel, and A. J. Nederhof. 2002. Towards research performance in the humanities.

[13] D. Hicks. 2004. The four literatures of social science. Handbook of quantitative science and technology research, pp. 473-496.

[14] É. Archambault, É. Vignola-Gagné, G. Côté, V. Larivi? re, and Y. Gingrasb. 2006. Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. Scientometrics, vol. 68, no. 3, pp. 329-342.

[15] A. J. Nederhof. 2006. Bibliometric monitoring of research performance in the social sciences and the humanities: A review. Scientometrics, vol. 66, no. 1, pp. 81-100.

[16] G. Bertocchi, A. Gambardella, T. Jappelli, C. A. Nappi, and F. J. R. P. Peracchi. 2015. Bibliometric evaluation vs. informed peer review: Evidence from Italy. vol. 44, no. 2, pp. 451-466.

[17] V. A. Traag, M. Malgarini, and S. Sarlo. 2020. Metrics and peer review agreement at the institutional level. arXiv preprint arXiv:.14830.

[18] A. Geuna and M. Piolatto. 2016. Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). Research Policy, vol. 45, no. 1, pp. 260-271.

[19] G. Abramo, C. A. D'Angelo, and F. Di Costa. 2011. National research assessment exercises: a comparison of peer review and bibliometrics rankings. Scientometrics, vol. 89, no. 3, pp. 929-941.

[20] A. Baccini and G. De Nicolao. 2016. Do they agree? Bibliometric evaluation versus informed peer review in the Italian research assessment exercise. Scientometrics, vol. 108, no. 3, pp. 1651-1671.

[21] A. Anfossi, A. Ciolfi, F. Costa, G. Parisi, and S. Benedetto. 2016. Large-scale assessment of research outputs through a weighted combination of bibliometric indicators. Scientometrics, vol. 107, no. 2, pp. 671-683.

[22] A. Bonaccorsi. 2020. Two decades of research assessment in Italy. Addressing the criticisms. Scholarly Assessment Reports, vol. 2, no. 1.

[23] G. Abramo and C. A. D'Angelo. 2016. Refrain from adopting the combination of citation and journal metrics to grade publications, as used in the Italian national research assessment exercise (VQR 2011–2014). Scientometrics, vol. 109, no. 3, pp. 2053-2065.

[24] G. Abramo and C. A. D'Angelo. 2015. The VQR, I taly's second national research assessment: Methodological failures and ranking distortions. Journal of the Association for Information Science Technology, vol. 66, no. 11, pp. 2202-2214.

[25] J. Mingers and L. Leydesdorff. 2015. A review of theory and practice in scientometrics. European journal of operational research, vol. 246, no. 1, pp. 1-19.

[26] G. A. Lozano, V. Larivière, and Y. Gingras. 2012. The weakening relationship between the impact factor and papers' citations in the digital age. Journal of the American Society for Information Science Technology, vol. 63, no. 11, pp. 2140-2145.

[27] A. Ciolfi, D. Checchi, G. De Fraja, I. Mazzotta, and S. Verzillo. 2019. Have You Read this? An Empirical Comparison of the British REF Peer Review and the Italian VQR Bibliometric Algorithm.

[28] S. Benedetto, D. Checchi, A. Graziosi, and M. Malgarini. 2017. Comments on the paper "Critical remarks on the Italian assessment exercise", Journal of Informetrics, 11 (2017) and pp. 337–357. Journal of Informetrics, vol. 2, no. 11, pp. 622-624.

[29] F. Franceschini and D. Maisano. 2017. Critical remarks on the Italian research assessment exercise VQR 2011–2014. Journal of informetrics, vol. 11, no. 2, pp. 337-357.

[30] G. Abramo, C. A. D'Angelo, and F. Di Costa. 2014. Inefficiency in selecting products for submission to national research assessment exercises. Scientometrics, vol. 98, no. 3, pp. 2069-2086.

[31] E. Giménez-Toledo, J. Mañana-Rodríguez, and E. Delgado-López-Cózar. 2013. Quality indicators for scientific journals based on experts opinion. arXiv preprint arXiv:1307.1271, 2013.

[32] D. Hicks and J. Wang. 2011. Coverage and overlap of the new social sciences and humanities journal lists. Journal of the American Society for Information Science Technology, vol. 62, no. 2, pp. 284-294.

[33] H. Willmott. 2011. Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. Organization, vol. 18, no. 4, pp. 429-442.

[34] M. Alvesson and J. Sandberg. 2013. Has management studies lost its way? Ideas for more imaginative and innovative research. Journal of management studies, vol. 50, no. 1, pp. 128-152.

[35] J. Mingers and H. Willmott. 2013. Taylorizing business school research: On the 'one best way'performative effects of journal ranking lists. Human Relations, vol. 66, no. 8, pp. 1051-1073.

[36] J. Mingers and L. Yang. 2017. Evaluating journal quality: A review of journal citation indicators and ranking in business and management. European Journal of Operational Research, vol. 257, no. 1, pp. 323-337.

[37] I. Rafols, L. Leydesdorff, A. O'Hare, P. Nightingale, and A. Stirling. 2012. How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. Research policy, vol. 41, no. 7, pp. 1262-1282.
[38] M. J. Grisorio and F. Prota. 2020. Italy's national research assessment: some unpleasant effects. Studies in Higher Education, vol. 45, no. 4, pp. 736-754.
[39] D. Checchi, I. Mazzotta, S. Momigliano, and F. Olivanti. 2020. Convergence or polarisation? The impact of research assessment exercises in the Italian case. Scientometrics, vol. 124, pp. 1439-1455.
[40] P. Dahler-Larsen. 2014. Constitutive effects of performance indicators: Getting beyond unintended consequences. Public Management Review, vol. 16, no. 7, pp. 969-986.
[41] R. Burrows. 2012. Living with the h-index? Metric assemblages in the contemporary academy. The sociological review, vol. 60, no. 2, pp. 355-372.
[42] E. Cartlidge. 2016. Italian science to be led by optics pioneer. Physics World, vol. 29, no. 4, p. 13.
[43] P. Galimberti. 2012. Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia. JLIS. it, vol. 3, no. 1, pp. 1-25.