



ELSEVIER

Contents lists available at ScienceDirect

Critical Perspectives on Accounting

journal homepage: www.elsevier.com/locate/cpa

A fatally efficient machine. Insights into the ‘banality’ of the research evaluation exercise in Italy[☆]

Rosanna Spanò^{a,1}, Enrico Bracci^{b,2}, Francesca Manes-Rossi^{a,3,*}, Vincenzo Sforza^{c,4}^a Department of Economics, Management, Institutions, University of Napoli, Federico II, Via Cinthia, Monte Sant'Angelo – 80126 Napoli, Italy^b Department of Economics & Management, University of Ferrara, Via Voltapaletto, 11 – Ferrara, Italy^c Department of Economy, Engineering, Society and Business, University of Tuscia, Via Santa Maria in Gradi, 4 – Viterbo, 01100, Italy

ARTICLE INFO

Keywords:

Performance measurement
 Research evaluation
 Banality of evil

ABSTRACT

This study is framed in the debate concerning the measurement of academic performance, and particularly the strand of studies that explores the risks associated with the metrification of research. The objective, guided by the conceptual framework of the banality of evil (Arendt, 1964), is to delve into how research evaluation can shape banal and unoriginal evaluative practices. These practices, in turn, can trigger a fatally efficient machine within the academic system and institutions, and among researchers.

The paper focuses on examining the recently concluded Research Quality Assessment 2015–2019 (VQR3) exercise in Italy, using an autoethnographic approach. The results highlight the risks stemming from the growing dependence of research quality assessment on automatism, which can cause its commodification at the cost of intellectual innovation and, eventually, force actors to conform to the *rules of the game*.

This work contributes to the ongoing academic debate by offering an innovative and multilevel (i.e. macro, meso and micro) theoretical perspective. Not only does this perspective conceptualise and present the dynamics, processes, instruments and actors at play in the phenomena under scrutiny but also provides a deeper understanding of the dynamics that promote the widespread application of research evaluation systems, despite their well-known weaknesses and potentially undesirable practical and ethical effects.

1. Introduction

The debate on performance measurement and evaluation in universities is composite and has highlighted interesting dynamics resulting from the shift from an academic to a market-driven logic (Martin-Sardesai et al., 2020; Parker et al., 2023; ter Bogt & Scapens, 2012). Numerous contributions have underscored the alarming dangers of the *metrification* of academic performance (Maran et al., 2023; Michelon, 2021), which can have counterproductive effects on researcher motivation and values (e.g., Kallio et al., 2017; ter

[☆] An Italian version of this article is provided as online supplementary material.

* Corresponding author.

E-mail address: manes@unina.it (F. Manes-Rossi).

¹ 0000-0003-0557-3011.

² 0000-0001-5048-9260.

³ 0000-0001-9617-4379.

⁴ 0000-0001-9213-1400.

<https://doi.org/10.1016/j.cpa.2024.102742>

Received 20 December 2022; Received in revised form 20 May 2024; Accepted 22 May 2024

1045-2354/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Bogt & Scapens, 2012). This is cause for concern, as it can lead to deep contradictions and splits between and within institutions and the actors involved (Gebreiter, 2022). Moreover, recent studies have highlighted how the increasing use of innovative technologies (big data analytics and artificial intelligence) conceived to facilitate processes may instead contribute to exacerbating conflicts (Beime et al., 2023; Bracci, 2023; Gendron et al., 2022). However, due to the claims of objectivity of research evaluation systems, the freedom of research seems to be threatened (Argento & van Helden, 2023), with negative implications for innovative and risky research agendas (Gendron & Rodrigue, 2021; Marques et al., 2017). In this complex picture, it is relevant to point out Gendron and Rodrigue's (2021) concerns regarding the possibility that evaluation systems and the use of rankings may generate adverse effects for the academic system, institutions and researchers.

The present paper provides evidence highlighting how the potential negative consequences of research evaluation systems, as feared by Gendron and Rodrigue (2021), are already in place. Specifically, the study focuses on research quality evaluation processes through the framework of the so-called *administrative evil* (Dillard & Ruchala, 2005) and the conceptualisation of the *banality of evil* proposed by the political philosopher Hannah Arendt (1964).⁵ Drawing on previous literature analysing the masking conditions that are typical of administrative evil (e.g., Adams & Balfour, 1998; Dillard & Ruchala, 2005; Hoffman et al., 2012; Neu, 2001), the study aims to investigate how research evaluation can configure banal and unoriginal evaluative practices capable of setting in motion a *fatally efficient machine*, and the impacts of this for the academic system, institutions and researchers. The recourse to the expression *fatally efficient machine* is linked to the need to highlight how an efficient evaluation process is not necessarily effective in achieving the purposes for which it was crafted. On the contrary, as we will discuss, it could even lead to a means–end inversion whereby the goal of evaluation becomes the strengthening of evaluation itself.

Crucially, the paper's contribution is not limited to the focus on the masking strategies and instruments that drive a *fatally efficient evaluation machine*; it also introduces a novel multilevel perspective of analysis (i.e. macro, meso and micro). This approach, which is not commonly used, offers a fresh and intriguing way to comprehend the complex interplay of factors involved in research evaluation. These phenomena are generally scrutinised from a single perspective such as the micro level, with a focus on changes in researcher identities (e.g., Beime et al., 2023; Marques et al., 2017), or the meso level, by investigating the effects on university business models (Gebreiter, 2022). However, these approaches fail to fully and holistically capture the intricate interconnections between the different levels.

The study focuses on the research evaluation technologies and practices that were utilised in the context of the 2015–2019 Research Quality Assessment (VQR3) exercise in Italy, which concluded in 2022. The Italian university system, with its turbulent regulatory environment, the implications on resource allocation, and the constant need for academics to adapt and renew their approaches, presents a fascinating context (Manes Rossi et al., 2022). This is a scenario that piques interest due to the incomplete *marketisation* process, a feature not found in Anglo-Saxon countries (Parker et al., 2023), and the frequent absence of the ex-ante provision of the criteria underlying evaluation processes (Rebora & Turri, 2011).

The article uses an autoethnographic methodological approach (Alvesson, 2003) based on the authors' direct knowledge and involvement in the evaluations and/or the university system subject to these, as well as the results of 19 semi-structured interviews conducted between June 2022 and October 2022. The interviewees were selected to obtain a balance in the views and perspectives of various actors involved in the processes under investigation. The scope of the study is limited to the humanities and social sciences fields, due to the deep-rooted differences in research, publication patterns and evaluation processes from those characteristic of the so-called hard sciences.

The paper notes how VQR, like other research evaluation systems, brings about several (un)intended effects. At the university and departmental level, VQR has instigated a heightened emphasis on the need for accountability in research activity. However, it has also significantly influenced decision-making processes, extending its reach to recruitment, career paths and the very identities of researchers. This influence, particularly in terms of the ethical principles of autonomy and freedom of research, has also affected the distribution of research funds from the national to the departmental level. All these changes are closely tied to the emergence of technical rationality and measurement capacity through VQR, sidelining the content aspects, the relational dimension and the long-term perspectives that are typical of the academic research tradition.

In addition, the theoretical lens adopted allows for the interpretation of the results by unmasking the dynamics potentially leading to an increasing dependence on performance measurement systems, which can reinforce the conformity and superficiality of scientific contributions (Gendron, 2015), as well as bend intellectual innovation to logics of a different genesis and nature. Such dynamics force actors to adapt and play by the 'rules of the game', even though they recognise the possible adverse effects of the process, (Gendron, 2008; Beime et al., 2023). Examining the research quality assessment process provides further support for the theoretical construct of the banality of evil and the fundamental role that bureaucracies, administrative processes and quantification tools such as *accounting* play in implementing programs and reforms, regardless of any moral or ethical implication. Finally, the simultaneous analysis of the macro, meso and micro levels supersedes the partial views of the extant debate, offering a more systemic understanding. In this way, the study enriches the debate on performance measurement issues in universities, discussing the distorting effects and limitations of quantitative measurement processes possibly unrelated to – if not in contradiction with – the purposes promoted by the system that led to their introduction. In short, the paper highlights how research evaluation systems such as VQR, although capable of acting as

⁵ In her report on the Eichmann trial, Arendt (1964) uses the expression 'banality of evil' to describe the practices of the Nazis during the Holocaust, with the term 'banal' referring to practices that were unoriginal, biased, and used and reproduced 'without thinking.' The reference to thinking in Arendt presupposes the absence of moral action and not the absence of reasoning (which was essential to the efficiently functioning machine the Nazis desired to create). It is because of the absence of moral thinking that evil is configured as a 'vacuum of good' (Neu, 2001).

effective tools for improving the academic system at an aspirational level, may end up proving to be a *fatally efficient machine* that facilitates and fuels the marginalisation of researchers who do not meet productivity requirements, the standardisation of research and institutions (in a fallacious quest for excellence) and, last but not least, the disappearance of pluralistic approaches that favour the isolation of the academy from the real world. Therefore, the work offers insights that are not only relevant for policy-makers, scientific societies and researchers in Italy but also for the broader international scene, adding relevant pieces to the complex understanding of the topic at hand.

The paper is organised as follows. The second and third sections respectively offer a critical review of the literature concerning research evaluation systems and an explanation of the theoretical framework and its operationalisation for the objectives of the study. The fourth section outlines the research design, with a focus on the portrayal of the context, the methodological approach and the methods of data collection and analysis. The fifth section presents the results, which are then discussed in the sixth section in light of the theory. Finally, some critical considerations relevant to theory, practice and policy-makers, together with the identification of possible future developments of the research, are offered to conclude.

2. The evaluation of research quality: Literature review

Within the framework of New Public Management (NPM), the topic of academic performance measurement and evaluation has been the subject of extensive debate (Guthrie & Neumann, 2007), fuelling numerous critiques. It is worth premising that the literature identifies a growing acceptance of the need for performance measurement systems in the university context, as levers for better university governance (Barnabè & Riccaboni, 2007) and a more efficient allocation of public resources (ter Bogt & Scapens, 2012), as well as for improved decision-making and staff motivation (Esposito et al., 2013). However, increasingly harsh criticism has emerged concerning the ways in which measurement and evaluation are pursued, and their effects (Marques et al., 2017). Indeed, multiple studies have revealed an obsession with measurability and the striving for productivity as harbingers of a change in research values and potentially negative consequences (e.g., Agyemang & Broadbent, 2015; Gendron, 2015; Gendron et al., 2022; Humphrey & Gendron, 2015; Lapsley & Miller, 2004; Manes Rossi et al., 2022; Michelon, 2021).

This highlights the importance of audit culture models in academia that have already revealed their limitations in other contexts (Argento et al., 2020; Argento & van Helden, 2023). The *metrification* of research leads to a greater importance placed on the number of publications and the journals in which they is published, rather than to the content of articles (Paolini & Quagli, 2013; van Helden & Argento, 2020). This can dangerously undermine research and its quality from different perspectives.

The exaggerated use of journal rankings in performance evaluation has been seen as the main factor in the change in ‘research discourse from a language of a discipline subject, discovery, and implications to a language of journal hits, journal scores, journal ranks’ (Parker & Guthrie, 2012, p. 7). In this regard, Gendron (2015), for instance, has pointed out the risk of the obsessive use of journal rankings to evaluate national and international funding request proposals, to make decisions concerning the careers of academics, or even to define teaching loads. The author observed that the aforementioned tendencies lead researchers to consider themselves successful when their publications receive large numbers of citations, which elevates them to *star* status or, conversely, to feel ‘abnormal’ or fear being considered ‘lazy’ if they do not achieve adequate productivity standards. All this can undermine creativity in research (Kallio & Kallio, 2014) and engender a sense of anxiety in scholars (Argento et al., 2020) and a tension to increase the number of publications (Martin-Sardesai et al., 2021), undermining the ethics of academics and the very meaning of an academic career (Grossi et al., 2020, Härström, 2023).

Similarly, the evaluative obsession implies a danger of exalting or, on the contrary, rejecting research topics that are considered *peripheral* (Gendron & Rodrigue, 2021; Picard et al., 2019), with the result of limiting or circumscribing the research opportunities of young researchers and restricting the trajectories of research development (Humphrey & Gendron, 2015). The tendency for this type of behaviour is now creeping into doctoral courses, where the diktat of scientific output to advance in one’s academic career, and the risk of being a ‘loser’ if one does not conform, dominates (Courtois et al., 2020). The focus of young and senior scholars is on publication opportunities rather than on the subjects and methods of research most appropriate to contribute to theory and practice (Gendron & Rodrigue, 2021; Malsch & Tessier, 2015; Raineri, 2015; Sargiacomo, 2003). Such dynamics carry the risk of impoverishing the reach of academic research and opportunities for innovation (Michelon, 2021; Maran et al., 2023). This approach also ends up distancing scholars from nationally relevant topics linked to the culture of a single country in favour of topics that attract the interest of the mainstream international community (Härström, 2023; Seger et al., 2023; van Helden & Argento, 2020).

What emerges, then, is a mutation of academic culture in which there is a move away from the university as a *genius loci* where ideas are freely exchanged and there is room for fertile contamination of knowledge (Mazza et al., 2008). Moreover, the prevalence of the interests of individuals over those of their institutions is consolidated, which often induces a reduced commitment to teaching and institutional and knowledge-exchange activities (Dobija et al., 2019; Lapsley & Miller, 2004). Studies also highlight a kind of schizophrenia: while journal rankings are criticised in departmental meetings or consultations between members of scientific societies, academics themselves take pride in demonstrating their ability to receive positive evaluations thanks to these metrics, in obedience to the dictate of *publish or perish* (Humphrey & Gendron, 2015; van Helden & Argento, 2020). Universities themselves re-prioritise research areas according to the maximisation of bibliometric metrics (Gebreiter, 2022). This configures a cultural and relational problem (Edgar & Geare, 2013; Field, 2015) that permeates the academic system and institutions, eventually impacting the identities of researchers in terms of the selection of research topics and the creation of new academic community *beliefs* (Guarini et al., 2020).

Developments in the most recent literature also highlight some emerging themes relevant to the present study. Artificial intelligence has increased the ability to trace research products through the use of keywords but also progressively tends to circumscribe the expression of autonomous judgements by those who are called upon to play the role of editor or reviewer, subjecting them to the results

of metadata analyses (Gendron et al., 2022). Specifically concerning research evaluation processes, it has been observed that academics end up refraining from open resistance (Becker & Lukka, 2023; Gebreiter, 2022), becoming progressively accustomed to a mercenary mentality or, in contrast – but ultimately with the same effect – to an attitude of *resilience*, i.e. pretending to resist the new logics but de facto complying with them (Härström, 2023; Rintamäki & Alvesson, 2023; Seger et al., 2023). Attempts to tie the dimensions of evaluation more closely to qualitative logics, taking into account a concrete impact on society, have frequently ended up creating a greater distance between academics and the real world (Willmott, 2003). This happens, for instance, with the use of (informed) peer review, instrumentally adopted to legitimise research evaluation exercises, as it is considered capable of ensuring an alignment with bibliometric evaluation (Bertocchi et al., 2015) but going beyond its merely quantitative character. What emerges is how the progressive dehumanisation of the processes of scientific research assessment risks ending up denying the very role of the academy, depriving it of its cognitive habitus based on independence and intellectual freedom, making (semi)automatic mechanisms of research performance evaluation become the main means to ensure proper evaluations (Nørreklit et al., 2019).

In short, the picture described leads one to reflect on the danger of a further drift triggered by evaluation systems and the use of rankings with negative systemic effects on the academic system, institutions and researchers. Gendron and Rodrigue's (2021) critical examination warns of the possibility that it may become impossible for individual academics to pursue what motivates or interests them, simply because the control systems marginalise certain types of research or engagement in practice. Furthermore, they express concern for academic institutions that, committed to pursuing increasing levels of excellence, might instead be led to exaggerated standardisation due to isomorphism. Finally, for the academic system this could create the conditions for a significant reduction in the pluralism of knowledge and a progressive isolation of scholars from reality. However, the growing literature on the subject has so far conducted investigations by looking at the phenomenon along one single dimension of analysis. The present study aims to contribute by delving into the mechanisms through which research evaluation systems, despite the well-known limitations and criticisms mentioned above, manage to create a *fatally efficient evaluation machine*, and its consequences at the level of the academic system (macro), individual organisations (meso) and researchers (micro).

3. Theoretical framework

This study focuses on the processes of research quality assessment by drawing on the framework of the so-called *administrative evil* (e.g., Adams & Balfour, 1998; Dillard & Ruchala, 2005; Hoffman et al., 2012; Neu, 2001) and the conceptualisation of the *banality of evil* proposed by the political philosopher Hannah Arendt (1964). Arendt, in her critical reading of the Eichman trial, mobilises the idea of *administrative evil* to explain the mechanisms behind the actions of human beings that inflict pain, suffering and even death (Adams, 2011; Adams & Balfour, 1998; Arendt, 1964). The scholar focused on the ordinary people involved in the machinery of the Holocaust, those who made it possible and *efficient*. This understanding is also shared by Bauman (1989). Both scholars highlight the involvement of thousands of people who fulfilled their responsibilities, carried out their tasks and ensured a fatally efficient organisational machine, without feeling any moral obligation other than the need to execute orders. Arendt's (1961) view, which is not without criticism, is that Eichmann does not represent a perverse or sadistic subject but a terrifyingly normal man who performed evil acts without having evil intentions. Specifically, Arendt (1961) emphasises how Eichmann's defining characteristic is *thoughtlessness* (verbatim from the original writing). She highlights how he reasonably adapted himself to unoriginal and biased practices in the context of the Holocaust by recklessly using them for the sole purpose of a career in the Nazi hierarchy built on efficiently undertaking his job according to orders and in the absence of any moral consideration of the nature and consequences of their execution.

In the wake of these considerations, Adams and Balfour (1998) make it clear that administrative evil can be distinguished from other manifestations of evil and ethical failures because its appearance tends to be masked by a series of conditions that can lead people to unwittingly perform evil acts. The topic of masking conditions was also the subject of a contribution by Hoffman et al. (2012) who, examining a series of real-life cases, highlight how administrative evil is the result of a series of conditions such as the renaming of subjects and/or objects to obscure or downplay their (un)ethical content, compartmentalisation, instrumental rationality, legalism, the creation of hierarchical accountability structures, dehumanisation, the supremacy of a certain mission, moral inversion, ethical fading and the provision of rewards and sanctions. These factors determine an undeniable complexity that tends to ambiguously mask administrative evil, even taking advantage of the persistence in modern society of roles and assumptions that are taken for granted (Adams & Balfour, 1998).

In accounting studies, the banality of evil framework has been increasingly employed (e.g., Antonelli et al., 2018). Neu (2001) discusses the state of accounting research, highlighting a drift characterised by the creation of subordinate voices and phenomena of institutional inertia. Furthermore, Dillard and Ruchala (2005) take up the so-called masking strategies commented on by Adams and Balfour (1998) and provide some conceptual categories of reference to understand the role of accounting in the materialisation of administrative evil. The authors recognise how instrumental rationality based on pervasive hierarchies of accountability, controls, rules and procedures, and accompanied by quantitative logics, can be decisive in inducing a detachment between ends and moral evaluations, leading to the de-humanisation of actors and hiding organisational violence.

Specifically, instrumental rationality lends itself to the (un)intentional tendency towards de-humanisation, where individuals, their values and their dignity disappear for the benefit of technical progress and cost-effectiveness in the narrowest sense (Adams & Balfour, 1998). This view is in line with the understanding of calculative practices as instruments capable of contributing to the reduction of subjects to quantified objects, thus eliminating their qualities of humanity (Rosenberg, 1983). This is consistent with the idea of extensive evil taken up by Neu (2001), based on a pervasive and well-designed organisation that adopts non-original, biased and thoughtlessly used practices. In the configuration of extensive evil, technologies (among which accounting is rightfully included) may reach a supreme finalism. As an example, in the view of the actors the goal of an evaluation process becomes the evaluation itself and

not the underlying objective for which it was conceived (e.g., the improvement of research quality).

In the conceptualisation of administrative evil, instrumental rationality represents one of the main components operating at both the system (macro) and individual institution (meso) levels, and together with a meticulous regulation of the process, they can convey the new imperatives and moral principles from the macro level to the meso and micro levels. At the latter level, it is possible to observe the effect on individuals and their reactions in terms of the creation of subaltern voices and individual behaviour of a mercenary or respliant nature. All these concepts are employed in an integrated manner in this study, operationalising the idea of the *banality of evil* by adopting a multi-level perspective to analyse the research evaluation processes.

4. Research design

4.1. The study context

To achieve the research objective, we focused on the last Italian research quality assessment (VQR) process, also known by the acronym VQR3. VQR was introduced in Italy following the establishment of the National Agency for Evaluation and Research (ANVUR) by Presidential Decree 76/2010. ANVUR has the function of evaluating the quality of the research outputs of universities and research institutions, mainly through peer review. ANVUR also defines the rules and the process for nominating the members of commissions for the National Scientific Habilitation (Abilitazione Scientifica Nazionale, ASN).⁶

Ministerial Decree 17 of July 15th 2011 officially launched the first evaluation exercise, which covered the 2004–2010 period and ended in 2013. This first round of evaluation was followed by two subsequent five-year evaluation rounds, the last of which – the subject of this study – covers the 2015–2019 period.

VQR has been the subject of fierce criticism and discussion within the institutional and scientific debate (Abramo & D'Angelo, 2015; Reborà & Turri, 2010, 2011). As a result, certain changes to the system were made by ANVUR, but without ever questioning its overall existence. Italy has become part of an international process of dissemination of research evaluation systems, along with countries such as the UK, 11 EU countries, China, Australia and New Zealand (Abramo & D'Angelo, 2015; Reborà & Turri, 2010, 2011).

VQR is aimed at assessing the research quality of universities and their departments. This assessment has had a gradually increasing impact on the amount of ordinary funding conferred to universities (the so-called Fondo di Finanziamento Ordinario, FFO). The results of VQR are used to allocate the premium share of the FFO. By way of example, in 2018 the premium share of the FFO was 23 %, 80 % of which was allocated based on the results of the 2011–2014 VQR. Law 98/2013 established an increasing trend in the premium share of the FFO up to a maximum of 30 %, which was reached in the year 2021. VQR is increasingly becoming a critical variable for the management of universities and their prospects for growth and sustainability in terms of funding. Considering the multi-year permanence of VQR and the impact of its results on the share of FFO allocated to each university, the Italian context represents an ideal case for the aims of the present paper.

VQR development process evolved without changing its original structure. The main stages of VQR3 process can be summarised as follows:

- Following a special ministerial decree, ANVUR proceeded to issue the call for applications in September 2020, setting the minimum number of research outputs that each university and research organisation was required to submit. In VQR3, the minimum number was set at 3 products per researcher, with the minimum and maximum numbers being 0 and 4, respectively. The general evaluation criteria and guiding principles were also made explicit.
- ANVUR subsequently provided for the selection and appointment of the members of the Group of Expert Evaluators (Gruppo di Esperti Valutatori (GEV) in Italian), which represent the committees entrusted with the task of evaluating the research products submitted for each identified disciplinary area. The number of members of each GEV was proportional to the number of researchers who submitted research outputs for the disciplinary area. The GEV, once established, issued specific criteria for evaluating the products.
- Universities and research organisations then had a time window in which to select and submit products for evaluation.
- The submitted products were evaluated by GEV members and/or external reviewers. Each product was evaluated by at least two reviewers.
- ANVUR published the outcomes of the evaluation with information on the performance achieved at the university, department and discipline level. Each researcher also privately received the outcome of the evaluation of the conferred products, to which the universities did not have access.

Fig. 1 outlines VQR research evaluation system as a process that goes from the macro level down to the micro level, resting in bureaucratic-administrative mechanisms developed *ad hoc* and supported by new actors (e.g., ANVUR, GEV). Considering these features, and in line with the research objective of investigating *how research evaluations can configure banal and unoriginal evaluation practices capable of setting in motion a fatally efficient machine and the impacts on the academic system, institutions and researchers*, the analysis embraces multiple levels of observation (macro, meso and micro).

In line with the theoretical framework, this study focuses on understanding possible conflicts between moral imperatives and

⁶ In the Italian system, the Abilitazione Scientifica Nazionale (National Scientific Habilitation) is required for academics to access the positions of Associate Professor and Full Professor. It is granted by a panel of five Full Professors based on an evaluation of candidates' scientific achievements.

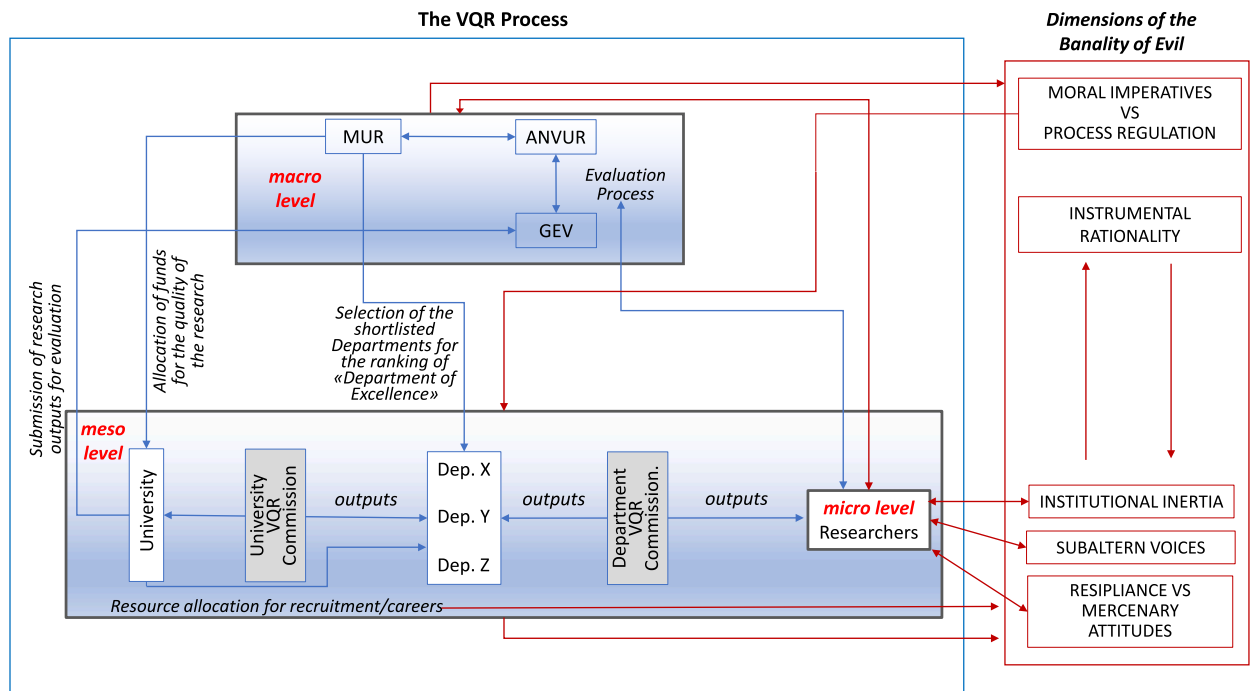


Fig. 1. The conceptual framework.

existing laws and regulations, especially regarding the purposes and impacts of the evaluation exercise. It also investigates the possibility that institutional logics and purposes are set aside to respond to regulatory pressure, short-term ministerial pressures and expectations of new *credos* defined by the academic community. Furthermore, following Neu (2001) and Dillard and Ruchala (2005), we focus on the characteristics of instrumental rationality of the process under scrutiny, delving into its design, implementation and subsequent impacts. A further aspect that the analysis captures is the perceived level of engagement of individuals, in terms of the transparency of information and active and proactive involvement in the processes, so as to identify any areas of institutional inertia (Neu, 2001) or exacerbation of the tensions related to compliance with the required performance levels. Moreover, we investigate whether and in what way the constituent elements of the process have potentially promoted dehumanisation routines and, on the other hand, the different types of resistance triggered in the academic world, understood as absolute rejection of the process and its aims and consequences, as well as utilitarian adherence or total abandonment to it due to its inevitability. This also constitutes an essential premise for exploring the relational profile connected to VQR: on the one hand, to identify any compartmentalisation and, on the other, to understand whether tendencies of resilience – or a *mercenary* attitude – are to be found, with repercussions on the organisational climate and research culture in terms of the possible creation of subordinate voices (Neu, 2001). The risk is that the increasing dependence on performance measurement systems may reinforce the conformity and superficiality of scientific contributions, bending intellectual innovation to new principles and moral imperatives, through the awareness that actors, although conscious of the negative effects of the process, must still act within the rules of the game. In this way, the results of the analysis allow us to unmask the implications of administrative evil connected to VQR, highlighting the rhetoric underlying the founding concepts of VQR, identifying the degenerative effects on moral imperatives and relations, as well as the consequences for ethical autonomy and the relevance of research.

4.2. Methodology, data collection and analysis

We used a qualitative approach by following an autoethnographic methodology (Alvesson, 2003). Autoethnography is suitable in contexts where the researcher studies his or her habitual working context, of which he or she therefore has personal and profound knowledge (Alvesson, 2003). This makes it possible to create '*intimate relations between the field, significant others and the private self*' (Coffey, 1999, p. 1) through a reflexive process carried out by researchers based on their personal experience and contextual knowledge, and integrated with that of the researched subjects (Hickey & Smith, 2020). All authors have participated in one or more of the previous VQRs and have held positions at the university level (e.g., research quality committee member, general manager, member of departmental committees) or in ANVUR (e.g., as a GEV member), giving them a lived awareness of the context and institutional dynamics characterising VQR. The researchers' direct knowledge and experiential background were accompanied by 19 semi-structured interviews conducted between June and October 2022. The interviews were conducted with individuals who were involved in VQR3 process, either actively (e.g., as GEV members) or passively (e.g., as professors/researchers whose products were evaluated). All interviewees were associated with different universities and non-bibliometric research areas in the social sciences. The

interviews lasted between 25 and 100 min. Table 1 shows the types of interviewees chosen, who can be grouped into the following categories:

- University governance representatives such as the rector, general manager of the university, dean of the department, academics involved in research evaluation committees, budget delegates;
- University professors from non-bibliometric subject areas;
- Institutional representatives of the academic context, such as experts in university evaluation systems, members of the Groups of Expert Evaluators (GEVs), members of the National University Committee (CUN), representatives of academies and scientific societies (Accademia Italiana di Economia Aziendale (AIDEA), Società Italiana di Ragioneria e Economia Aziendale (SIDREA) and Società Italiana di Storia della Ragioneria (SISR)), members of commission for the National Scientific Habilitation (Abilitazione Scientifica Nazionale (ASN)).

One of the most crucial aspects of the qualitative interview research process is the definition of the type and quantity of interviewees (Steccolini, 2023). In the present work, the choice of interviewees was made using an intentional approach (so-called *purposive selection*) in order to reach a satisfactory degree of representation of the different actors involved in the research evaluation process. The end of the data selection and analysis process was based on the authors' assessed achievement of what Saunders et al. (2018) refer to as data saturation, i.e. the level at which researchers begin to observe that new interviews do not generate significant incremental knowledge (Qu & Dumay, 2011). This was possible because the same interview structure was followed in all interviewees. The use of a common interview structure also prevented the individual researcher from influencing or being influenced by the interviewee according to his or her personal experience.

The autoethnographic methodology allowed for a process of data collection and analysis linked to a deeper, conscious and systematic understanding of VQR3 process, through an *insider's view* that enabled a focus on the participants' concepts of reality and shared meanings (Harding et al., 2010). It should be noted that the benefits of such a deep connection with the field of investigation can be traced back to the characteristics of reflexivity and self-reflexivity typical of the approach adopted (Davie, 2008; Ellis et al., 2011; Harding et al., 2010; Haynes, 2006, 2011) and require taking into account the need to adapt the traditional meanings of reliability, generalisability and validity that are usually associated with social science research (Haynes, 2011). For this reason, the preparation and analysis of the interviews involved a shared process among the researchers, who also collected and analysed official documents and analyses concerning VQR. These included, for example, the regulations instituting VQR and ANVUR, and publications and analyses conducted by ANVUR regarding VQR3. This documentation made it possible to deepen the reference context, leading to the definition of a coding system for the collection and analysis of content, which was developed in accordance with the conceptual framework represented in Fig. 1.

The interviews addressed several relevant areas of focus, according to which the interview transcripts were coded:

- the purposes of VQR and how its process is implemented ex-ante, in itinere and ex-post;
- the level of engagement of individuals in the process;
- impacts on relationships, 'culture' and the context;
- impacts on research planning at the university, department and individual level;
- impacts on resource planning;
- connections with other directly/indirectly related evaluation processes (e.g., ASN, recruitment).

Table 1
Types interviewees and descriptive statistics.

Interviewee	n	Encoding	Duration minutes
Rector	1	RE	25
General manager of the university	1	DG	44
Dean of the department	1	AI	28
Representative of a scientific society	3	RSS1 RSS2 RSS3	35
			90
			30
Member of the National University Committee (CUN) area 14	1	CUN	26
Member of a group of expert evaluators (GEV) VQR15-19	1	GEV	71
Member of the Commission for the National Scientific Habilitation (ASN)	1	ASN	90
University budget delegate	1	DBU	32
Member of the university evaluation board (NDV)	1	NDV	40
University evaluation expert	1	EVU	100
University professor	4	PU1 PU2 PU3 PU4	45
			40
			35
			75
University professor and member of departmental VQR commission	3	RVQR1	85
		RVQR2	87
		RVQR3	90
Total	19		1,068

The interviews were recorded, transcribed and sent to the interviewees for confirmation and as an opportunity to make changes to their statements. Each author individually coded the approved transcripts according to the conceptual categories of the theoretical framework, i.e. highlighting themes such as prevailing rationality and, therefore, the role of rules and regulations with respect to moral and ethical principles; institutional inertia; the emergence of the search for *super-heroes* and the sidelining of low performers; the emergence of c.d. *mercenary* research or resistance phenomena complicit in the creation of subaltern voices and the creation of new moral imperatives. When coding the transcripts, particular attention was paid to identifying harmony and dystonia, to collectively arrive at an unambiguous view of the meaning of the texts and thereby reduce the risk of bias and influence with respect to one's own experience and the interviewee. The analysis allowed us to not only understand the way the entire VQR process was managed but also the perceptions of the various actors involved. Subsequently, the authors initiated a reflective and self-reflective process incorporating their own personal experience and knowledge, the empirical evidence, and theoretical concepts from the conceptual framework adopted. Like any work with an autoethnographic component, the results should not be read as a closed and final interpretation. The ample material obtained should not exempt authors and readers from seeking further interpretative perspectives (Ahrens & Mollona, 2007).

5. Results

The results are organised so as to fully grasp the elements underlying the complexity of VQR process, following the theoretical framework chosen and enclosing the dynamics at the macro, meso and micro levels as well as the interconnections between these levels.

5.1. Aims of VQR and methods of *ex-ante*, *in itinere* and *ex-post* implementation

At a macro level, it should be noted that the academic components of each scientific area were involved in the operational phases only at the design stage of VQR3 process. During the implementation of VQR, academics were included both directly in assessment coordination roles, through the establishment of the GEVs, and indirectly (not always with the same intensity) by playing a steering role through scientific societies and their committees dedicated to assessment and journal rankings.

The interviews provide a composite picture of the purpose of the process and how it was implemented. In terms of purpose, some actors trace VQR back to a need for resource allocation.

It seems to me that from the very beginning, VQR was linked to the need to distribute, through its result, a portion of the FFO – a rather significant part. So, I believe that the need was very much linked to a dynamic of distribution of the resources of the Ordinary Financing Fund that has characterised the relationship between the State and universities since the (issuance of the law) 537/1993; therefore, the logic in which it was developed is first and foremost a logic of that type. (NDV)

Other interviewees identify a broader logic, however, denouncing a lack of clarity and/or the absence of explicit identification of the ultimate goal and a vague reference to quality and merit.

VQR invokes the usual rhetoric of eliminating nepotistic mechanisms and conforming to international models in the search for legitimacy and the resolution of the historic question of the atavistic ills of Italian university recruitment and progression. The unanswered questions remain: what are merit and quality and how are they measured? (GEV)

Some concern emerged about the risk of VQR3 being downgraded and the limiting of the results to its stated purposes.

There was confusion between the ultimate goal of an evaluation process and the use of evidence downstream of that process. (EVU)

VQR is an important tool that should be used according to its stated purposes and not for anything else, as is being done. It is as if ANVUR had created the tool to provide researchers and departments, in the aggregate, with a measure of research quality, but in the end it is (incorrectly) used for everyone. (CUN)

Referring to the theoretical framework, it is possible to identify an example of how *technique* constitutes a resource for operating by dissociating ends from moral evaluations. The regulation of the process, while remotely recalling the moral imperatives of quality and merit, then leaves (perhaps deliberately) the prevailing rationality to be instrumental. Worthy of note is the widespread opinion that VQR3 was constructed and conducted without considering VQR2 and, above all, with evaluation criteria that were not known a priori, thereby *asking players to take to the field without knowing the rules of the game* (RE).

Thus, VQR promotes the joint and coordinated use of technologies (especially of quantification), subjects and hierarchical organisational structures capable of separating collective action from the moral context in which they operate. Consequently, it lends itself to the (in)deliberate tendency towards de-humanisation and the potential transformation of research values for the benefit of technical process progress and cost-effectiveness in the narrow sense. On the other hand, several interviewees pointed to the indiscriminate use of results for different purposes (e.g., the identification of members of doctoral boards).

The interviewees argued for the need to *recognise the value of the research produced* to bring about systemic improvement that can affect the allocation of resources to individuals, and not only to individual universities.

It would be much more useful if this system could have an even greater impact, if it were useful in allocating resources to individuals and not just to individual universities. [...] If you wanted, you could allocate resources according to individuals,

which is not done for political reasons and academic bureaucracy. [...] In the current research evaluation system there is no link between the utility that the researcher brings to the university and the utility that comes back to it. This impact system fails to capture where the research groups are producing the most. (RSS2)

Faced with the risk that the moral imperative of research quality would be reduced, scaled down or deformed by the game of quantification, leading to a flattening of knowledge, the role played by GEVs, including the creation of sub-groups (sub-GEVs) to preserve the valorisation of differences, appeared positive, and in the opinion of many should be supervised and strengthened.

The creation of the sub-GEVs has helped; the evaluation has been differentiated, and it remains rigid. Harsher than in other cases, but it works. (RSS1)

The GEV, on which I served, decided to set two guiding principles: inclusiveness, embracing a non-absolute concept of quality that could enhance the heterogeneity of approaches and methods, and the need to avoid fundamentalism, by playing a balancing role with respect to ANVUR's forced approach to the distribution of evaluations among the different bands that would have wanted only a few excellent ones. (GEV)

The actors involved in the management of VQR3 in their institutions revealed conflicting feelings and perceptions. At the local level, dissatisfaction with the opaqueness of VQR's aims is even more pronounced. On the one hand, VQR is seen as a necessary process.

It seems to be accepted that research should have a moment of evaluation. Apart from a few 'particular' colleagues (who would perhaps abolish ANVUR and the system), most academics have realised that VQR is the only way to overcome the self-referentiality that often characterises the world of the university. (RE)

On the other hand, VQR seems to have become a new moral imperative, the result of a cumbersome process compared to mechanisms already in place and as an (unstated) methodology for allocating resources that suffers from political influences and arbitrariness.

If one considers research evaluation as an end in itself, there are other criteria/rankings that inform research quality; there is no need to burden the structure/resources with a further evaluation process, often subjective and with such a time lag. If one considers the distribution of resources (FFO prize share) as the goal, VQR can result in non-coherent and deviant logics that often leave too much room for 'political' evaluations. (RVQR2)

In essence, consistent with instrumental rationality, VQR is a process that must be complied with and is 'possibly useful but largely perfectible'. (RVQR3).

In terms of VQR's purpose, the perceptions were heterogeneous but shared the view that a problematic divergence persists between VQR logics, which are only partially made explicit, and the impacts generated. Some see VQR as playing an important role in creating and contributing to an *accountability* logic (broadly understood) but point to problems in its application.

The purpose, theoretically agreeable, does not appear well defined in practice. (RVQR2)

VQR has an important role [...] let us remember that academics were not used to accountability [...] and it has helped to foster internationalisation, accountability and quality. We need to reflect on the divergences between declared logic and generated impacts. [...] If the aim is to allocate resources based on research quality more selectively, paradoxically and inexplicably three VQRs have ended up reducing it. (PU4)

For others, evaluation is instrumental to legitimising their institution, even if only 'cosmetically'.

My university is newly established and, until 2020, it was largely governed by a teaching staff composed of professors on a fixed-term contract. Although the aims of VQR are in principle shareable and should represent a useful moment of reflection to guide the improvement of the university's performance, in our context VQR will not have any impact or usefulness since the results achieved refer to the production of a teaching staff that is no longer part of the university. (PU1)

For others, it is a mere lever of power.

The theoretically acceptable aim should be to steer towards a continuous improvement of research. In practice, I understand it to be a tool for deciding (in a not entirely transparent manner) how to allocate the premium share, without duly considering the implications in terms of the progressive impoverishment of knowledge (one works to maximise indicators, not one's ability to understand and explain a phenomenon in a relevant way). It is based on a partial claim of objectivity that greatly complicates the political dynamics associated with it, both internal to the groups and external to them. (PU2)

In short, it emerges that VQR configures the creation of new moral imperatives for which evaluation becomes the ultimate goal, rather than a tool for pursuing research quality. The same moral imperatives may also end up compressing, rather than fostering, research quality and merit.

5.2. VQR process

At the macro level, the interviewees made it clear that the main limitations of the process are related to the delayed specification of indicators, criteria, targets and modes of application.

Only the broad trends are clear. (RSS3)

Even after months, the fundamental results are clear but other aspects that would arouse an indubitable right to information are not so obvious. The process, as it unfolds, has remained fairly unknown. (RSS1)

It must be clear that the presence of quantitative measures can give the illusion of objectivity in the evaluation process, when in truth those measures are linked to symbols and phenomena that require an intersubjective interpretation. Clarity of process, therefore, implies clarity of key terms and concepts on the part of all those who enter the evaluation process in the role of key actors. (EVU)

The interviewees commented on certain aspects that should be improved in terms of transparency, clarity, timeliness and engagement of the different actors in the system.

Analytical perception is not always applicable. Rules are set ex-post, whereas they should be clear at the beginning of the evaluation period to ensure consistency and reasonableness, to understand what will be evaluated and how. Establishing rules ex-post involves a certain degree of irrationality. (RSS3)

Our evaluation process raises several questions. Prior communication was neither clear and transparent nor effective and timely. [...] The process needs to be more shared and transparent: panellists should be identified by the scientific societies, starting with the president, among those with evaluation expertise, and not simply extracted from lists. (RSS2)

The interviews reveal a flaw in VQR bureaucratic regulation and a strong pervasiveness of instrumental rationality, with an exasperated level of proceduralisation that induces the acceptance of irrational approaches and implementation rules by the scientific community. In this regard, the interviewees referred to examples of late and unsystematic approaches in the selection of GEVs and (downstream) reviewers, as well as a substantial imbalance in the attribution of products awarded by researchers who changed their affiliation over the evaluation period.

Preparation needs to be more careful because the reviewers as the terminals of the evaluation process need to be more aware, not only of the impact in general but of the individual scores they assign to the evaluated products. [...] There was also difficulty in finding evaluators. (CUN)

In the case of researchers who changed their affiliation, the products went to the university where the researcher worked at the time of submission. If ANVUR had allowed the time aspect to be taken into account (as was possible for sickness, maternity or assignment exemptions), some contradictions would have been avoided. In Italy, where universities are predominantly public, the effect may seem minimal but a phenomenon of 'cross-subsidisation' has been de facto allowed. (RSS2)

In addition, there were no mechanisms for rebalancing different subject areas in the pursuit of a common approach that did not properly value the obvious differences.

I Noticed that the scoring was not homogeneous between areas, impacting mainly on the non-bibliometric areas. [...] the scale values (identical for all areas) risked not corresponding to the quality of the products. (CUN)

There are also critical issues related to the levels of engagement of individuals, from which institutional inertia emerges. This is a transversal issue and is closely linked to the elements of clarity and communication. The problem lies in two fundamental issues: adherence to the evaluation process, its rules and its logic.

In some cases, I had the distinct feeling that there was not good coordination at the local level. Sometimes looking at the types of products submitted (e.g., teaching manuals) I wondered: why? (GEV)

The engagement of individuals is a necessary objective. This also implies that there is a need to report on the results achieved in relation to the objectives set, in accordance with the well-known principle of accountability. (EVU)

As VQR manager, I tried to have a shared process with more meetings and more minutes. Many [colleagues], however, suffered from the process. Since it is not a transparent process and fearing its use for other goals than its stated purpose, people were concerned about 'protecting themselves'. At the preliminary stage, the process was not clearly communicated. This led to the defence of certain areas 'regardless of' and even to the detriment of the department. The engagement led to some real conflict, especially for researchers still in the pursuit of their careers, who wanted to be represented with three products. Many, for example, did not understand how recruits were evaluated. The delay and the opacity of communication increased the backlash. Especially those climbing the career ladder who felt committed to (submitting) three products. I achieved a reduction in conflict by having many meetings and ensuring that there would be no single report card or that something like this would not be used. (RSS2)

The interviews show no shortage of criticism, as well as a perception of risk on a personal level and an emphasis on individual instrumentalisation.

The natural consequence of an unclear process is that the impact and purpose itself is distorted, manipulated and misrepresented at the departmental level. (RVQR2)

In line with the pervasive instrumental rationality emerging, 'the "paper-pusher" role of the dean and his team is reaffirmed' (PU4). 'There is a debasement of the evaluated figures, on whom the performance of a "modest task" is imposed' (PU2).

A more active involvement of the academic components in the design phase, even before implementation, would be desirable. In my opinion, this has been lacking. We were told to perform a modest task, which, moreover, if we did not comply as asked, someone else would take care of for us. It honestly seems to me a bit demeaning of the dignity of researchers. (PU2)

The problem of local implementation was very much felt by the 'insiders', on whom most of the implementation burden was placed. The operational part was borne by a few members of *ad hoc* VQR panels set up within the departments, which were responsible for determining which products to submit for evaluation. In some cases, authors were asked to indicate a preferred ranking of products; in others, the panels proceeded independently. Most interviewees noted how 'in one or the other approach there was no lack of controversy or need for "resensitisation" and "training" of colleagues' (RVQR2).

The issue of exempting some researchers from submitting a research product was controversial, with very different approaches between individual universities, many of which urged departments to maximise quantity (in the absence of clarity regarding the consequences of exploiting exemptions), frequently at the expense of product quality.

The problem of non-exploitation of exemptions was mainly related to inactive and very young people who could not deliver three quality products. There was some discontent on the part of the people who had to submit four products, who either felt that they were a cut above or felt put at risk. The department's choice was not to put too much emphasis on it, also because every time individuals were pushed, potential hotbeds of controversy were generated and further questions were raised, which VQR commission had no way of answering in the absence of clarity at the national level. Disagreements were created even though the progression planning had already been done. (RVQR1)

It was generally agreed upon that the management of the selection and submission process by the deans and support teams was complex, very time-consuming and mentally demanding.

It happens, as this time, that the choice to compensate is managed by people helping the department deans, with the selection of a fourth product not always up to scratch. In the big departments it goes better. But it is the game of competition between universities – that's where the FFO comes from. Deans and delegates get their hands on it, but in some cases also other people who may not have fully understood the mechanisms. (RE)

Thus, two types of problems emerged: the need for clear and a priori rules and reasons for engagement, and the issue of IT innovations/limitations.

In this game, it becomes important to have predictive possibilities about possible outcomes based on the rules of engagement, which should be known long beforehand and remain stable. (PU4)

The reactions of academics were heterogeneous. Some, adopting a mercenary mentality, manifested adaptation and refrained from open resistance. Others adopted attitudes of *resipliance*, pretending to resist the new logic when they spoke to colleagues but adhering to it in their actual behaviour. Both attitudes fostered the process of dehumanisation and the creation of new moral imperatives.

The perceived level of involvement of individuals in the process was very low, and the process remained confined to the responsibility of the few involved. The dialectic that developed in some local contexts revealed not only a lack of depth in the process but also a marked utilitarian profile in various aspects, from recruitment planning to the positive and/or negative impacts on young people's careers.

The reaction of individuals has been different, and this has, in my opinion, depended on the individual scientific production. I noticed that colleagues with a more conspicuous and valid scientific production showed a greater propensity to collaborate, while the others were more reluctant. (AI)

Most colleagues are only passively affected by VQR fulfilments, sometimes suffering them. Hardly any of the researchers involved thinks in terms of the common good of the department and risk letting personal perceptions prevail over the (more or less clear-cut) criteria used by the GEV. (RVQR3)

Whenever engagement was attempted, there was no shortage of questions (which VQR committee could not answer) about how much funding would be obtained and which positions would be opened. The very young, far from caring about VQR, are more oriented towards their personal career pathway; the department's performance is only of interest to the extent that the institution receives resources (generated by others) to enable career advancement. For most, in fact, I believe that VQR is a matter for the department's top management and the few colleagues who – poor souls – assist it in its fulfilment. (RVQR1)

Also interesting was the topic of peer review, which started as an important moment of involvement and turned into something else.

There is also a risk of perpetuating, through informed peer review, subjective judgments that are not based on an appreciation of the quality of research products. Evaluation includes a margin of discretion. For area 14, discretion has, according to some, been used a little too flexibly. I would call it 'the pebble-in-the-shoe syndrome' [...] Perhaps the mistake of evaluating from an incorrect perspective was made, where the reviewers thought about how they would do the research. (CUN)

5.3. The impacts of VQR and its connection with other evaluation processes

The interviews also revealed the impacts of VQR3 on university culture. At the macro level, VQR3 prompted universities to team up and scientific societies to take on an important role in settling interests, generating possible 'witch hunts' and conflicts between

research areas.

The extremes stand out or are penalised. Those who are poorly evaluated certainly then pay a price in the internal logic of the department. The others who stand out make their role count. The whole is no longer evaluated. Those who get better results feel they are owed and complain about others. The others feel mistreated and feel they are still being watched. (RSS1)

However, the process also highlighted the importance of research teams rather than individual ‘super-hero’ researchers.

Team building was easier where there was excellence. (RSS3)

The super-heroes didn’t do it. The groups won. For those who have a group, it is the group that wins. (RSS2)

While cultural profiles may suffer from yet another unannounced ranking, the results of the evaluations gave hope for a rethinking of strategies more focused on the content of what is published and the outlet of publication. However, there is strong pressure from departments to publish, and this, from a cultural perspective, impacts young people. The perceived danger is that evaluation will create new moral imperatives and new principles of authority related to the choice to publish for and by virtue of evaluation. This can influence the selection of topics, approaches and methodologies and, in extreme cases, even reverse the quantity–quality relationship. VQR may constitute a lever of extra-scientific orientation capable of legitimising ‘peripheral’ research topics or delegitimising other ‘central’ topics. This can generate dangerous choices, triggering topics and journal selection processes no longer guided by reasons of curiosity, innovation and relevance but by the utilitarian ends of individuals and groups, creating subaltern voices not aligned with what becomes relevant given the evaluation.

There is a generational difference between those who grew up without VQR and those who were born with VQR and journal rankings. The impact on culture and climate is massive. For those of us who study these evaluation systems, this is nothing new. Incentives guide behaviour; there has been a mad rush to try to publish in journals included in VQR, based on these evaluation mechanisms. VQR3 for area 13b has helped to reduce the strength of this race and make it clear that we are not bibliometric, but for young scholars the prevailing message remains. (RSS3)

Right now, scholars choose the journal more than the topic. It is not only the effect of VQR, but more broadly the composition of the journal ranking that has given the research a strong imprint. Researchers choose the journal, not a theme. (RSS2)

In my opinion, the exasperated search for products to be placed in ‘ranked’ journals has led to the choice of collateral research topics, which are considered more easily publishable, even beyond a real scientific interest in the topic and possible impacts. (RVQR1)

The cultural profile, on the other hand, seems to be intimately linked to aspects of research and resource planning, and to the connection with other evaluation processes charged to individuals (as in the case of the ASN) or institutions (evaluation of recruitment).

What was supposed to be a collective evaluation is also reflected in the individual evaluation. There is certainly no witch hunt, but when a professor who came out badly from VQR goes to ask for an investment of resources, he comes out badly because there is a political evaluation. The fact that he was evaluated negatively certainly has political weight for the department, especially when the results are particularly bad. (RSS1)

There is pressure from the departments to incentivise publication; this has consequences, especially on young people who do not give weight to VQR and think mainly of the ASN. This is natural but produces a dyscrasia. (RSS3)

Research and resource planning are difficult to separate, and a more decisive intervention is needed to align and enhance the evaluation of research, teaching and knowledge exchange, making individual evaluation processes (i.e. ASN) uniform and with known and stable criteria. By making the impacts of evaluations more explicit, the engagement of individuals and the cultural training *tout court* of young researchers are seen as possible.

The exclusive focus on the final result, i.e. the products being evaluated in VQR exercises, could stimulate undesirable behaviour which aims at improving the evaluation by contrivance. From this point of view, the increasing focus on knowledge exchange with industry and civic society also increases reflection regarding the impact of the work. Certainly, emphasising the idea that one must publish a lot regardless of quality is wrong, as is the idea that what one writes is of no use. That is why between the V [valuation] and the R [research] there is a Q [quality], to mean the intention of precisely assessing the quality of scientific production. The Q deserves more thought! Indeed, before questioning the criteria, we should ask ourselves what the relevant dimensions of research quality are today and what they should be in the future. (EVU)

The interviewees highlighted the central role of scientific societies, in particular in offering an unambiguous message to those who need to pursue a career. The risk is the reinforcement of institutional inertia and the phenomenon of moral inversion, leading towards results that are quite different from those of quality and merit.

The scientific societies play a key role precisely in defending the quality of research, given the risk of ‘article factories’ boosting the system’s numbers, for example, through opportunities for paid publications or exchanges of authorship in multi-name publications. Transparent processes and the safeguarding of the accountability principle are the only possible brakes for such deviations. (EVU)

Focusing on the meso- and micro-levels, the interviewees pointed out that the competition inherent in VQR makes it a non-neutral process that can greatly influence research, but also other aspects. The competitive climate has significant reverberations on teaching

and institutional commitment, especially for young people. Competition and conflict, which feed on the above-mentioned mercenary logic, must be interpreted in light of the fact that the research climate and culture are affected by VQR and may foster attitudes that are not quality-oriented.

Young scholars do nothing that is not consistent with competition. They develop this awareness as early as their PhD. They get into a whirlwind process to get A-ranked publications [the top journals for the purposes of ASN] [...]; they know that already in their PhD years they have to achieve a few important publications. Teaching is much less important. Commitment to the institution is time wasted or, in any case, has a very limited value. It is a mechanism that pushes individualism and does not foster the creation of local research groups. Evaluation (VQR) is not neutral, it shifts where research is done, how research is done (in the sense of methods) and also the third mission. In a recent report by the Conference of Italian University Rectors (CRUI), VQR that assesses the third mission evaluates public engagement. The various scores that have been created and the benchmarks chosen clearly to point towards this direction. (RE)

VQR represents a potential incentive for a race for publication (*quantification game*) characterised by opportunistic behaviour, self-plagiarism, constrained research choices, trends of the moment, the search for *accessible* journals and paths constructed to exploit any favourable circumstances.

VQR (like the ASN and other evaluation processes) often has nothing to do with research. It may have induced a process of internationalisation, but not necessarily of quality. On the contrary, it can be assumed that the evaluations, without serious guidance from scientific societies, induce research that is consistent with the 'tick the box' logic. But the researcher who wants to do basic research (and this mainly concerns the hard sciences) that has no impact on society and the economy is discouraged, which is a problem, by the way, even in our disciplines. VQR highlights a situation that does not work. It tries to assess the impact of our work, but this impact cannot be measured except by the judgement of our end users. [...] Let's have our end users evaluate the research and see how many understand what we write and whether they need it! (RVQR1)

In line with what has already been said about the danger of institutional inertia and mercenary behaviour or resiliance, there was a consensus among the interviewees that VQR impacts both research and resource programming, albeit indirectly. However, several issues arose, such as how to monitor this supposed impact, how to make it comprehensible at all levels and how to prevent the evaluation from self-referentially selecting what is important by cutting off entire fields of study and theoretical and methodological approaches. The principle of authority connected to VQR is the major concern in terms of research culture and, above all, the training of young people. This is pervasive, as there is a close connection between productivity guidelines, research planning and the impact of evaluation on funding.

You risk principles of authority by allowing people to feel good based on rules that have no real basis. A good researcher is so because he/she has published in A-ranked journals. It is a devastating landscape in disciplinary terms. It does not encourage interdisciplinary or interdepartmental research. VQR as it stands destroys departmental research groups. This is also inconsistent with the stated aims of VQR. A stable relationship between researchers should be stimulated. [...] VQR might even reinforce opportunistic behaviour due to non-transparent and replicable criteria. (PU4)

The most frequent complaint was the instrumental (ir)rationality linked to the lack of consistency between ASN and VQR, especially in some areas. In addition to VQR and the ASN, which are frequently cited together even though they often generate discordant behaviour, we cannot forget, then, the funding for 'departments of excellence' that is linked to the National Resilience and Recovery Plan (PNRR). The reference to the PNRR, moreover, is a cue for a broader reflection – recurring in the interviews – on the real capacity of Italian universities, in terms of intellectual resources but also administrative capacity, to attract funding in the absence of adequate organisational support.

The ASN and VQR are distinct processes but proceed in parallel in terms of the logics used. The orientation of the research and the required targets follow quite similar logics. There is a research orientation that goes beyond VQR and ASN, which ultimately permeates both mechanisms. That of departments of excellence is a more interesting game, in the logic of rewarding research capacity. We have been one of them and are even now among the candidates on the recent shortlist. It can be an interesting evaluation mechanism, also for recruitment mechanisms. In this historical phase, we then have the PNRR research projects that will commit researchers to the given directives, so we need to understand how this will be reflected in the outputs.

The PNRR stipulates that research must have specific and usable outputs with respect to the economic system. It is yet to be understood to what extent it is possible to transfer the output of projects to some journals included in the high end of the journal ranking, which have high standards of contribution to the theory of their publications. If this happens, this effort will yield positive results. (PU4)

The issue of internal departmental evaluations is also becoming central. The problem of the strategic management of the evaluation process, consistent with the idea that VQR represents 'a potential generator of financial resources' (DBU) is common to all universities seeking to maximise the outcome. The point here, once again, is related to the dehumanisation brought about by VQR, which by pursuing fully instrumental rationality creates an orientation towards the maximisation of quantitative and procedural results, often contingent and frequently of a financial nature, instead of providing a coherent and well-conceived set of tools and logics capable of enabling longer-term quality dynamics.

Since we receive a large number of resources as a result of VQR, we try to incentivise researchers by achieving alignment. The attempt at alignment goes in the direction of improving the contribution that individuals make to the institution. Of course, it will take time to see the effects. We are like a big ship: even a small movement can produce big shifts in the long run. (DBU) The parameters of the next VQR are not yet known. In fact, we are now halfway through the five-year period (if it is to be five years again) and general principles guiding actions for the remaining time must be identified. Recruitment is certainly one of the elements that must be improved in terms of quality and will probably remain among the criteria for the next VQR. Action must be taken regarding the quality of recruitment to achieve better results. (RE)

VQR evaluation should be linked to the salary increases, which are currently almost automatic despite the bureaucratic formalities required: the need to submit three products over a very long period, the quality of which is apparently not subject to any control, goes in this direction. The paradox, therefore, is that a researcher rated 0 in VQR can benefit from salary increases thanks to the same products submitted for the assessment. (VQR3)

6. Discussion

The analysis conducted shows how VQR, like other international research quality assessment processes, is part of the attempt to modernise the Italian university. Like the concept of *who can be against efficiency?* (Hopwood, 1984), VQR conveys the principle of quality in research, delegating to instrumental rationality the enabling logic capable of achieving it (Dillard & Ruchala, 2005). The techniques, technologies and administrative procedures aimed at quantifying what is or is not quality research have played a prominent role in all editions of VQR. The administrative bureaucratic system has also been linked to the need to allocate an increasing share of public funding to universities on the basis of 'merit', thereby creating a strong pressure for universities and departments to try to maximise the outcome and the funds transferred.

Interesting insights emerged from the interviews regarding the effects generated by VQR in terms of the diffusion of competitive forces and tensions at various levels. This is similar to the processes already in place in other contexts where research evaluation has been practised for years (Martin-Sadersai et al., 2021). Fig. 2 highlights these effects and the link with transmission mechanisms at the macro, meso and micro levels, which are then traced back to the adopted framework.

At the macro level, despite its obvious limitations VQR process is accepted by virtue of positive rhetoric related to the importance of research evaluation, the transformation of research teams and the increase in internationalisation. Despite criticism from various sources, VQR is not questioned as such but, rather, is seen as a part of the academic system requiring improvement. The vision of an evaluation that is necessary because it is rhetorically a harbinger of (substantial?) improvements, however, unveils the dynamics of *masking* the so-called administrative evil (Adams & Balfour, 1998). The process under consideration is aimed at maximising the results of the individual university and/or the individual department and sees little involvement of individual researchers who are subjected to the choices and criteria for evaluation. As Dillard and Ruchala (2005) point out in their study, instrumental rationality obscures and over time eliminates alternative principles and values, manipulating institutional, organisational and individual instruments, processes and structures.

VQR thus becomes the means – legitimised and accepted in spite of everything – for introducing an instrumental rationality based on economic and commercial principles. It operates by adhering to the abstract principle that efficient quantification and the introduction of incentive mechanisms can guarantee the university system increasing levels of research quality through the self-regulation of researchers. This is irrespective of the distorting effects on the identity of researchers, their careers, the organisational culture and the quality of the organisational contexts in which they operate.

Within universities (the meso level), the process of technical rationalisation is made possible by the construction of administrative mechanisms and processes supported by increasingly efficient quantification techniques, and the people involved have limited chances/willingness for independent action (Arendt, 1964). This instrumental perspective blurs ethical issues, leading to a moral inversion whereby even immoral actions are reformulated as positive actions undertaken in the name of the public interest and are qualified as the improvement of performance (Adams & Balfour, 1998). The concept of routine assumes relevance in this picture as the specification of roles, and the rules embedded in bureaucratic hierarchies protect participants from having to deal with the moral consequences of administrative actions (e.g., Dillard, 2003; Funnell, 1998). This generates a real detachment, leading to distance between intention, action and outcomes, with a focus on technical meanings of responsibility and quality (Baumeister, 1997).

As Dillard and Ruchala (2005) state, human beings are represented as objects to be manipulated according to the norms and logics that are an integral part of economic calculation. This emerges in VQR process at the university and/or departmental level, where the aim of maximising VQR outcome leads to the introduction of internal governance tools and mechanisms to enable this. The processes of quantifying research quality become efficient and rationally applied tools, which cause a loss of focus on the fate of individuals, communities and, more generally, on ethical principles such as the autonomy and freedom of research and the values linked to disciplinary and organisational traditions (Becker and Lukka, 2023). The consequence is the rise of internal tensions between fields and between departments in relation to internal resource allocation processes.

In some cases, individual researchers are neither involved nor informed about which and how many research products their institution has submitted, which limits or hinders the opportunity for involvement of the actors. All this is managed by an efficient administrative machine in which individual researchers are turned into numbers, beyond any consideration of human and contingent aspects. In some contexts, this can go so far as considering and treating researchers as objects and their publications as products and '*as mere numbers that can be easily manipulated and casually disposed [...] It allows individuals to manipulate fellow human beings as things until they [...] are no longer perceived useful*' (Rosenberg, 1983, p. 12). The concept of the *inactive* researcher, the penalties for those who do

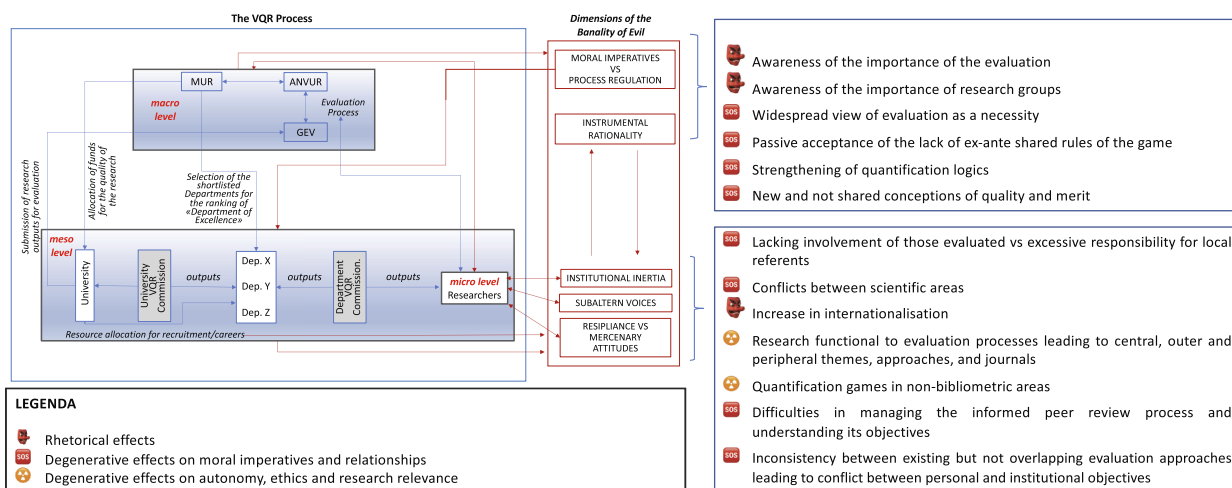


Fig. 2. The research evaluation system in Italy: the effects.

not contribute to VQR or the rewards for those who excel become stigmas or medals, generating forms of psychological and organisational violence and, at the same time, shaping the organisational culture. Reward and incentive mechanisms have the function of making new values more acceptable, impacting the individual interests of the researcher and not the collective interests (Hoffman et al., 2012).

Thus, phenomena of so-called ‘authorised violence’ (Arendt, 1964) are facilitated by virtue of the action(s) that promote the organisation’s end goals and which are pursued through rules and procedures designed to induce actors to act towards these ends rather than towards their well-being. Regarding VQR, discipline is imposed through an evaluation system based on quantitative and financial criteria that is an integral part of the hierarchy of controls. The hierarchy authorises actions, actors ensure that technically rational knowledge is applied appropriately, and technology facilitates the efficient and effective performance of duties (Becker and Lukka, 2023). In this sense, the adaptation of researchers to such processes is also observed, with the consequent limitation or elimination of forms of explicit resistance, even to the point of developing a mercenary mentality (Becker and Lukka, 2023; Seger et al., 2023).

It is not surprising, therefore, that young researchers consider this evaluation mechanism to be an element of organisational discipline (if not organisational control) to which they need to contribute, within a context where roles and practices are defined. Thus, we see the emergence of tools and procedures for the rational quantification of the potential contribution of the individual researcher, research team, department or other actors involved. Dillard and Ruchala (2005) note how the bureaucratic administration of *evil* is not hidden but is so well entrenched in administrative processes that it is not overtly visible and permeates the academic system (Edgar & Geare, 2013; Field, 2015). In line with Bauman (1989), researchers are reduced to a set of quantitative measures and thus undergo a true process of dehumanisation – the separation of the object from the subject, giving the object (the quality of research) primacy over the fate of the subject (the effects on the researcher). The critical point here is that the moral imperative of research *quality*, in this case, is reduced, downsized and deformed by the game of quantification. More specifically, the Italian research system moves towards a concept of quality that is extra-scientific in terms of research practices and yet is capable of determining the conduct of researchers.

In this sense, by deviating from its stated objectives VQR ends up (at the micro level) impacting the career paths of individual researchers, constraining researchers’ choices regarding what to research, where to publish and with whom to publish, thus changing organisational contexts and the university system. It also ends up impacting the identities of researchers who self-identify as *super-heroes* or as low-quality or inactive researchers. This only confirms how instruments – sometimes imperfect – introduced with the only aim of promoting economic and administrative rationality inevitably, or deliberately, lead to the modification of the identity and behaviour of individual researchers (Argento & van Helden, 2023; Seger et al., 2023).

Indeed, journal rankings, together with the associated competitive mechanisms for measuring the performance of institutions (VQR) and individuals (ASN and career progression processes), incentivise the introduction of ever-higher performance targets, with pressure on young researchers to publish in *top* journals, creating new categories of identity such as the *super-hero* researcher (Argento & van Helden, 2023). This happens even though everyone recognises the limitations of rankings, despite the fact that many universities recognise the principles of the Declaration on Research Assessment (DORA),⁷ and even though the same rankings are sometimes known after the evaluation exercise, without any success (or desire) in curbing this practice. The researcher becomes associated with metrics, in terms of the score obtained in VQR, their H-Index or the impact factor (even in non-bibliometric fields), disassociating him/her from what he/she researches, from the scientific project he/she pursues and from the innovativeness of the topics he/she deals

⁷ The Declaration on Research Assessment (DORA), established in 2012, is an initiative to disseminate best practices and principles in research evaluation processes. Among its main recommendations is the rejection of bibliometrics as a proxy for the quality of individual articles, and even less so for individual researchers.

with, with future individual opportunities also being associated with those metrics. The researcher becomes functional, or dysfunctional, to the pursuit of the goal of a predefined concept of quality. As time goes by, this could be a problem due to the flattening and excessive homogenisation of knowledge brought about by external rankings and judgments, undermining the preservation and enhancement of traditions and differences and forcibly creating areas deemed *central*, with others being *peripheral* or outside the boundaries of what is deemed permissible (Gendron & Rodrigue, 2021).

The modification of individual identity and the risk of loss of sociality and collective responsibility in the research process underlies all attempts throughout history to solve a problem through a bureaucratic-administrative system capable of efficiently managing and quantifying new processes and practices and identifying roles and responsibilities (Arendt, 1964). All this is then placed within a legislative framework that legitimises and institutionalises these mechanisms. Although with necessary distinctions, what we observed in the context of VQR resonates mechanisms and logics already known to have been mobilised to efficiently manage ethnic groups or entire populations, displacing a sense of individual and ethical responsibility through bureaucratic efficiency and quantification (Antonelli et al., 2018; Davie, 2000).

It is interesting to observe how no regulatory or organisational act has undermined the principles of autonomy and freedom of research, but it is equally evident that these come in second place following the primacy of certified quality according to institutionalised evaluation mechanisms. VQR, therefore, does not emerge as a neutral instrument that rationally applies administrative techniques and procedures, but as a means of influencing moral and ethical values in the operation of research in universities, creating what is called a *moral inversion* (Adams & Balfour, 1998). Autonomy and freedom of research remain, but the process of assessing research quality inevitably directs it forcibly. The pursuit of ‘quality’ scientific production runs the risk of displacing the development of alternative research paths, which are by their very nature uncertain. Where an article must be published to be of quality is also predefined, without regard for its content. What appears to be happening is a displacement or repositioning of the priority scale of the public values underpinning scientific research, changing the academic profession in ways already occurring in other contexts (Gebreiter, 2022).

If it can be said that VQR has substantially achieved its objectives, leading to an increased awareness of the importance of research evaluation as well as to the internationalisation of research, it is also true that the moral imperatives that arise increasingly entail the materialisation of new risks (e.g., ‘*quantification game*’) ascribable to administrative evil. In short, the analysis suggests that the idea of quality and merit underlies administrative evil and that together with the tools, processes and systems applied at different levels, it can mask the organisational violence which affects individuals (Dillard and Ruchala, 2005). The multi-level view adopted, which combines micro (individual), meso (organisational) and macro (university system) perspectives, has allowed for the unveiling of the mechanisms and conditions of the masking of administrative evil (Hoffman et al., 2012) – which stems from the concept of the banality of evil (Arendt, 1964) – in the context of research evaluation systems.

7. Conclusions

The present article is part of a stream of studies that have variously highlighted the worrying thrusts towards the logic of quantifying academic performance, leading towards the development and implementation of increasingly complex systems and ‘apparent’ tendencies towards objectification (Gendron et al., 2022). In particular, the study focused on the recent VQR3 process in Italy. Within the conceptual framework of the so-called *administrative evil* (Dillard & Ruchala, 2005) and by drawing upon the conceptualisation of the *banality of evil* proposed by Hannah Arendt (1964), the paper analyses the Italian VQR3 in detail. The examination highlights how VQR3 has fuelled the progressive affirmation of unoriginal and biased evaluation technologies and practices already widely experimented with in other contexts (Marques et al., 2017). These practices were first applied and then used in an un-critical manner, gradually creating new and restrictive moral imperatives and principles of authority which are far from the traditional values that have always inspired academic culture (Maran et al., 2023).

At the macro level, the study reveals the need to reflect on the competitive tensions concerning access to the premium share of funds distributed to universities by the Ministry of University and Research, which represents an increasing share of the FFO. These tensions are exacerbated by the widely acknowledged underfunding of the Italian university system. Moreover, the results of VQR affect the allocation of funds that universities will receive for five years, creating significant operational limits for those structures that do not achieve excellent levels. It is evident how this competitive dynamic, as it is structured, spills over dangerously at the level of internal governance – at the meso level – in the relationships and competition between departments and in the allocation of resources, which increasingly follow criteria linked to how well or poorly they perform in VQR. Departments, in turn, are increasingly using and/or considering VQR results and VQR targets in their decisions concerning the recruitment and career processes of researchers, as well as decisions concerning research groups or research fields, essentially producing a further unstated effect.

Finally, at the micro level, researchers (whether individual or coordinated in groups) have in turn come to understand the values and ethical principles that VQR conveys, sometimes even in spite of explicitly criticising them. In particular, new researchers – the so-called ‘VQR generation’ – consciously develop their own research and publication strategies with an eye on the evaluation mechanisms. Together with the pressure exerted by certain particularly attractive journals as research outlets, this can generate dangerous dynamics leading to the creation of central and peripheral research areas in each scientific field. There is a risk of triggering processes for the selection of topics and journals no longer driven by curiosity, innovation and relevance but by utilitarian ends of individuals and groups, as it has already been demonstrated in other countries such as Britain (Marques et al., 2017). We do not claim that competition cannot have positive effects and improve the quality of research. Rather, we note that the lack of clear rules and the presence of unstated intentions, together with the lack of coordination with other evaluation systems in the same university system, may engender an obsession with rankings provided at the institutional level for different purposes (e.g., ASN, VQR, recruitment). We also indicate

that the positive potential of research groups within the same department should not be overlooked or taken for granted. In fact, our results show that VQR can limit academic collaboration within a department, penalising the daily discussions and exchange of ideas and reflections that can have a positive influence not only on research but also on other missions.

Based on these considerations, therefore, it is possible to identify several contributions of a theoretical nature in this paper that enrich and expand the literature in the direction of a more consolidated awareness of the multiple implications of accountability in universities. The study draws attention to the fact that the quality of academic research is a new responsibility for the university system, for individual universities, for departments and for researchers, from which the full legitimacy of accountability processes emerges (Dillard & Ruchala, 2005). The various actors in the system have to account for this ethical finalism, thus accepting the measurement and evaluation systems and aligning themselves with the logics that characterise the evaluation exercises. The so-called ‘metrification’ of research (van Helden & Argento, 2020) leads not only researchers but also institutions to place more importance on the number of publications and the reputation of the journals in which one publishes rather than to aspects still considered fundamental to academic research (Malsch & Tessier, 2015). As discussed in the literature, the mutation of academic culture is progressively stimulating the prevalence of the interests of individuals over those of the institution and induces a progressive detraction of commitment to teaching and institutional activity (Dobija et al., 2019; Lapsley & Miller, 2004). Similar situations have already occurred in other contexts, such as in the UK (Marques et al., 2017), where following some initial hostility from academics, it was the researchers themselves who reified the system, distorting the ways of doing research and disseminating it (Harley, 2000). The ineluctability of the use of research evaluation for the primacy of quality and the possibility of qualifying it objectively is in fact refuted in countries such as Germany or France where research evaluations are not undertaken and the ability to perform research of value for cultural, economic and social development does not appear significantly penalised.

In this context, this paper contributes to the literature by proposing an innovative, multilevel (macro, meso, micro) view to better conceptualise the dynamics, processes, instruments and actors at play and the relationships between them. Previous studies have often focused on the effects of research evaluation systems at individual levels, and in particular the micro level, thus losing the overall view of the subject of academic research metrification. Understanding the constituent elements of the ‘administrative evil’ and its dynamics is conducive to the critical breakdown of complex phenomena, such as the reform (in a neo-liberal sense) of public services. On a conceptual level, a dangerous tendency emerges from our work, as it unveils the primacy of technical rationality and measurement capacity to the detriment of the innovativeness of research, the artefactual and utilitarian creation of frameworks and boundaries, and the risk of endangering the relationships between researchers. The long-term perspective impacts of these changes require further theoretical investigation. The multilevel perspective of the framework adopted allows for the reflections to be extended beyond the Italian academic context, as well as to other public service reform processes.

On a practical level, moreover, the paper’s contribution extends across several areas. First, the analysis conducted highlights the spread of criticalities within the university system. The mechanisms behind research evaluation systems, and in particular the negative effects resulting from the failure to adequately identify purposes, timeframes and methods prior to the carrying out of the process, have led to an increasing misalignment of the various evaluation processes applied to university performance, if not to substantial schizophrenia. The results of this research can serve as inspiration for the competent bodies in revisiting the entire evaluation system. Consistently, the added value of the study is to warn the reader of the concrete risk of emphasising a short-term vision aimed at maximising reward and career opportunities that is not necessarily linked to the free development of innovative but risky research paths (Gendron, 2015). Increasingly, it emerges that what would have been ethically unacceptable to past generations of researchers (e.g., measuring the quality of a scientific article), now becomes ethically legitimate or even necessary and no longer an object of debate (Hoffman et al., 2012). This acceptance ends up distorting the explicitly stated aims of the evaluation process. The role of scientific societies and, above all, their awareness of the importance of preserving the ethical and cultural foundations of disciplines in such a landscape, also emerges as an aspect of primary interest and one that will be interesting to observe in the years to come. A fruitful example is that of the economic-business disciplines, characterised by a growing role of scientific societies that have for some time, and more incisively in recent periods, set up coordinated activities aimed at monitoring and improving journal rankings and raising researcher awareness of the need to adhere to ethical principles, preserving the freedom of research but also the cultural roots to which it belongs.

Lastly, although exhaustive in terms of theoretical saturation, this work may suffer from a limitation in terms of the chosen perimeter of analysis within an intrinsically complex phenomenon, as well as the focus on a non-bibliometric area. Future research should further address the issue of research assessment in a manner not limited to the periodic application of VQR but reconciling the emerging logics deriving from the academic system of self-evaluation and accreditation that is gradually expanding to research and knowledge exchange beyond the traditional teaching and education area.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

The authors are deeply grateful to all those who were willing to offer their time by granting interviews. Thanks are also due to the anonymous referees who provided constructive and valuable suggestions. Special thanks go to the guest editors for their continued support and professional guidance during the complex review process. However, responsibility for the content of the article remains solely with the authors.

References

- Abramo, G., & D'Angelo, C. A. (2015). VQR, Italy's second national research assessment: Methodological failures and ranking biases. *Journal of the Association for Information Science and Technology*, 66(11), 2202–2214.
- Adams, G. B. (2011). The problem of administrative evil in a culture of technical rationality. *Public Integrity*, 13(3), 275–285.
- Adams, G. B., & Balfour, D. L. (1998). *Unmasking Administrative Evil*. New York, NY: Sharpe.
- Agyemang, G., & Broadbent, J. (2015). Management control systems and research management in universities: An empirical and conceptual exploration. *Accounting, Auditing and Accountability Journal*, 29(7), 1018–1046.
- Ahrens, T., & Mollona, M. (2007). Organisational control as cultural practice—A shop floor ethnography of a Sheffield steel mill. *Accounting, Organizations and Society*, 32(4–5), 305–331.
- Alvesson, M. (2003). Methodology for close up studies - Struggling with closeness and closure. *Higher Education*, 46(2), 167–193.
- Antonelli, V., D'Alessio, R., Rossi, R., & Funnell, W. (2018). Accounting and the banality of evil: Expropriation of Jewish property in Fascist Italy (1939–1945). *Accounting, Auditing & Accountability Journal*, 31(8), 2165–2191.
- Arendt, H. (1964). *Eichmann in Jerusalem, A Report on the Banality of Evil*. New York, NY: The Viking Press.
- Argento, D., Dobija, D., & Grossi, G. (2020). The disillusionment of calculative practices in academia. *Qualitative Research in Accounting & Management*, 17(1), 1–17.
- Barnabè, F., & Riccaboni, A. (2007). Which role for performance measurement systems in higher education? Focus on quality assurance in Italy. *Studies in Educational Evaluation*, 33(3–4), 302–319.
- Bauman, Z. (1989). *Modernity and the Holocaust*. Ithaca, NY: Cornell University Press.
- Baumeister, R. F. (1997). Esteem threat, self-regulatory breakdown, and emotional distress as factors in self-defeating behaviour. *Review of General Psychology*, 1(2), 145–174.
- Becker, A., & Lukka, K. (2023). Instrumentalism and the publish-or-perish regime. *Critical Perspectives on Accounting*, 94, Article 102436. <https://doi.org/10.1016/j.cpa.2022.102436>
- Beime, K. S., Englund, H., Gerdin, J., & Seger, K. (2023). Theorizing the subjectivizing powers of market-based technologies: Looking beyond coercion and seduction. *Critical Perspectives on Accounting*, Article 102662. <https://doi.org/10.1016/j.cpa.2023.102662>
- Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, 44(2), 451–466.
- Bracci, E. (2023). The loopholes of algorithmic public services: an “intelligent” accountability research agenda. *Accounting, Auditing & Accountability Journal*, 36(2), 739–763. <https://doi.org/10.1108/AAAJ-06-2022-5856>
- Coffey, A. (1999). *The Ethnographic Self: Fieldwork and the Representation of Identity*. Thousand Oaks, CA: Sage.
- Courtois, C., Plante, M., & Lajoie, P.-L. (2020). Performance in neo-liberal doctorates: The making of academics. *Qualitative Research in Accounting & Management*, 17(3), 465–494.
- Davie, S. K. (2000). Accounting for imperialism: A case of British-imposed indigenous collaboration. *Accounting, Auditing & Accountability Journal*, 13(3), 330–359.
- Davie, S. K. (2008). An autoethnography of accounting knowledge production: Serendipitous and fortuitous choices for understanding our social world. *Critical Perspectives on Accounting*, 19(7), 1054–1079.
- Dillard, J. F. (2003). Professional services, IBM, and the Holocaust. *Journal of Information Systems*, 17(2), 1–16.
- Dillard, J. F., & Ruchala, L. (2005). The rules are no game: From instrumental rationality to administrative evil. *Accounting, Auditing & Accountability Journal*, 18(5), 608–630.
- Dobija, D., Górska, A. M., Grossi, G., & Strzelczyk, W. (2019). Rational and symbolic uses of performance measurement: Experiences from Polish universities. *Accounting, Auditing & Accountability Journal*, 32(3), 750–781.
- Edgar, F., & Geare, A. (2013). Factors influencing university research performance. *Studies in Higher Education*, 38(5), 774–792.
- Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: an overview. *Historical social research/Historische sozialforschung*, 36(4), 273–290.
- Esposito, V., De Nito, E., Pezzillo Iacono, M., & Silvestri, L. (2013). Dealing with knowledge in the Italian public universities: The role of performance management systems. *Journal of Intellectual Capital*, 14(3), 431–450.
- Field, L. (2015). Appraising academic appraisal in the new public management university. *Journal of Higher Education Policy and Management*, 37(2), 172–189.
- Funnell, W. (1998). Accounting in the service of the Holocaust. *Critical Perspectives on Accounting*, 9(4), 435–464.
- Gebreiter, F. (2022). A profession in peril? University corporatization, performance measurement and the sustainability of accounting academia. *Critical Perspectives on Accounting*, 87(1–14), Article 102292. <https://doi.org/10.1016/j.cpa.2021.102292>
- Gendron, Y. (2008). Constituting the academic performer: the spectre of superficiality and stagnation in academia. *European Accounting Review*, 17(1), 97–127.
- Gendron, Y. (2015). Accounting academia and the threat of the paying-off mentality. *Critical Perspectives on Accounting*, 26, 168–176.
- Gendron, Y., Andrew, J., & Cooper, C. (2022). The perils of artificial intelligence in academic publishing. *Critical Perspectives on Accounting*, 87(1–12), Article 102411. <https://doi.org/10.1016/j.cpa.2021.102411>
- Gendron, Y., & Rodrigue, M. (2021). On the centrality of peripheral research and the dangers of tight boundary gatekeeping. *Critical Perspectives on Accounting*, 76, Article 102076.
- Grossi, G., Kallio, K. M., Sargiacomo, M., & Skoog, M. (2020). Accounting, performance management systems and accountability changes in knowledge-intensive public organizations: A literature review and research agenda. *Accounting, Auditing & Accountability Journal*, 33(1), 256–280.
- Guarini, E., Magli, F., & Francesconi, A. (2020). Academic logics in changing performance measurement systems: an exploration in a university setting. *Qualitative Research in Accounting & Management*, 17(1), 109–142.
- Guthrie, J., & Neumann, R. (2007). Economic and non-financial performance indicators in universities: The establishment of a performance-driven system for Australian higher education. *Public Management Review*, 9(2), 231–252.
- Harding, N., Ford, J., & Gough, B. (2010). Accounting for ourselves: Are academics exploited workers? *Critical Perspectives on Accounting*, 21(2), 159–168.
- Härström, M. (2023). Logics in situations of lower or higher ‘perceived control exposure’: the case of researchers. *Journal of Public Budgeting, Accounting & Financial Management*, 35(6), 78–99.
- Harley, S. (2000). Accountants divided: Research selectivity and academic accounting labour in UK universities. *Critical Perspectives on Accounting*, 11(5), 549–582.
- Haynes, K. (2006). Linking narrative and identity construction: using autobiography in accounting research. *Critical Perspectives on Accounting*, 17(4), 399–418.
- Haynes, K. (2011). Tensions in (re)presenting the self in reflexive autoethnographical research. *Qualitative Research in Organizations and Management*, 6(2), 134–149.
- Hickey, A., & Smith, C. (2020). Working the aporia: ethnography, embodiment and the ethnographic self. *Qualitative Research*, 20(6), 819–836.
- Hoffman, M. C., Pyne, J., & Gajewski, M. (2012). The ten masks of administrative evil. *Administrative Theory & Praxis*, 34(1), 125–132.
- Hopwood, A. (1984). Accounting and the Pursuit of Efficiency. In Hopwood and Tomkins (Eds.), *Issues in Public Sector Accounting*. London: Philip Allan.
- Humphrey, C., & Gendron, Y. (2015). What is going on? The sustainability of accounting academia. *Critical Perspectives on Accounting*, 26, 47–66.

- Kallio, K. M., & Kallio, T. J. (2014). Management-by-results and performance measurement in universities-implications for work motivation. *Studies in Higher Education*, 39(4), 574–589.
- Kallio, K. M., Kallio, T. J., & Grossi, G. (2017). Performance measurement in universities: Ambiguities in the use of quality versus quantity in performance indicators. *Public Money & Management*, 37(4), 293–300.
- Lapsley, I., & Miller, P. (2004). Foreword: Transforming universities: The uncertain, erratic path. *Financial Accountability & Management*, 20(2), 103–106.
- Malsch, B., & Tessier, S. (2015). Journal ranking effects on junior academics: Identity fragmentation and politicisation. *Critical Perspectives on Accounting*, 26, 84–98.
- Manes Rossi, F., Spanò, R., Martin-Sardesai, A., & Guthrie, J. (2022). Performance management change: discourses at play in an Italian university. *Journal of Public Budgeting, Accounting & Financial Management*, 34(4), 534–556. <https://doi.org/10.1108/JPBAFM-11-2021-0156>
- Maran, L., Bigoni, M., & Morrison, L. (2023). Shedding light on alternative interdisciplinary accounting research through journal editors' perspectives and an analysis of recent publications. *Critical Perspectives on Accounting*, 93, Article 102420. <https://doi.org/10.1016/j.cpa.2022.102420>
- Marques, M., Powell, J. J. W., Zapp, M., & Biesta, G. (2017). How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986–2014. *European Educational Research Journal*, 16(6), 820–842. <https://doi.org/10.1177/1474904117730159>
- Martin-Sardesai, A., Guthrie, J., & Parker, L. (2021). The neoliberal reality of higher education in Australia: how accountingisation is corporatising knowledge. *Meditari Accountancy Research*, 29(6), 1261–1282.
- Martin-Sardesai, A., Guthrie, J., & Tucker, B. P. (2020). What you see depends on where you look: performance measurement of Australian accounting academics. *Accounting, Auditing & Accountability Journal*, 33(6), 1193–1218.
- Mazza, C., Quattrone, P., & Riccaboni, A. (2008). Found in translation: On the persistence of the university as institution. In *European Universities in Transition: Issues, Models and Cases* (pp. 1–16). Edward Elgar Publishing Ltd.
- Michelon, G. (2021). Accounting research boundaries, multiple centres and academic empathy. *Critical Perspectives on Accounting*, 76(1–9), Article 102204. <https://doi.org/10.1016/j.cpa.2020.102204>
- Neu, D. (2001). Banal accounts: subaltern voices. *Accounting Forum*, 25(4), 319–333.
- Nørreklit, L., Jack, L., & Nørreklit, H. (2019). Moving towards digital governance of university scholars: Instigating a post-truth university culture. *Journal of Management and Governance*, 23(4), 869–899.
- Paolini, A., & Quagli, A. (2013). A reflection on bibliometric tools for research evaluation and a proposal: the real impact factor. *Management Control*, 3, 115–128.
- Parker, L. D., & Guthrie, J. (2012). Accounting scholars and journals rating and benchmarking - Risking academic research quality. *Accounting, Auditing and Accountability Journal*, 26(1), 4–15.
- Parker, L., Martin-Sardesai, A., & Guthrie, J. (2023). The commercialised Australian public university: An accountable transition. *Financial Accountability & Management*, 39, 125–150.
- Picard, C. F., Durocher, S., & Gendron, Y. (2019). Desingularization and de-skilling: A foray into ranking production and utilisation processes. *European Accounting Review*, 28(4), 737–765.
- Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting & Management*, 8(3), 238–264.
- Raineri, N. (2015). Business doctoral education as a liminal period of transition: Comparing theory and practice. *Critical Perspectives on Accounting*, 26, 99–107.
- Rebora, G., & Turri, M. (2010). The development of research evaluation systems: a critical analysis of the Italian experience. Paper presented at the AIDEA Conference. Available at: https://gianfrancorebora.org/wp-content/uploads/2010/10/rebora-turri_52.pdf.
- Rebora, G., & Turri, M. (2011). Critical factors in the use of evaluation in Italian universities. *Higher education*, 61(5), 531–544.
- Rintamäki, J., & Alvesson, M. (2023). Resisting whilst complying? A case study of a power struggle in a business school. *Academy of Management Learning & Education*, 22(2), 257–273. <https://doi.org/10.5465/amle.2020.0070>
- Rosenberg, A. (1983). The philosophical implications of the Holocaust. In R. L. Brahm (Ed.), *Perspectives on the Holocaust* (pp. 1–18). Holocaust Studies Series, Dordrecht: Springer.
- Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52(4), 1893–1907.
- Sargiacomo, M. (2003). Best practices in doctoral studies at the G. d'Annunzio University. *Public Enterprise*, 16(5–6), 495–525.
- Seger, K., Englund, H., & Härström, M. (2023). Researchers' hate-love relationship to performance measurement systems in academia-a Foucauldian perspective. *Qualitative Research in Accounting & Management*, 20(1), 38–71.
- Steccolini, I. (2023). What counts as 'good' qualitative accounting research? Researchers' perspectives on assessing and proving research quality. *Accounting, Auditing & Accountability Journal*, 36(3), 1032–1057. <https://doi.org/10.1108/AAAJ-05-2022-5808>
- ter Bogt, H. J., & Scapens, R. W. (2012). Performance management in universities: Effects of the transition to more quantitative measurement systems. *European Accounting Review*, 21(3), 451–497.
- van Helden, J., & Argento, D. (2020). New development: Our hate-love relationship with publication metrics. *Public Money & Management*, 40(2), 174–177.
- Willmott, H. (2003). Commercialising higher education in the UK: The state, industry and peer review. *Studies in Higher Education*, 28(2), 129–141.