ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Evaluating Research and Scholarly Impact in Criminology and Criminal Justice in the United Kingdom and Italy: A Comparative Perspective

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Lavorgna Anita, Ugwudike Pamela, Vianello Francesca (2023). Evaluating Research and Scholarly Impact in Criminology and Criminal Justice in the United Kingdom and Italy: A Comparative Perspective. JOURNAL OF CONTEMPORARY CRIMINAL JUSTICE, 39(3), 354-370 [10.1177/10439862231170966].

Availability:
This version is available at: https://hdl.handle.net/11585/924559 since: 2023-05-02

Published:
DOI: http://doi.org/10.1177/10439862231170966

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

14 October 2026

This is the final peer-reviewed accepted manuscript of:

Anita Lavorgna, Pamela Ugwudike, Francesca Vianello (2023): Evaluating Research and Scholarly Impact in Criminology and Criminal Justice in the United Kingdom and Italy: A Comparative Perspective, Journal of Contemporary Criminal Justice, 39 (3): 354-370

The final published version is available online at:

https://doi.org/10.1177/10439862231170966

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it)

When citing, please refer to the publisher version.

Evaluating research and scholarly impact in Criminology and Criminal Justice the UK and Italy: A comparative perspective

Anita Lavorgna (University of Bologna, Italy and University of Southampton, UK), Pamela Ugwudike (University of Southampton, UK) and Francesca Vianello (University of Padova, Italy)

1. Introduction

In a research context characterised by scarce resources, especially over the last couple of decades, several ways to assess research quality, particularly to measure scholarly productivity and the scientific impact of academic institutions, funding agencies and other parties, have been proposed. Key measures include citation analysis, at times supplemented by other means to try and assess readership data, and each measure has its merits and shortcomings (e.g., Moed, 2005; Bollen et al., 2005; Rowlands & Nicholas, 2007; Priem & Hemminger, 2010; Schloegl & Gorraiz, 2011; Mohammadi & Thelwall, 2014). As we will see, the measures are not neutral, but they are – at least implicitly – aligned with a certain approach to framing and understanding academic research, and have both direct and indirect repercussions for how certain research endeavours are prioritised or even implemented.

This contribution focuses on the two main nation-wide systems used in the United Kingdom (UK) and Italy to measure or evaluate research and assess scholarly impact beyond academia. Both countries have similar overall and student population sizes (Rebora & Turri, 2013) and the systems used to assess impact are the Research Excellence Framework (hereafter, REF) in the UK and the Evaluation of Research Quality (*Valutazione della Qualità della Ricerca*, hereafter, VQR) in Italy. After presenting a descriptive overview of both systems, this article will draw on the relevant literature to critically discuss some systemic limitations affecting them. The chapter will discuss the specific deficiencies of

the current systems, with a focus on their application in the field of criminology and criminal justice (hereafter, CCJ).

Before moving to the core of this contribution, a brief definitory premise is necessary. CCJ is a broad field of teaching and research. Its definition and disciplinary divisions vary across diverse countries according to their unique theoretical traditions and linguistic preferences. In the UK, for instance, CCJ is more clearly detectable as a stand-alone discipline in research publications and teaching programmes (for a relatively recent overview, see Harris et al., 2019). It is mostly recognised as an interdisciplinary and evolving field that brings together theories, principles, and methods from law, psychology, economics as well as sociology and other social sciences. It finds its core in its focus on the aetiology of crime and deviance, and the operation of criminal justice systems (e.g., Downes, 1988; Garland & Sparks, 2000; Zedner, 2007; Brisman et al., 2018). Over time, the boundaries of the discipline have been stretched, to encompass relative new areas such as the study of social harms. The discipline has also become more open to inputs from disciplines from outside the social sciences (consider, for instance, the computer and data sciences), giving rise to the nascent field of digital criminology. In Italy, on the other hand, CCJ stems from legal and psychological traditions. At the time of writing, CCJ in most universities tends to be academically split between a range of other disciplines including 'sociology of law, deviance and social changes', legal medicine, criminal law, philosophy of law, and psychology. This more sectorial approach possibly stems from a traditional understanding of CCJ as skewed towards acritical/legalistic stances (e.g., Gridelli Veligogna, 1986, pp. 83-84; Baratta, 2019, pp. 203ff), and at times critical of the inherently interdisciplinary nature of criminology, which has been described as a 'hybrid discipline' (Prina, 2020, p. 16). As we will see later, this apparently

negligible difference in how boundaries are set can have direct implications for the evaluation of research in CCJ in both countries[1].

2. A descriptive overview of the two systems

2.1. The Research Excellence Framework (UK)

Quality-related (QR) funding in the UK is considered to be 'the invisible force holding up the UK's research ecosystem' (Russel Group, 2019). It has four main components (mainstream QR; research degree programme supervision; business research; and charity support). The mainstream QR component is the main one; it relies on a 'dual support system' for university research, with money coming from both institutional funding, mostly from the REF (which the universities can spend as they prefer), and competitively won, project-based external funding (Arnold et al., 2018).

The REF is the current iteration of the UK-wide performance-based research assessment system for UK Higher Education providers. It has been held every few years since 1986, becoming increasingly more complex, demanding, and costly (Stern et al., 2016; Wildson & Weinstein, 2018). In 2014, the REF replaced the Research Assessment Exercise. It is generally perceived as one of the world's most comprehensive research assessments and considered an example by many countries of how to review research and assess its impact (Stern et al., 2016; Else, 2022). REF outcomes influence how public

[1] It is worth noting that existing CCJ literature has already dealt with research evaluation, but mostly in terms of scholarly productivity and academic 'influence' or 'prestige', and with a strong US focus (with a few exceptions, e.g., Moeller, 2019). Consider, for instance, many publications in the *Journal of Criminal Justice Education* such as Cohn & Farrington, 2007, 2014; Walker & Raptopoulos, 2008; Kleck & Barnes, 2011; Long et al., 2011; Copes et al., 2012).

funds are allocated to universities for research (UKRI, 2022). For the sake of this contribution, we will only focus on the last REF exercise (REF, 2021), and occasionally on the previous one (REF, 2014). It is expected that the next REF will be held in 2028 although this is yet to be confirmed by Research England which oversees the REF.

As we have seen, the REF informs UK research, being at the basis of its funding (Arnold et al., 2018). UKRI, a key REF stakeholder, estimates that REF outcomes inform the distribution of approximately £2 billion of public funding annually across universities for research. As such, the stakes in the REF game are very high. Overall, the distribution of REF funding has been relatively static over time (with about 50% of funds going to the top 10 Universities, and about 70% going to the larger Russell Group of research-intensive universities). But even within this trend there can be 'wins' and 'losses' for individual universities (Arnold et al., 2018). For example, the considerable resources invested in preparing REF submissions are estimated to be a few million pounds per institution. Nevertheless, the REF is considered to be, financially, a very good return on investment by administrators, particularly for research-oriented institutions.

Apart from the (potential) financial gain, there are a number of other declared benefits of participating in the REF. These include the potential benefits to broader society, institutions and individual researchers alike. In this respect, the REF is considered by some to be 'an essential tool for accountability and information' (e.g., Stern et al., 2016, p. 8) as it is expected to provide accountability for public funding of research, evidence its public benefits by explicitly discussing research impact, and provide quality benchmarks across the sector. Additionally, as the REF and Open Access UK Research and Innovation (UKRI) have introduced a policy requiring all published articles and proceedings submitted to the REF to be made open access, the REF could indirectly increase opportunities for the findings of publicly funded research to be accessed, shared and reused. The REF could also be a driver for research quality, and could raise the profile and reputation of research activity within the UK in

general and also within research institutions. Moreover, institutions can use REF results to better acknowledge and understand their strengths and weaknesses, and improve their reputation in the 'academic market' to attract the best researchers. Last but not least, the REF could benefit individual researchers by providing rewards for strong performance (Stern et al., 2016). From this perspective, the mandatory inclusion of ALL staff (>0.2 FTE[2]) with significant responsibility for research aims to make REF more inclusive (as, for instance, the type of job title or contact does not matter). This is a change introduced for the latest REF. Previously, institutions were permitted to put forward only top performers but this tended to skew ratings in their favour. With the change, the last REF saw an increase of 46% in the number of staff submitted for assessment compared with the previous one (Else, 2022).

In practice, the REF is based on a process of peer review undertaken by expert panels, comprising (national) senior academics, international members and research users. Each panel focuses on one of 34 units of assessment (UOAs). The UOAs are subject-based, cutting across different disciplines, with institutions able to make submissions to as few or as many UOAs as they wish. The UOAs are arranged into 4 Main Panels, which have slightly different rules for submission. CCJ does not have a specific UOA, but it is generally submitted under 'Social Work and Social Policy' or 'Sociology', all belonging to the Main Panel C (which includes, among other fields, the social sciences, law, and geography). Overall, the panels are broad and multidisciplinary, and they can refer research outputs to other panels, for example, to handle interdisciplinary work (Arnold et al., 2018).

[2] FTE (Full Time Equivalent) is a common unit of measurement equivalent to a worker one unit of a work day, which is generally used to 'measure' an employee and their workload, and to determine how many full-time employees are on staff. The FTE is calculated by adding the hours paid to all employees, and dividing that by the number of hours a full-time employee would work over a given period.

In this peer review process, three distinct elements are assessed:

(1) *The quality of research outputs.* Assessment of research is based on five quality levels, each providing a certain number of 'stars'. As such, research can be considered word leading (4*); internationally excellent (3*); recognized internationally (2*); recognized nationally (2*); or unclassified, or falling behind standards of nationally recognised work (U). Only research assessed as 3* and 4* is funded, with funding allocation based on the volume of activity assessed as 4* and 3* at a ratio of 4:1.

(2) *Research impact*. Since REF 2014, Impact Case Studies have been introduced – that is, short reports (4 pages each) describing how a specific piece of research has brought positive societal change. Impact is defined by Research England as 'an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia'. The number of case studies to be submitted by each institution is limited, and depends on the number of staff submitted to each unit of assessment.

(3) *The environment that supports research*. Environment statement – consists of between 8,000 and 20,000 words depending on the number of staff returned.

REF funding, at the time of writing, depends on the following: 65% on research outputs; 20% on impact (an increase of 5% from the previous REF); and 15% on the research environment. These pots are distributed between subjects (taking into account, for instance, a cost-weighting for different subjects depending on whether they are lab-based or library-based; and the volume on research). The distribution of REF funding is also based on the number of submitted research-active staff, and it is distributed to Higher Education institutions. UOA profiles can be aggregated to produce overall

profiles for the University, weighted by the number of staff returned in each UOA. The scores for individual outputs or researchers are not published.

2.2. The Evaluation of Research Quality (Italy)

Founded in 2006, the Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR) manages the evaluation of the quality of the outcomes of the research of the public universities and research institutes, as well as those of private institutions that voluntarily submit their research outcomes for evaluation. The VQR process, which was mandated by the Presidential Decree no. 76/2010, is overseen by the Ministry for Education, Universities and Research (MIUR), and takes place approximately every five years. The first one took place in 2011 and considered research outputs from the period 2004-2010 while the last one took place in 2021 and considered the period 2015-2019) (see ANVUR, 2022a; see Bonaccorsi, 2020 for a recent overview of all the VQR exercises that have taken place). In the last VQR, a total of 98 Universities, 14 public research centres, and 22 private institutions participated in the process.

Similar to the REF, a number of reported benefits are ascribed to the VQR system. In a context of scarce resources, there is an emphasis on its role in monitoring how: public money is invested in research; public funding is allocated (both ordinary research budget and reward budget); and public evidence is produced to the taxpayer to the benefit of scientific research (Buoso, 2021). To borrow the ANVUR's words, 'the VQR exercises provide an up-to-date assessment of the state of research in the various scientific fields, in order to promote the improvement of research quality in the assessed institutions and to allocate the performance-based share of the *Fondo di Finanziamento Ordinario* (FFO) – the Ordinary Financing Fund for the Italian University system' (ANVUR, 2022a). There is also an expectation that the introduction of evaluation practices would stimulate research efficiently. Indeed, there is some evidence that the lowest performers in the system achieved the greatest

improvement (Abramo & D'Angelo, 2021). Additionally, a performance-based criterion for funding should foster greater international collaboration and the rate of internationalization of Italian research has been increasing but whether or not that depends on the VQR or is rather part of a more general, global phenomenon remains unclear – see Abramo & D'Angelo, 2022.

Further, the VQR performance evaluation is based on the evaluation of a selected subset of products. But it adopts a more complex, hybrid evaluation approach. It brings together bibliometric approaches, As will be discussed in more detail below, for many disciplines including psychology, citation and journal indexes are also considered as are peer review systems (mostly relevant for the legal and the social sciences, and hence for a majority of CCJ-adjacent sectors in Italy) (see also Abramo & D'Angelo, 2015a; Franceschini & Maisiano, 2017). The procedure is described by ANVUR (2022b) as based on 'informed peer review' – i.e., the process is mostly based on the work of peer reviewers, who can use bibliometric information to form their opinion when appropriate.

VQR evaluation is split in practice across 18 Expert Groups (*Gruppi di Esperti della Valutazione*, or GEV) representing disciplinary macro-areas formed by approximately 630 researchers based both in Italy and abroad, and supported by about 11,000 external evaluators and 24 administrative assistants. Hence, while in the REF assessment is directly performed by the field panel components, with the VQR external reviewers are recruited by the disciplinary panels (Rebora & Turri, 2013). CCJ-related subjects, as mentioned above, can be linked to different disciplinary 'areas' in this context, with 'political and social sciences' being probably the main one. One of the 18 GEV (formed, in this case, by professors with specific expertise alongside representatives of the industrial, financial and cultural arenas) focuses on impact and 'third mission' activities. These activities are broadly defined as the 'opening towards the socio-economic context through the valorization and the transfer of skills' (ANVUR, 2022c), or even more broadly as those outward-facing activities that emphasise the role of research institutions to promote the social, economic and cultural development of the country (ANVUR,

2022d). This broad field encompasses a wide range of outreach activities (e.g., public engagement, the creation of educational material) as well as entrepreneurial ones.

In the VQR procedure, research outputs (publications) and case studies are considered. The research outputs are assessed according to their, (1) *originality* (i.e., research must add knowledge to the state of the art), (2) *relevance* (refers to scientific impact, in a purely scientific-scholarly perspective – as discussed in Bonaccorsi, 2020), and (3) *methodological rigor* (previously internationalization. But the latter criterion has been eliminated as it is considered problematic for certain disciplines where publications in Italian are common – see Bonaccorsi, 2020). In the last VQR, the number of scientific outputs presented increased to 182.000, for a total of about 65.000 researchers. The number of outputs to be presented is not set, but there is a certain flexibility in the sense that each Institution should present a number of outputs equal to three times the number of their researchers – with a maximum of four outputs to be attributed to each individual researcher. The case studies are assessed by looking (1) at the social, economic and cultural dimension of their impact; (2) their relevance with regard to their context of reference; (3) the added value they could bring to the beneficiaries, and (4) the contribution of the proposing institution.

Each research output and each case study can be awarded one category and score: (a) excellent and extremely relevant, 1 point; (b) excellent, 0.8 points; (c) standard, 0.5 points; (d) sufficient relevance, 0.2 points; and (e) scarce relevance or not acceptable, 0 points. A number of equations are then used to determine a number of quantitative indicators such as: the total points obtained per reference area ('total score'); the average quality of the outputs per reference area within each institution ('average score'); the relative quality per institution, normalised according to the different scientific areas ('qualitative indicator'); and the relative quality per institution, normalised according to the number of outputs and case studies presented by each institution ('quali-quantitative indicator'). A combination of these indicators is then used to determine the reward funding to be allocated to the

best-scoring institutions. It is worth noting that, in Italy, *all* professors are required to carry out both research and teaching (there are no 'teaching-only' universities). As such, even if the VQR is oriented towards the assessment of institutions (the use of VQR indicators for the evaluation of individual researchers is inhibited), the process is based on the involvement of all academic staff (Bonaccorsi, 2020).

Table 1: Indicators used in research assessment exercises

		UK (REF 2021)	ITALY (VQR 2021)
Output indicators	Academic outputs	x	x
	Non-academic outputs	x	x
	Innovation-related outputs (IPR)	x	x
External funding indicators	National competitive funding	x	x
	International competitive funding	x	x
	Contract research funding	x	x
	Non-competitive funding	x	x
Systemic indicators	Esteem (conferences, editorships, rewards, …)	x	x
	National collaborations	x	x
	International collaborations	x	x
	International mobility	x	x
	Research-industry collaborations	x	x
	PHD-recruitment / awarding	x	x
Outcomes / impact indicators	Academic impact (citations)	x	Missing in the UK

	Socio-economic outcomes / impacts (e.g., spin-offs)	x	x

Table adapted from Stern et al., 2016, p. 47

3. A critical view

3.1. Systemic limitations

Overall, the general view of research administrators, as exemplified by the independent Reviews available, is that that 'the REF is useful' and its broad structure should not change, as 'the processes used to assess research excellence […] have delivered well in their objective to improve quality' (Stern et al., 2016: 18). However, it is also recognised that the system suffers from severe limitations. Similarly, a number of shortcomings have been identified in the VQR procedure. Specifically, the small number of papers considered, the incorrect use of journal metrics, and the use of conceptually misleading criteria for normalizing bibliometric indicators, have all been identified as the major vulnerabilities of this approach (Franceschini & Maisano, 2017). Indeed, despite the previously mentioned assumptions about the capacity of the REF and VQR to drive research quality and raise both the profile and reputation of research activity across research institutions and beyond, some contend that there is limited evidence to suggest that the performance-based systems are associated with the increased production and quality of scientific outputs (see for example, Arnold et al., 2018).

On the contrary, it has been argued that the REF may have some negative effects including the risk that some institutions may (1) focus more on research that can produce quick, measurable results; (2) impose a heavy financial burden on institutions as they invest considerable resources in assembling REF submissions; (3) marginalise staff considered unable to submit outputs assessed as 3* and above, in some cases jeopardizing their career progression; and (4) foster a counterproductive 'publish or

perish' culture that perversely incentivises staff to focus on the quantity rather than the quality of outputs (Sayer, 2014; Watermeyer & Derrick, 2022).

What the foregoing insights from critical perspectives suggest is that the REF might have negative effects on the nature of the research done. For example, it could affect how researchers select research topics or methodological approaches as research outputs are expected to appear in high-status, indexed journals. The critical perspectives also suggest that it could affect the amount of interdisciplinary and transformative (i.e., of high-risk research challenging established theory or method) research done, as university research managers are less likely to submit interdisciplinary outputs to the REF. Similarly, research managers seem to prefer low-risk research that produces outputs in the short term, or outputs published in mainstream journals that often tend to drive out unconventional work, suggesting that the REF might negatively impact the attitudes and behaviours of both research managers and researchers (Arnold et al., 2018). Second, the REF has become very costly (Stern et al., 2016): consider, for instance, that the 2014 REF cost approximately £250 million. It is also considered 'disruptive and time-consuming' (Mryglod et al., 2014).

Third, the critics contend that the REF could negatively affect researchers' careers, which extends to the recruitment process (as the search is for academics with a REFable research profile). As such, other academics, such as those who seek more flexible career structures or work part-time, might suffer negative consequences (Stern et al. 2016). Hence, unsurprisingly, while comments on the REF coming from research managers tend to be positive, comments from individual academics, and especially from younger researchers and women, are generally negative (Arnold et al., 2018). Similarly, the REF might further marginalise certain groups including minorities as structural issues around discriminatory practices might still permeate REF selection and assessment processes (e.g., REF, 2018; Yarrow, 2018; Verma, 2020; AdvanceHE, 2021). There have been calls to recognise with greater rigour

the diverse set of perspectives and experiences different researchers have on the REF (Wildson & Weinstein, 2018), but no major initiatives in this direction have been taken yet.

Fourth, there is also scope for 'gaming' the system, as Higher Education institutions can use tactics to maximise REF performance – e.g., short-termism in publication strategies, going for 'safer' publications (Stern et al., 2016) – that might have negative effects in the long run (Mryglod et al., 2014; Stern et al., 2016, p. 12). Although researchers currently submit case studies – a measure to limit institutions' attempts to game the process, and to try and widen the approach to research assessment – the REF is still considered by some to be problematic in the sense that it fails to properly recognise, promote and reward a positive research culture. Consider, for instance, how negative results or replication studies, which are fundamental for the research process, are deemed unworthy of REF submission, hence disincentivising researchers from pursuing them (Else, 2022).

These systemic problems associated with the REF have led to complaints which have on occasion spiralled into threats of institutional boycotts (UCU, 2013). The complaints and threats predominantly hinge on the unreasonable expectations concerning the required number of outputs and a perceived lack of transparency and fairness in selection processes.

Again, similar to the REF, the VQR system has several shortcomings (see Franceschini & Maisano, 2017, above) and negative spillovers. These can directly affect the research context, creating a culture of risk avoidance that disincentivises innovative, unconventional, unpredictable or interdisciplinary research, while encouraging a 'publish or perish' culture that might reward quantity at the expenses of quality (Buouso, 2021). Good practices of academic citizenship can also be affected, as individuals (or, at least, those who have a choice) could avoid unrewarding activities – regardless of the fact that 'research' is generally considered just one dimension of academic life, together with 'teaching' and 'service' (Walker & Raptopoulos, 2008). Moreover, as achieving a good score on the indicators becomes a

target in itself (Buoso, 2021), the system can be rigged, with implications for research ethics. Rigging could for instance, include promoting mutual citation practices, not citing the work of 'competitors', or relying on 'salami publishing'. Additionally, in the VQR system, the financial costs of the evaluation itself and the time resources it demands (by increasing bureaucratic burdens) cannot be ignored (Buouso, 2021).

Last but not least, as recently emphasized by Buoso (2021), the VQR is worsened by additional factors endemic in the Italian system. Examples include the National Scientific Habilitation (ASN) and the connected rating system of scientific journals. While these measures have the appreciable aim to prevent some of the inefficiencies and misbehaviours that have been afflicting the Italian research system, their development and implementation are proving to be problematic. The ASN – which also occurs under ANVUR's aegis – is a prerequisite for applying for permanent positions (full and associate professorship) in Italian universities. According to this system, depending on the discipline, the candidate needs to reach certain publication thresholds. They should also be considered a proper 'fit' with the disciplinary area considered (a qualitative assessment carried out by a disciplinary panel). The indicators used for so-called 'non-bibliometric' disciplines (e.g., the social sciences) are: (1) the number of monographs (with ISBN); (2) the number of articles published in a set of journals indicated by ANVUR, or as part of monographs with ISBN; (3) the number of articles in top-ranked journals, as selected by ANVUR. For 'bibliometric' disciplines, the indicators consider, beyond the number of articles published, the number of citations received for the overall scientific production, normalized for the academic age of the candidate, and the contemporary h-index of the overall scientific production – as such, using these indicators as proxies for the value of a researcher's scientific activity (Abramo & D'Angelo, 2015b). As research outputs are strictly assessed depending on the discipline in which the candidate tries to qualify, because of the sectorial approach found in many disciplines relating to the social sciences in Italy (including CCJ), the choice of research topics and approaches by individual researchers is constrained, and interdisciplinary endeavours end up being penalized. This

assessment approach also affects the format and the venue of the publication, and can obviously affect editorial choices for those (Italian) journals with closer links to the VQR (as discussed also in Buouso, 2021).

3.2. Disciplinary issues

With respect to CCJ specifically, research evaluation exercises such as the REF and the VQR suffer from a number of limitations. Some are common with other disciplines in the social sciences and the humanities. For instance, as emerges from the metareview carried out by Huang and Chang (2008), there are several inherent differences among the disciplines traditionally referring to these macro-fields and those traditionally linked to the natural sciences, that go beyond the generally recognised differences in subjects and epistemologies, and rather impact the practicalities of output production and assessment. When it comes to output production particularly, while the natural sciences (and the sectors of the social sciences and the humanities more aligned with the natural sciences) tend to publish more articles and fewer books, in the social sciences and the humanities book publishing is still often considered a more appropriate form for presenting scholarship, and the amount of journal articles published is generally lower (a practice that is changing, possibly as a result of research evaluation mechanisms – see Savage & Olejniczak, 2022). Similarly, the citation studies reviewed by Huang and Chang (2008) demonstrate the significant role of books as reference sources in the social sciences and the humanities. The studies also highlight an important trend: on average, researchers in the social sciences and the humanities tend to cite older literature than do researchers in the natural sciences. A consequence of this is that their research outputs tend to generate impact after several years, while showing a low citation count in the immediate years after publication (see also Sivertsen, 2022).

Impact in the social sciences can be too indirect, non-linear, and diffuse to be fully captured by a REF case study, or by the VQR 'third mission' (consider, for instance, conceptual impacts – see Muhonen & Tellmann, 2022). Building 'impact' in the social sciences often depends on a number of strategies that have to be embedded in the research design to work. But the strategies are only occasionally promoted by Universities, which tend to focus on supporting areas that can bring more direct commercial benefits. Consider for instance, when it comes to the co-creation of knowledge with practitioners and recipients of research results; or the focus on a 'what works' perspective to motivate practitioners to engage with academics and foster evidence-informed practice. Consider also the need to allow open-ended opportunities for discussion; and the use of services and resources such as translations and the production of policy briefs (Aiello et al., 2021).

Furthermore, there is a linguistic aspect that is worth noting. For certain disciplines in the social sciences and the humanities (and CCJ are likely to be among those), publishing in the native language is preferred because researchers' thinking might be deeply intertwined with their language expressions. Their research may also concern a local issue, and so publishing in the native language might help to better reach the intended audience. Unfortunately, publishing in a language other than English or with a local publisher can imply exclusion from international index databases, and also has an effect on the international visibility of certain research (Huang & Chang, 2008), impacting the (measurable, and locally or internationally perceived) scholarly influence and prestige in the sector (see e.g. Gabbidon et al., 2018; Cohn & Farrington, 2011). As such, as explained by Huang and Chang (2008, p. 1826), an ideal evaluation for research should address issues including the diversity of publication types, the emphasis on book publishing, the heavier reference to books than to journal articles, the local and regional concerns, and the tendency to cite older literatures in the social sciences and the humanities – something that the system in place has not yet fully considered.

Further, the issue of interdisciplinarity remains an elephant in the room, particularly for disciplines such as CCJ that have an inherent interstitial nature (Zender, 2007) and are increasingly open to collaborations with non-neighbouring disciplines in the attempt to make better sense of a complex social reality and technological changes. An example is the importance and the challenges linked to the increasingly common collaborations between criminology and computer and data science when researching the digital field – see Lavorgna, 2021).

4. Concluding thoughts

While the VQR has been inspired by the REF system, with Italy being relatively a latecomer to academic evaluation (Rebora & Turri, 2013; Di Berardino & Corsi, 2018; Buouso, 2021), the two systems have also profound differences. The VQR is potentially more complex and problematic. As stressed by Rebora and Turri (2013), despite the methodological similarities in British and Italian research assessment exercises, there are a number of important differences. According to these authors, while British assessment exercises are historically linked to new public management and are influenced by organisational control theory[3], Italian assessment exercises are mostly affected by neo-institutional sociology[4], and underestimate reactivity to evaluation by researchers and institutions. Again, according to Rebora and Turri (2013), this would stem from the fact that the REF originated from a more solid conceptual base. By contrast, in Italy, an imitative approach prevailed, with 'assessment' being simplistically advocated as 'the' solution to some deep-rooted issues in the Italian research system. A risk here is that managerial ideology might negatively impact academic culture, while failing

[3] Which focuses on the link between the characteristics of assessment systems and the behaviours that they trigger.

[4] Which analyses the behaviour of organisations in terms of legitimacy and approval from their general environment and culture.

to ‘fix’ some malpractices. Research evaluation is not only a neutral, technical procedure, but stems from social constructions and produces consequences at the level of both organizational and individual behaviours. In the UK, for instance, there has been a longer debate on the methods and effects of assessment both in academic-scientific and political-institutional contexts, and pilot exercises were conducted before adopting the REF system. Conversely, in Italy, there was less debate and no trials (Rebora & Turri, 2013). Interestingly, however, empirical research has shown extremely high correlations between the two assessment approaches (Checchi et al., 2021).

We do recognise that if the limitations we have identified can be addressed, both the REF and the VQR have the potential to serve as drivers for ensuring that higher education institutions in the UK and Italy maintain high standards of accountability in relation to public spending on research. Both evaluation frameworks can also encourage researchers to consider the social and economic benefits of their work. But impact requirements and measurements should not stifle academic freedom or restrict research to the boundaries set by those with whom researchers align or partner to deliver outcomes.

In describing and comparing the REF and the VQR, we have stressed some specific challenges affecting CCJ. These are more serious in the Italian case, because of some linguistic and interdisciplinary issues. But we have also noted that most of the problems discussed are systemic and impact various disciplines. Indeed, in other disciplines, many researchers and academic leaders believe that the current system for evaluating research is somehow misaligned with societal needs and disconnected from the challenges and practicalities of many research endeavours. Consider, for instance, the set of recommendations (even if mostly focused on the ‘hard sciences’ and bibliometric sectors) referred to as the San Francisco Declaration on Research Assessment (DORA). They stemmed from a meeting of a group of editors and publishers of scholarly journals during the Annual Meeting of The American Society for Cell Biology in 2012 (DORA, 2022). While recognising the importance of accurately

measuring and evaluating research outputs, the Declaration questions some of the metrics currently in use for funding, appointment, and promotion considerations. Examples include journal-based metrics. Similarly, again in the clinical and life sciences, some academic leaders, funders, and scientists have detailed six principles for assessing scientists and associated research and policy implications, that should be (better) implemented to improve the current approaches. The principles are: rewarding the social benefits of research; assessing researchers on evidence and indicators that can incentivise best publication practices; recognising that all research should be published completely and transparently, regardless of the results, and that collaborative endeavours should be better rewarded; facilitating the dissemination and use of research data and results by others; investing in research to provide the necessary evidence to guide the development of new assessment criteria and to evaluate the merits of existing ones; and rewarding researchers for intellectual risk-taking that might not be reflected in early successes or publications (Moher et al., 2018). While a thorough discussion on alternative evaluation systems would exceed the scope of our contribution, we can only echo the DORA recommendations, which would certainly improve the research culture also in CCJ settings, once integrated with systemic corrections taking into account the disciplinary peculiarities reviewed above (i.e., the variety of possible quality outputs; a broader understanding of what 'impact' is or can be; linguistic needs; and interdisciplinary challenges).

Beyond being a cross-disciplinary issue, the call to improve the mechanisms in place to evaluate research is also being issued globally. For example, Xu and colleagues (2021) observe that the current metrics used for research incentives and evaluation mechanisms in China are creating tensions in research cultures. Humanities and the social sciences are particularly suffering from the reinforcement of a managerial culture in higher education, and from the rising tensions between 'internationalization and indigenization, quality and quantity, integrity and instrumentalism, equity and inequity' (Xu et al., 2021). They fear that the choice for output-oriented, rather than human-oriented, evaluation approaches have undesirable effects on research values, norms, and standards. This resonates with

the concerns raised in the UK and Italy. As such, we hope that our short contribution in the context of this special issue will help further these discussion in a constructive direction. If evaluation exercises can be a physiologic exercise in a healthy system, it is of upmost importance to address the pathologies that hinder academic culture and freedom of research.

References

Abramo G. and D'Angelo C.A. (2015a). The VQR, Italy's second national research assessment: Methodological failures and ranking distortions. *Journal of the Association for Information Science and Technology*, 66(11), 2202-2214.

Abramo . and D'Angelo C.A. (2015b). An assessment of the first 'scientific habilitation' for university appointments in Italy. *Economia Politica*, 32, 329-357.

Abramo G. and D'Angelo C.A. (2021). The different responses of universities to introduction of performance-based research funding. *Research Evaluation*, 30(4), 514-528.

Abramo G. and D'Angelo C.A. (2022). The impact of Italian performance-based research funding systems on the intensity of international research collaboration. *Research Evaluation* (online first).

AdvanceHE (2001). Research Excellence Framework (REF) 2021 and equality. Available at: https://www.advance-he.ac.uk/guidance/equality-diversity-and-inclusion/research-excellence-framework-ref-2021-and-equality.

Aiello E., Donovan C., Duque E., et al. (2021). Effective strategies that enhance the social impact of social sciences and humanities research. *Evidence & Policy*, 17(1), 131-146.

ANVUR (2022a). VQR. Available at: https://www.anvur.it/attivita/vqr/.

ANVUR (2022b). Valutazione della Qualità della Ricerca (VQR) 2015-2019. Press Release. Available at: https://www.anvur.it/wp-content/uploads/2022/04/comunicato-stampa-VQR-2015_2019.pdf.

ANVUR (2022c). Terza Missione/Impatto. Available at: https://www.anvur.it/attivita/temi/.

ANVUR (2022d). Valutazione della Qualità della Ricerca (VQR) 2015-2019. Presentation. Available at: https://www.anvur.it/wp-content/uploads/2022/04/Risultati_VQR_2015_2019.pdf.

Arnold E., Simmonds P., Farla K. et al., (2018). Review of the Research Excellence Framework. Evidence Report. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/768162/research-excellence-framework-review-evidence-report.pdf.

Baratta A. (2019). *Criminologia critica e critica del diritto penale. Introduzione alla sociologia giuridico-penale.* Milan: Meltemi.

Bollen J., Van De Sompel H., Smith J.A. and Luce R. (2005). Toward alternative metrics of journal impact: A comparison of downloadand citation data. *Information Processing & Management*, 41(6), 1419-1440.

Bonaccorsi A. (2020). Two decades of experience in research assessment in Italy. *Scholarly Assessment Reports*, 2(1).

Brisman A., Carrabine E. and South N. (2018). A rendezvous subject? A fragmented discipline?: An introduction to The Routledge Companion to Criminological Theory and Concepts. In *The Routledge Companion to Criminological Theory and Concepts*. London: Routledge.

Buoso E. (2021). Automated Decisions, Research Quality Assessment and Step-Based Salary System: Potential Dangers for Academic Freedom in Italy. In *Academic Freedom Under Pressure?* Cham: Springer, pp. 153-162.

Checchi D., Ciolfi A., De Fraja G., Mazzotta I. and Verzillo S. (2021). Have you read this? An empirical comparison of the British REF peer review and the Italian VQR bibliometric algorithm. *Economica*, 88(352), 1107-1129.

Cohn E.G. and Farrington D.P. (2007). Scholarly Influence in Criminal Justice: Foreword to the Special Issue. *Journal of Criminal Justice Education*, 18(3), 337-339.

Cohn E.G. and Farrington D.P. (2011). Scholarly Influence and Prestige in Criminology and Criminal Justice. *Journal of Criminal Justice Education*, 22(1), 5-11.

Cohn E.G. and Farrington D.P. (2014). Publication Productivity of Criminologists. *Journal of Criminal Justice Education*, 25(3), 275-303.

Copes H., Khey D.N. and Tewksbury R. (2012). Criminology and Criminal Justice Hit Parade: Measuring Academic Productivity in the Discipline. *Journal of Criminal Justice Education*, 23(4), 423-440.

Di Berardino D. and Corsi C. (2018). A quality evaluation approach to disclosing third mission activities and intellectual capital in Italian universities. *Journal of Intellectual Capital*, 19(1), 178-201.

DORA (2022). San Francisco Declaration on Research Assessment. Available at: https://sfdora.org/read/.

Downes D. (1988). The sociology of crime and social control in Britain 1960–1987. In Rock P (ed.) *A History of British Criminology*. Oxford: Oxford University Press.

Else H. (2022). Mammoth UK research assessment concludes as leaders eye radical shake up. *Nature* (news), 12 May. Available at: https://www.nature.com/articles/d41586-022-01310-0.

Franceschini F. and Maisano D. (2017). Critical remarks on the Italian research assessment exercise VQR 2011–2014. *Journal of informetrics*, 11(2), 337-357.

Gabbidon S.L., Higgins G. and Martin F. (2010). Press Rankings in Criminology/Criminal Justice: A Preliminary Assessment of Book Publisher Quality. *Journal of Criminal Justice Education,* 21(3), 229-244.

Garland D. and Sparks R. (2000). Criminology, social theory and the challenge of our times. *British Journal of Criminology*, 40, 189-205.

Gridelli Veligogna N. (1986). *Scipio Sighele. Dalla Criminologia alla Sociologia del Diritto e della Politica*. Milan: Giuffrè.

Harris C., Jones H. and Squires P. (2019). How Criminology is Taught and Researched Today. Available at: https://www.britsoccrim.org/wp-content/uploads/2019/12/How-Criminology-is-Taught-and-Researched-Today-PBCC19.pdf.

Huang M.H. and Chang Y.W. (2008). Characteristics of research output in social sciences and humanities: From a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, 59(11), 1819-1828.

Kleck G. and Barnes J.C. (2011). Article Productivity among the Faculty of Criminology and Criminal Justice Doctoral Programs, 2005–2009, *Journal of Criminal Justice Education*, 22(1), 43-66.

Lavorgna A. (2021). Epistemologies of cyberspace: notes for interdisciplinary research. In Lavorgna A. and Holt T.J. (eds.) *Researching Cybercrimes*. Cham: Palgrave Macmillan, pp.5-22.

Long H., Boggess L.N. and Jennings W.J. (2011). Re-Assessing Publication Productivity among Academic "Stars" in Criminology and Criminal Justice. *Journal of Criminal Justice Education*, 22(1), 102-117.

Moed H.F. (2005). *Citation analysis in research evaluation* (Vol. 9). Dordrecht: Kluwer.

Moeller K. (2019). Ranking Nordic Criminologists by Impact and Prestige. *Journal of Criminal Justice Education*, 30(4), 536-550.

Mohammadi E. and Thelwall M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, 65(8), 1627-1638.

Moher D., Naudet F., Cristea I.A., Miedema F., Ioannidis J.P. and Goodman S.N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS biology* 16(3):e2004089.

Mryglod O., Kenna R., Holovatch Y. and Berche B. (2014). Predicting the results of the REF using departmental h-index: A look at biology, chemistry, physics, and sociology. *LSE Blogs*. Available at: https://blogs.lse.ac.uk/impactofsocialsciences/2014/12/16/predicting-the-results-of-the-ref/.

Muhonen R. and Tellmann S.M. (2022). Challenges of reporting societal impacts for research evaluation purposesthe case of sociology. In Engels C.E. and Kulczycki E. (eds.) *Handbook on Research Assessment in the Social Sciences.* Edward Elgar Publishing.

Priem J. and Hemminger B.M.H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social web. *First Monday*, 15(7).

Rebora G. and Turri M. (2013). The UK and Italian research assessment exercises face to face. *Research policy*, 42(9), 1657-1666.

REF (2018). Equality briefing for panels. Available at : https://www.ref.ac.uk/media/1017/equality-briefing-for-panels-ref-2018_05.pdf.

Rowlands I. and Nicholas D. (2007). The missing link: journal usagemetrics. *Aslib Proceedings*, 59(3), 222.

Russel Group (2019). Why we need QR funding. Available at: https://russellgroup.ac.uk/news/why-we-need-qr-funding/#:~:text=Quality-related%2C%20or%20%E2%80%98QR%E2%80%99%20research%20funding%2C%20is%20the%20invisible,treatments%20for%20diabetes%2C%20depression%2C%20dementia%20and%20many%20more.

Savage W.E. and Olejniczak A.J. (2022). More journal articles and fewer books: Publication practices in the social sciences in the 2010's. *PLoSONE*, 17(2), e0263410.

Sayer D. (2014). Five reasons why the REF is not fit for purpose. Available at: https://www.theguardian.com/higher-education-network/2014/dec/15/research-excellence-framework-five-reasons-not-fit-for-purpose.

Schloegl C. and Gorraiz J. (2011). Global usage versus global citationmetrics: The case of pharmacology journals. *Journal of the AmericanSociety for Information Science and Technology*, 62(2005), 161-170.

Sivertsen G. (2022). Publishing in the social sciences and its representation in research evaluation and funding systems. In Engels C.E. and Kulczycki E. (eds.) *Handbook on Research Assessment in the Social Sciences* Edward Elgar Publishing, pp.238-261.

Stern N., Black J., Borysiewicz L., et al. (2016). Building on Success and Learning from Experience. An Independent Review of the Research Excellence Framework. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf.

UCU (2013). Survey finds widespread dissatisfaction amongst university staff with research assessment. Available at: https://www.ucu.org.uk/article/6776/Survey-finds-widespread-dissatisfaction-amongst-university-staff-with-research-assessment.

UKRI (2022). How Research England Supports Research Excellence. Available at: https://www.ukri.org/about-us/research-england/research-excellence/research-excellence-framework/.

Verma A. (2020). *Anti-Racism in Higher Education*. Bristol: Bristol University Press.

Walker J.T. and Raptopoulos K.E. (2008), Journal Editors and Journal Writers: Service vs Publication as a Measure of Faculty and Department Productivity. *Journal of Criminal Justice Education*, 19(2), 251-274.

Watermeyer R. and Derrick G. (2022). Why the party is over for Britain's Research Excellence Framework. Available at: https://www.nature.com/articles/d41586-022-01881-y.

Wildson J. and Weinstein N. (2018). Why we need a real-time REF Review to plan for 2027. *WonkHE*, April 12. Available at: https://wonkhe.com/blogs/why-we-need-a-real-time-ref-review-to-plan-for-2027/.

Xu X., Oancea A. and Rose H. (2021). The impacts of incentives for international publications on research cultures in Chinese humanities and social sciences. *Minerva*, 59(4), 469-492.

Yarrow E. (2018). *Gender and the Research Excellence Framework*. Rotterdam: Brill Sense.

Zedner L. (2007). Pre-crime and post-criminology? *Theoretical criminology*, 11(2), 261-281.