

Editorial

Stefania Panebianco
Francesco Zucchini

ITALIAN POLITICAL SCIENCE CO-EDITORS

In the last decade, research assessment has become a crucial dimension of academic life, in Italy as elsewhere in Europe. Today, career advancements, research fund granting or department ranking, are increasingly dependent on research assessment as one of the key indicators. Individual and collective evaluations are thus mixed up to achieve the declared objectives of more efficient use of public funds, on the one hand, and more innovative and advanced research, on the other.

This IPS issue hosts the contributions presented at the Round Table on research assessment in a comparative perspective, held at the SISP annual conference, in September 2016. The IPS community will certainly benefit from the experience of the distinguished international scholars who participated in the Round Table, namely Matthew Flinders (PSA), Rudy Andeweg (ECPR and Dutch association for political science), Catherine Paradeise (Professor emerita UPEM-LISIS) and Daniele Checchi (member of GEV-13 for Research Quality Evaluation (VQR) 2004-2010). SISP President Simona Piattoni, who organized this Round Table, is the guest editor of this issue. We thank them all for having contributed to the on-going debate, providing food for thought from the European academic environment.

This is not the first time that IPS has focused on research evaluation and VQR. The very first issue coordinated by the current IPS co-editors was devoted to “Evaluating the Evaluation. The pros and cons of ‘VQR’ in social and political research”, in IPS Volume 8, Issue 1 of June 2013, which addressed controversial issues such as the conceptualization of quality (i.e. quality as an objective fact or as a social construction), the ex post adoption of evaluation criteria, the fact that “the number of citations may reflect the level of subordination of scholars who cite, instead of indicating the degree of innovation, originality and explanatory power of the cited publication”, etc.

Four years later, the third VQR assessment edition has ended, and research assessment as a tool has been generally accepted by Italian scholars (to be precise, fewer than 10% of them abstained from the VQR). Yet, the debate on research assessment in social sciences is still faced with several dilemmas: qualitative versus quantitative evaluation, quantitative indicators privileging single products or journal quality (e.g. journals with IF), the definition of internationalization of research, which does not simply mean publishing in English, etc. Alongside that, periodic research assessment has an impact on academia both in individual and systemic terms, on an individual’s academic profession and on departments in terms of ranking and fund distribution.

It remains to be seen whether an article is quoted because of the popularity of the author, the relevance of the topic, the impact of the journal or the quality of the article itself.

Paradoxically, once peer-review has become widespread, recruiting peer-reviewers has become a hard task, because peer-review is not regarded as a rewarding activity, it is a 'hidden' prestige and not many scholars are happy to remain behind the scenes. Moreover, peer-review methods rely upon the subjectivity of reviewers. The introduction of a peer to peer-review might be useful to evaluate the evaluators.

The comparative perspective provided by this IPS issue helps to foster the Italian debate on research assessment and to better understand the criticisms of research assessment, which can be useful in academic advancement and should not risk becoming an individual or systemic hindrance.

FOCUS ON:

The Impact of Research Assessment on the Profession and the Discipline of Political Science: a Roundtable

Simona Piattoni

PRESIDENT OF THE ITALIAN POLITICAL SCIENCE ASSOCIATION
UNIVERSITY OF TRENTO
GUEST EDITOR OF THIS ISSUE

*Proceedings of the roundtable held at the SISP Annual Congress,
Department of Social and Political Sciences,
University of Milan, 15 September 2016*

Several European countries – UK, Sweden, Spain, Norway, Netherlands, Italy, Ireland, Hungary, Germany, France, Finland and Belgium – conduct by now periodic research assessment exercises. In particular, the Italian university system has already conducted three such rounds of assessment, the last two (2004-2010, 2011-2014) according to particularly formalized procedures. And yet the very idea of assessing the scientific production as if it were a specific output of the university systems still meets with considerable resistance and skepticism. Many doubt that such assessment has any consequence at all (while it does have a financial repercussion on the distribution of funds from the Ministry of Education, which departments can use to expand their teaching staff) and most fear that it simply lends itself to constructing rankings of little scientific significance, but potentially great political import. The controversy is particularly intense in the social sciences and humanities, therefore in political science, an area in which the so-called bibliometric indicators are more difficult to apply and assessment is mostly qualitative. We think that the time has come for a collective reflection on the pros and cons of such exercises and on the potential repercussions that they may have on the academic profession and the discipline of political science.

We have invited four distinguished colleagues from four different university systems – UK, Netherlands, France and Italy – to discuss about these issues and to bring their particular experience and points of view to fruition of the Italian political science community. We asked them to comment on and report their experience on the following aspects of research assessment:

1. Have research assessment exercises in your country been met with enthusiasm and collaboration or with suspicion and resistance? What were the arguments pro and against? What was the return rate?
2. Which aspects have been pinpointed as being particularly problematic: a) use of quantitative indicators (such as single product/journal impact factor); b) interna-

tionalization (often coinciding with “publishing in English”); c) publishers’ prestige; etc.?

3. Which consequences – monetary or otherwise – have these exercises had on: a) single scholars; b) departments; c) universities?
4. Which aspects have been reformed/improved from one round to the next? Have the problems encountered in early rounds been amended in successive rounds?
5. What impact have these exercises had on the academic profession in political science? Have they prompted a higher rate of international submissions? Have they improved overall production rates? Have they encouraged publications of journal articles as opposed to monographs?
6. What is your overall assessment of research assessment in your country?

Roundtable participants:

- **Prof. Matthew Flinders** (*Political science, University of Sheffield, UK*), Chair of the Executive Committee of the Political Studies Association (PSA).
- **Prof. Rudy Andeweg** (*Political science, Leiden University, Netherlands*), Chair of the Executive Committee of the European Consortium for Political Research (ECPR).
- **Prof. Catherine Paradeise** (*Sociology, Université Paris Est-Marne-la-Vallée*), Professor emerita UPEM-LISIS (Laboratoire Interdisciplinaire Sciences Innovations Sociétés).
- **Prof. Daniele Checchi** (*Economics, University of Milan*), member of the Group Expert Evaluator in Economics (GEV-13) for the Research Quality Evaluation (VQR) 2004-2010.

Moderator:

- **Prof. Simona Piattoni** (*Political science, University of Trento*), President, Società Italiana di Scienza Politica, member of the Group Expert Evaluator in Political and Social Sciences (GEV-14) for the Research Quality Evaluation (VQR) 2011-2014.

The Impact of Research Assessment on the Profession and the Discipline of Political Science

Simona Piattoni

PRESIDENT OF THE ITALIAN POLITICAL SCIENCE ASSOCIATION
UNIVERSITY OF TRENTO

At the 2016 SISP annual meeting, in Milano, we held a roundtable on the topic of research assessment in comparative perspective. Leading European scholars, both expert of evaluation and with significant experience at the helm of their respective national associations, took part in this roundtable: Prof. Matthew Flinders, Chair of the Executive Committee of the Political Studies Association (PSA); Prof. Rudy Andeweg, Chair of the Executive Committee of the European Consortium for Political Research (ECPR) and former Chair of the Dutch political science association; Prof. Catherine Paradeise, Professor emerita UPEM-LISIS (*Laboratoire Interdisciplinaire Sciences Innovations Sociétés*) and expert of academic evaluation; and Prof. Daniele Checchi, member of the Group Expert Evaluator in Economics (GEV-13) for the Research Quality Evaluation (VQR) 2004-2010. They were asked to share their experience with research assessment in their countries and to contribute their points of view to the discussion of the impact of research assessment in the social sciences, and particularly in political science. As organizer of the roundtable, I asked them to comment on and report their experience on the following aspects of research assessment:

1. Have research assessment exercises in your country been met with enthusiasm and collaboration or with suspicion and resistance? What were the arguments pro and against?
2. Which aspects have been pinpointed as being particularly problematic: the use of quantitative indicators (such as single product/journal impact factor); the pressure towards internationalization (often coinciding with 'publishing in English'); debatable rankings of publishers' prestige, etc.?
3. What impact have these exercises had on the academic profession in political science? Have they prompted a higher rate of international submissions? Have they improved overall production rates? Have they encouraged publications of journal articles as opposed to monographs?
4. What impact have these exercises had on the academic profession in political science? Have they prompted a higher rate of international submissions? Have they improved overall production rates? Have they encouraged publications of journal articles as opposed to monographs?

5. Which aspects have been reformed/improved from one round to the next? Have the problems encountered in early rounds been amended in successive rounds?

The context for such an analysis is the fact that several European countries – UK, Sweden, Spain, Norway, Netherlands, Italy, Ireland, Hungary, Germany, France, Finland and Belgium – conduct by now periodic research assessment exercises. In particular, the Italian university system has already conducted three such rounds of assessment, according to particularly formalized procedures. We therefore thought that the time had come for a collective reflection on the pros and cons of such exercises and on the potential repercussions that they may have on the academic profession and the discipline of political science, as the very idea of assessing scientific production was met with considerable resistance and skepticism in many countries, Italy included. Some are opposed in principle to the idea of assessing scholarly products *as if* they were just any other product, implicitly rejecting both the logic of accountability (how are public or private/public funds spent) and the logic of monitoring. Others have misgivings about the specific way in which this assessment is carried out, and in particular about the construction of (increasingly) indicators-driven excellence rankings of departments, scholars and disciplines. A handful doubt that such assessment has any consequence at all (while it does have a small financial impact on the distribution of funds from the Ministry of Education, which departments can use to expand their teaching staff), while many object to the mostly undesired consequences that these exercises have on the development of the academic profession. The controversy is particularly intense in the social sciences and humanities, therefore also in political science, areas in which the so-called bibliometric indicators are more difficult to apply and assessment must therefore remain mostly qualitative.

The articles that follow are the much elaborated and refined texts of the interventions presented at the roundtable. The article by Flinders sketches the long history of British research assessment and warns against the subtle and paradoxical effects of the potential excesses of a productivity-driven assessment of academic activity. The articles by Andeweg and Paradeise show how other European countries tailor research assessment to the specific needs and particularities of the national organization of academic and research institutions. And finally, the article by Checchi provides abundant data on the Italian experience with the VQR (*Valutazione della Qualità della Ricerca*), allowing the readers to draw their own conclusions. In thanking again the participants for their generous contribution of time, knowledge and ideas, I would like to draw attention to a few common themes that emerge from these articles:

1. The managerial logic that inspired these assessment exercises, particularly in the UK but also in other European countries, was mostly implemented over the heads of the academic profession as a way of curtailing what were perceived to be outdated privileges; the academic profession has been mostly sidelined in the elaboration and implementation of these procedures and has shown either skepticism and resistance or indifference to the idea of assessing scholarly production; this is particularly shocking in the case of political science, as political scientists have been marginalized in one of their putative fields of expertise – the politics of academic policy-making;

2. Assessment of research products tends to increasingly rely on quantitative indicators and rankings of journals and publishing houses – an aspect which is contested also, e.g., in France and in Italy – in a mimetic attempt to emulate the hard sciences; such indicators increasingly acquire a life of their own, being often used as summary indicators for the scientific worth of Departments and scholars;
3. Research assessment exercises introduce a number of potential distortive elements: a) peer-reviewed journal articles tend to be assessed better than edited volumes and monographs, regardless of their real value; b) joint works tend to be preferred over single-authored works, inducing the artificial inflation of multi-author products; c) all other things being equal, works in English attract greater readership and gain higher impact factors than works in national languages, which affects particularly academic communities which do not use English as their first language; d) research assessment rankings of Departments and scholars induce ‘gaming strategies’ that create further distortions (strategic hiring, discouragement of teaching, creation of two-tier academic milieus) that do not necessarily secure better scholarship; d) the advantages of creating a culture of assessment, peer-review and accountability may be more than offset by the costs, in terms of time and money, of the assessment exercise itself;
4. The impact of research assessment on departmental funding is highly uneven across Europe – small but meaningful in Italy and France, inexistent in the Netherlands (where cuts and increases in funding follow a different logic) and indirect in the UK (through the effect that rankings have on the attractiveness of departments for scholars and students is remarkable) – while the impact on the nature and pressures of being in academia are momentous (described in one of the contributions as ‘going MAD’); the relevance of political science for society may have paradoxically suffered from this attempt to make it more socially accountable, as the energy and attention of scholars has been in part diverted from the pursuit of interesting, cross-disciplinary research questions to the production of formally more polished and marketable works;
5. Subsequent reforms of the exercise have, in certain cases, tried to correct some of the perceived distortions by adding, e.g., teaching assessment or by correcting the number and selection of works to be assessed, but these corrections run the risk of introducing new distortions of their own.

In conclusion, while research assessment throughout Europe addresses the issues of transparency, comparability and accountability in the academic world, it also carries challenges of its own that affect particularly the social sciences and the humanities. Italian political science is neither alone nor unique in experiencing some difficulties in having its production being assessed through such methods, yet it would be difficult to argue that the assessment should cease and that the academic world should deprive itself of this instrument of self-evaluation and accountability towards society. The one overarching lesson that we may perhaps draw from this comparative analysis is that political scientists need to pay greater attention to academic politics and policy, and should attempt to play a more proactive role in defining the standards and goals of academia.

The Tragedies of Political Science: The Politics of Research Assessment in the United Kingdom

Matthew Flinders

UNIVERSITY OF SHEFFIELD

The *Tragedy of Political Science* is the title of a 1984 book by David Ricci that made a bold argument concerning the evolution of the discipline. Ricci's thesis, put simply, suggested that as political science had become more 'professionalised' throughout the twentieth century so it had also become less relevant, more verbose, less engaged, more impenetrable, increasingly distant from practitioners of politics and the public. The discipline had simply not lived up to the high hopes of C. Wright Mills for the social sciences – as set out in *The Sociological Imagination* (1959) – but had, if anything, become ensnared in a trap of its own making. In 1967 the Caucus for a New Political Science (CNPS) was created in the United States in order to encourage social engagement and activism amongst political scientists in direct rejection of the American Political Science Association's (APSA) commitment to political neutrality and refusal to engage in major social debates. The 'tragedy' as both David Ricci and the CNPS argued was that at a historical point when American society desperately needed the evidence and insights that political science could deliver the discipline apparently either did not want to engage or had little to say. Political science – to paraphrase C. Wright Mills – had failed to deliver on its early promise.

To offer this disciplinary narrative is to offer little that is novel or new. The flaying of political science has emerged into a popular intellectual pastime in recent years and there are clear exceptions to this broad account in the form of individual scholars or sub-disciplines that have retained a clear social connection. Engaged scholars, however, arguably became very much the exception rather than the rule in a profession that incentivised sub-disciplinary balkanisation, methodological hyper-specialism, theoretical fetishism and the development of esoteric discourses. This is, of course, an account of American political science that came to a head with the emergence of the Perestroika Movement in 2000 with its demands for greater methodological pluralism (within the discipline) and greater social engagement (beyond the discipline). While political science *beyond* the United States was never quite so seduced by the promises of rat-choice, quantitative, large-n, mathematical methodologies to fuel a 'raucous rebellion', the issue of whether and how the discipline should be required to demonstrate its non-academic relevance, social impact or public value has become a global challenge for the discipline. This, in turn, has spawned a growing pool of scholarship on the structural and contextual factors underlying

political science's apparent 'relevance gap' and, from this, how to 'make political science relevant' (Gilberto Capano and Luca Verzichelli (2016; see also 2010).

The focus of this article is more specific, provocative and future focused. It concerns an analysis of the impact of arguably the most explicit and potentially far-reaching externally-imposed research audit process in the world – the United Kingdom's Research Excellence Framework (REF). As such this article reviews the REF assessment process before then exploring the impact and unintended consequences of this incredibly dominant and demanding 'meta-governance' framework for British higher education. It then locates the 'politics of' research assessment within a far broader and innovative account of the changing nature of scholarship in the United Kingdom. The core argument is that over thirty years since the publication of Ricci's *Tragedy of Political Science* it is now more appropriate to survey the *tragedies of political science* in the sense of an apparent failure by the discipline to adopt a strategically selective and politically astute approach to navigating a changing socio-political context. Political science, put simply, frequently appears very poor at utilising the insights of its own discipline. The tragedies of political science therefore exposes the discipline's own lack of political guile.

Could it be that this analysis makes the classic mistake of over-generalising from a single case study? What relevance does the REF have for scholars based beyond the shores of an island of less than a quarter-of-a-million square kilometres off the coast of Northern Europe? The answer is simple. As already mentioned, concern regarding the social benefit of publicly funded scholarship has moved up the political agenda in recent years and shows no sign of abating (quite the opposite). Add to this the fact that the UK has for some decades been a world-leader in terms of imposing market-based managerialist reforms in higher education that often have a subsequent ripple-effects beyond its shores and the relevance of this article for debates about scholarly relevance becomes clear. REF-like – or what might more accurately be described as REF-lite – procedures have or will be implemented in a large number of countries during the next decade and to be forewarned is to be forearmed. It is for this reason that this article is divided into three main parts.

Part I provides a descriptive account of the introduction and evolution of research assessment frameworks in the United Kingdom. Part II then dissects this chronological account to expose the unintended consequences and intellectual pathologies of this process. The argument is not that research assessment frameworks are a 'bad thing' of that they do not deliver positive impacts but it is to emphasise the manner in which new frameworks have to be very carefully calibrated in order to avoid negative over-steers and short-term gaming of the system. The final section – Part III – argues that any understanding of the impact of research assessment frameworks has to be located within an account of the wider context of higher education and the rapidly shifting sands of scholarship. The main concluding argument is therefore not about research assessment *per se* but about how research assessment is contributing to the break-up of academe and the splintering of disciplines along diverging pathways. The tragedies of political sciences are therefore converging to create the unravelling or unbundling of both the discipline and the nature of scholarship. It is this innovative macro-political argument that forms the main contribution of this article.

1. The Meta-Governance of Research

Although a far simpler and more accurate title for this section might be ‘Tragedy and Farce’, the main argument is simply that it is impossible to separate the introduction of externally imposed assessments of research by governments (either directly or indirectly) from broader debates concerning the distribution of political power in a polity. Therefore, although APSA may well have been at fault for failing to position political science more strategically or to ensure that evidence of clear social benefit and relevance for the discipline was not always conveniently at hand it is also true that large sections of the right-wing political elite in the United States treated universities, in general, and political science departments, in particular, with a mixture of disdain and distrust. The attempts between 2009 and 2014 to block or restrict federal funding to political science in the United States may therefore have been couched in terms of economic prioritisation in times of austerity but they actually veiled deeper ideological views about the role and independence of scholars and universities. In many ways, the decision by the Conservative Government to introduce a new Research Selectivity Exercise in 1986 was a similarly charged exercise in pressure politics. Mrs Thatcher had become Prime Minister in 1979 on the basis of a need for public sector control, discipline and cutbacks. The belief in the capacity of the market over the state led to what David Marquand (2014) termed ‘the decline of the public’ in the sense of an assault on those basic anchor institutions within society that were designed and intended to promote collective values over individualised notions of society. The politics of ‘the public’ was rapidly eviscerated by a new politics of ‘the private’ that not only limited the powers of collective-bargaining institutions such as the trade unions but that also imposed neo-liberal values across the public sector in the guise of ‘New Public Management’ (NPM).

NPM was a broad set of managerial methods that all shared a simple faith in the capacity of the market to drive-up performance, increase efficiency and expose shirking. The promise was ‘a government that worked better and cost less’ but – as the prize-winning scholarship of Christopher Hood and Ruth Dixon (2015) has revealed – rarely lived up to its rhetorical promises. A neo-liberal philosophy of management that was supposed to banish bureaucracy and increase dynamism actually gave birth to new forms of ever more elaborate and organisationally suffocating rules, regulations and red tape. The universities were, however, more of a challenge to the Conservative governments of the 1980s due to their historically embedded independence from direct government control. But Mrs Thatcher was very much a conviction politician. Her time at the University of Oxford has convinced her that universities were complacent institutions that were overly protected from market forces. The public deserved better – in terms of both performance and accountability – and it was her job as Prime Minister to find a way to make that happen. Constrained by intra-party tensions during her first government (1979-1983) her capacity to intervene was limited to introducing fees for overseas students in 1981. By her second term in office – 1983-1987 – Mrs Thatcher was in a far stronger position in terms of her grip on the Conservative Party and could therefore use this stable foundation in order to institute more radical measures.

Lacking direct control capacity the obvious lever for affecting change lie in relation to public funding. In short, if Mrs Thatcher (or, more precisely her government) could not easily impose reforms on the governance of universities she could impose requirements

and controls upon the distribution and use of public money. The shift was therefore one of meta-governance (i.e. ‘the governance of governance’ or ‘rules of the game’). The University Grants Committee (UGC) had existed since 1918 with a remit to act as a buffer between higher education and the government of the day. Its main role had been to distribute block Treasury grants to universities (with the remainder of university funding coming from tuitions fees paid in full on every student’s behalf by his or her local authority). The UGC was, Thatcher believed, a committee made-up of academics to distribute large amounts of public money to academics and although the body did oversee the implementation of the first ‘Research Selectivity Exercise’ in 1986 it was abolished in 1989 with its power transferring to a new University Funding Council (UFC) on which academics were a minority.

The shadow of central government control had therefore become far tighter and the UFC oversaw the introduction of a classic NPM framework involving contracts, performance monitoring and league tables. With the benefit of hindsight it is possible to suggest that one historical tragedy was the failure of British universities to divert, subvert or shape that Thatcherite revolution; their general position of obstinacy combined with predictions of impending crisis that ultimately proved woefully inaccurate simply reinforced the Conservative governments’ belief that universities leaders were out of touch and radical reform was necessary. And while possibly not immediately interpreted as a radical act the introduction of the first Research Selectivity Exercise (RSE) in 1986 can now be viewed as a crucial initial wedge or crack in the governance of universities that has subsequently developed into a core feature of academic life in the UK. The degree to which this initial ‘crack’ or ‘wedge’ has been expanded is clear from the manner in which the RSE involved the UGC creating subject specific sub-committees that would review just five research outputs – books, articles, papers, etc. – from the past five years on which the department in question would be ‘content for its total output to be assessed’. In addition to these outputs departments were invited to submit up to four pages of general description about their research strengths. This really was ‘light touch’ to the extent that one subject committee was so confident that it already knew all it needed to know about each university’s departmental quality, it produced a provisional classification before it received any submissions and ‘when it got all the extra evidence it saw no reason at all to alter any of the classifications’.¹ As a point of comparison the most recent assessment (REF2014, discussed below) required four outputs for every member of staff returned, plus extensive datasets on a whole range of topics, plus a range of environment documents plus a number of independently verifiable ‘impact case studies’.

It is therefore possible to identify a rather rapid process of ‘regulatory creep’ with all the usual unintended bureaucratic and organisational implications. The existence of an academic ‘expert body’ (i.e. the UGC) controlling a process that was implicitly designed to shed light on the previously murky world of academic funding allocation did not go unnoticed. In 1986 This was exactly the insider-elite sort of horse-trading, pork-barreling that Mrs. Thatcher was so personally committed to abolish across the public sector. By 1989 the UGC had been abolished and replaced by a ‘non-expert’ UFC and by the 1989 exercise – now labeled the Research Assessment Exercise (RAE) – universities were permitted to return two pieces of work for each member of staff and information was also sought on the total volume of a department’s research outputs. Other changes involved a shift from the

original 37 subject specific sub-committees were replaced by 152 subject units assessed by nearly 70 panels who would, in turn, apply a new five-point scale to assess the quality of research. (Interesting recommendation made by the chief executive of the UFC for non-academic impacts achieved by each department to be evaluated and therefore incentivized was rejected.)

The decision to end the binary division between research-focused universities and teaching-focused polytechnics in 1992 created new challenges for the assessment of research. Some 'new' universities were clearly committed to developing research-related reputations in some fields but the overall pot of research funding was not going to be increased. The RAE therefore had to become far more robust and rigorous which, in effect, meant the rapid creation of a body of administrative law around higher education research assessment. Some of the decisions that were produced by the 1992 RAE were subject to challenge in the courts and although the UFC successfully defended their decisions the view of the judiciary was clear: academics could no longer make decisions of what was 'good' or 'bad' research based upon their claimed professional expertise and subjective judgements.

The response was a doomed attempt to replace normative judgement with administrative and technical precision. Rules, regulations and methodologies were codified and procedures and processes formalized but all this achieved was an ever-greater bureaucratic burden on universities, research managers and academics. Peer review by specialist disciplinary panels remained the core method for assessing research quality. Moreover, as the results of RAE developed important (indirect) league-table implications for the recruitment of students, particularly international students, it became clear that a football-like transfer market was emerging within higher education. Inflated salaries could be demanded by a small number of research-intensive scholars who would, in turn, secure research-only positions that effectively ensured their 'elite' status, insulated them from pressures (such as teaching or administration) that might threaten their position and thereby reinforced their high-market value. By the 2001 RAE new rules were therefore being implemented about universities 'poaching' staff towards the end of an assessment period in order to claim outputs that had in reality been researched and written at a different institution. By 2008 the situation had become even more complex with an attempt to disaggregate departmental performances. Prior to this date all departments received a simple assessment grade – with 1* being the lowest and 5* the highest – but there was widespread game-playing in the sense that the overall grade could hide a multitude of weaknesses within a department. Many departments would have 'a long tail' in the sense of a fairly large number of staff who were simply not research active or undertaking work that was deemed of insufficiently quality. In 5* departments these 'long tail' staff would effectively be over-graded and over-funded because the department received (and were funded based upon) a flat score (i.e. 4*). Conversely in a largely teaching-focused department that did have a small number of excellent research active staff these academics would be unfairly penalized (and under-funded) by being captured within the overall grade of a weak department.

The answer was to adopt a more refined process based upon 'quality profiles' that reflected the performance of all staff and made more refined calculations on the basis of excellent research performance even if it was found in relatively small pockets. The aim

being to encourage dynamism and to penalize those departments that did in effect carry ‘a long tail’. The problem was that as the research assessment process became more ‘robust’, ‘fine-grained’ and ‘professional’ it also became more demanding upon academics and universities in terms of both administrative costs and emotional distraction. The benefit of this historical ‘long view’ is that it provides an almost perfect representation of Parkinson’s Law of Bureaucracy – that every reform to reduce bureaucracy and increase organizational performance will inevitably have the opposite effect. The RAE had become ‘the tail that wags the dog’ (a typically befuddling English phrase that simply means that a secondary or subservient object, process or operation is in fact dominating an issue). A major review was initiated under the chairmanship of Professor Sir Gareth Roberts in 2003 and led to the recommendation that teaching-focused universities be given the opportunity to opt out of the RAE in return for a guaranteed base level of funding. This was rejected by the institutions it was intended to help due to a concern by them that taking such an ‘opt out’ would send out the wrong message to potential students and research funders. Put slightly differently, the Roberts’ ‘opt out’ was interpreted as reflecting a lack of ambition by any university who opted for it and in an increasingly aggressive and market-driven environment this could be a suicidal strategy.

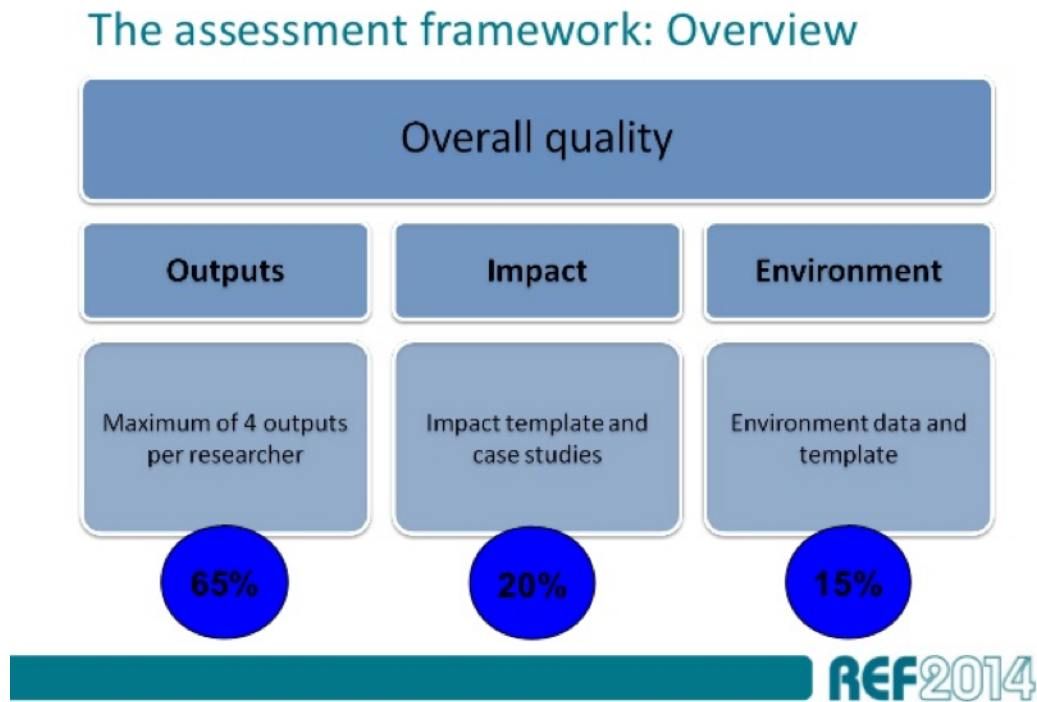
Even the efforts of the Treasury failed to trim the bureaucratic costs of research assessment and in 2008 the government estimated that simply participating in the exercise was costing English (note, not Irish, Welsh or Scottish universities) nearly 50 million pounds of public funding that could otherwise have been dedicated to the primary tasks of the institutions (i.e. Research and teaching). The Treasury did, however, isolate a new option in the form of metrics that could in theory reduce the bureaucratic burdens on universities. This would involve the adoption of a set of metrics such as citation statistics, journal Impact-Factor scores, and other quantitative methods as proxies for research quality therefore removing the need for time-consuming and elaborate procedures for peer review that were in themselves highly normative. The constant analysis of specific metrics could even remove the need for five-yearly research assessment cycles and provide more accurate and up to date information on which funding decisions could be made. In 2006 the then Chancellor of the Exchequer, Gordon Brown, made a surprise announcement that the next RAE was to be a metric only exercise and that it would be for the Higher Education Funding Council for England (HECFE) to decide the specific details for this process. Unfortunately the Chancellor had not forewarned HEFCE (which had itself been created in 1992 to assume the functions of the UFC) of this announcement and a period of intense confusion reigned until a compromise situation was agreed whereby although it was too late to change the criteria for the 2008 RAE the 2014 exercise – now retitled the ‘Research Excellence Framework’ (REF) in an attempt to escape from some of the negativity that had emerged around the RAE – would for the first time include a new ‘impact’ component in return for the Treasury dropping its proposals for metrics.

The meta-governance of research funding within higher education has therefore been transformed since the mid-1980s from essentially an internal, informal and elitist system of financial distribution (i.e. the UGC) through to the external imposition of an incredibly extensive audit and assessment framework with huge associated costs.

Table 1. The Evolution of Research Assessment Exercises in the UK, 1986-2014.

When?	What?	Who?	How?
1986	Research Selectivity Exercise	University Grants Committee	Subjects divided into 37 'cost centres'. For each one, universities were asked to submit five outputs – books, articles, patents, etc. – and up to four pages of general description of the unit's research and industrial strength. Assessments were made UGC's subject subcommittees on a four-point scale from 'below average' to 'outstanding'.
1989	Research Selectivity Exercise	Universities Funding Council	Subjects divided in 152 units of assessment, evaluated by nearly 70 peer review panels. Up to two publications for every member of staff plus information on research student numbers and research income. Results were presented on a five-point scale related to national and international standards.
1992	Research Assessment Exercise	Higher Education Funding Council for England	About 2,800 submissions in 72 units of assessment were rated by 63 subpanels. All 'research active' members of staff at the census date were required to submit up to two publications and two other forms of public output. Submissions were audited for accuracy. No funding was given to departments assigned the lowest of the five grades.
1996	Research Assessment Exercise	Higher Education Funding Council for England	Sixty panels assessed work in 69 units of assessment produced over four years for the sciences and six for the humanities. Universities filed up to four publications per academic submitted. Each sub-panel published its assessment criteria and working methods ahead of submissions and more outside assessors, including academics from abroad and people working beyond the sector, were recruited. Two extra levels were added to the now seven-point assessment scale: 3 was split into 3a and 3b and 5* was added at the top. Departments in the lowest two categories received no funding.
2001	Research Assessment Exercise	Higher Education Funding Council for England	Nearly 2,600 submissions were made to sixty-nine units of assessment. Five "umbrella groups" of panel chairs in related disciplines met to try to achieve greater consistency and fairer assessment of interdisciplinary work. Submissions awarded top grades were reviewed by international experts. Reductions in the number of research outputs required were allowed for some staff in certain circumstances. More feedback on results was given to vice-chancellors, and panels' assessment of the strength of their disciplines was published. Universities whose staff were poached ahead of the census date were permitted to submit two of their outputs, but all the funding still went to their new departments. Departments rated in the top two categories contained nearly 40 per cent of academics, compared with only 13 per cent in 1992. Funding was withdrawn from departments rated 3b and the amount of funding for those rated 4 was steadily reduced.
2008	Research Assessment Exercise	Higher Education Funding Council for England	The sixty-seven subpanels were overseen by fifteen main panels. Explicit criteria were introduced for the assessment of applied, practice-based and interdisciplinary research. Results were presented as 'quality profiles', setting out the proportion of each department's submissions that fell into five quality categories. Research in the top three categories was originally funded, but this has since been reduced to the top two.
2014	Research Excellence Framework	Higher Education Funding Council for England	The research at 154 universities was assessed. A total of 1,911 submissions were made including 52,061 staff, 191,150 academic outputs and 6,975 impact case studies. Thirty per cent of submissions were assessed as world-leading (i.e. 4*), forty-six per cent as internationally excellent (3*), twenty per cent 'recognised internationally' (2*) and just three per cent as 'recognised nationally' (1*).

The evolution of this framework was criticized by the unions as little more than the advancement of marketization into the university sector but was reluctantly accepted by university leaders who seemed almost unable to frame a coherent response or to influence the agenda in a manner that may have smoothed some of the rough edges of the process. However, before examining some of these costs and 'the politics of' this process in more detail, and therefore how the nature of academic life and scholarship has and is changing in the UK, it is worthwhile briefly charting the REF2014 framework and results. With this in mind, possibly the most significant element of REF2014 was the introduction of a significant assessment component for the social impact of scholarship (see Figure 1, below). Each department or unit would now have to deliver a number of 'impact case studies' that could be independently verified and clearly demonstrated the link between research outputs and some significance element of non-academic impact. This was significant for a number of reasons but not least due to the simple fact that 'impact' was a new and potentially game-changing part of the assessment process. Most institutions had become adept at managing the publication profiles of its staff, of managing the existence of 'long tails' and making claims regarding the existence of a dynamic, collegial and stimulating research environment. Demonstrating non-academic impact was a new piece of the assessment framework that would dwarf the marginal gains delivered by tinkering with publications management and could therefore transform the results and subsequent league tables. How exactly the introduction of this major new component could be reconciled with REF2014's stated objective of '[reducing] significantly the administrative burden on institutions in comparison to the RAE' was unknown but would (perhaps not surprisingly) surface as a major issue in the wake of the process.

Figure 1. The REF2014 Assessment Structure.

In order to fulfil the assessment framework each ‘unit of assessment’ (i.e. a department or part of an academic school that wished to be assessed in a certain discipline) was expected to make a formal submission consisting of five main elements. Part 1 related to information and data regarding the number of staff that were being submitted (proportions, exemptions, etc.); Part 2 detailed the publications (up to four) that were being submitted for assessment by each member of staff. Part 3 then required the submission to describe the unit’s approach to enabling non-academic impact from its research and case studies providing specific examples. Part 4 harvested a range of data about the broader research environment of an institution such as the number of research degrees awarded, research income, etc. Part 5 featured a narrative statement prepared by each unit about the research culture, environment and momentum that was in place (plus strategies for development in the future). As Table 1 illustrates, the scale of the exercise was extensive with 154 universities making nearly 2,000 submissions involving over 50,000 staff. The results were interesting in general terms and very positive for Politics and International Studies as a discrete discipline due to the manner in which more than sixty-eight per cent of the overall research quality of the discipline was assessed as 4* or 3* (i.e. either ‘world-leading’ or ‘internationally recognised’). When the various percentages are combined to produce a grade-point average (GPA) – a simple measure of the overall or average quality of research, which takes no account of the number or proportion of staff submitted -the overall score for the discipline of 2.90 reflects a marked increase on the comparable score of 2.34 in the 2008 RAE. Moreover, nearly all politics departments witnessed substantial improvements on their 2008 scores, with those at Leeds, Strathclyde, Southampton, Westminster and York enjoying the biggest increases. And, on the basis of the discipline’s GPA score of 3.22

in the specific area of ‘impact’, political scientists demonstrated that their research has real-world meaning and relevance.

Table 2. REF2014 Politics and International Studies: Top Ten Institutions.

Ranking	Institution	Overall GPA	%4*
1	Essex	3.54	68
2	LSE	3.45	54
3	Sheffield	3.40	52
4	Oxford	3.31	54
5	UCL	3.28	41
6	Warwick	3.19	43
7	Aberystwyth	3.18	44
8	York	3.16	37
9	Exeter	3.13	30
10	Edinburgh	3.17	38

Not surprisingly, however, a number of alternative formulae have been developed in order to tease out the deeper insights of the basic REF 2014 data. ‘Research power’, for example, relates to issues of scale and provides a measure of volume of research multiplied by quality. The effect, as shown in the first column of Table 3, is to reward the largest departments, with King’s College London jumping to top of the rankings thanks to the 98 researchers submitted to the Politics and International Relations sub-panel. ‘Research Intensity’ takes into account the proportion of full-time staff that were returned to the REF2014 process by a department and therefore attempts to correct for strategic submissions in which a significant number of staff are left out of the audit process.

Table 3. REF2014 ‘Research Power’ and ‘Research Intensity’: Top Ten Institutions.

Ranking	‘Research power’	‘Research intensity’
1	Kings [293.9]	Essex [3.31]
2	Oxford [253.7]	UCL [3.28]
3	LSE [214.8]	Oxford [3.21]
4	Warwick [154.8]	Warwick [3.04]
5	Birmingham [130.9]	Queens (Belfast) [2.84]
6	Manchester [107.0]	LSE [2.83]
7	Essex [106.0]	Strathclyde [2.81]
8	Edinburgh [104.5]	Edinburgh [2.75]
9	UCL [98.4]	Reading [2.74]
10	St. Andrews [96.8]	Royal Holloway [2.68]

Debates concerning the most appropriate or credible way of understanding and presenting research assessment data has almost spawned its own sub-field of political science. The simple facts are that: (1) no research assessment process is perfect; (2) different formulae will inevitably produce different results; and (3) institutions will cherry pick

those interpretations of performance that benefit them the most. But the problem is that the debates go far beyond the analysis and presentation of the results and down into procedural issues of ‘measurement controversy’ over how specific outputs are graded by a REF sub-panel (procedures for the ‘double-weighting’ of books, for example); through to ‘patronage controversy’ over who was appointed to serve on or chair the sub-panels; and ‘sampling controversy’ over who was selected by departments to form part of the submission, and on what basis. Some departments were inclusive and returned almost 100 per cent of staff on the basis of a mixture of collegiality and confidence. In some cases, ‘universal returns’ were the product of a failure of senior staff to make tough decisions and ‘blaming the REF’ became a useful lightning-rod for long-standing institutional weaknesses. In other departments a rather centralized and uniform decision-making system was imposed whereby anyone with outputs that were deemed to fall below the 3* threshold were simply not returned. Under ‘research intensity’ those departments that were more selective fell back down the rankings but the long-and-short of it is that due to institutional selectivity the units of assessment were not being assessed on ‘a like-for-like’ basis as that would have required all units to return all eligible staff.

What then does this largely descriptive account of research events in the UK tell us about the tragedy of political science? The main answer to this question must be that there has been no one singular ‘tragedy’ and it might therefore be more appropriate to explore the existence of an inter-woven set of *tragedies*.

- **Tragedy 1:** The inability of British university leaders to influence, shape, moderate or control the evolution of increasingly bureaucratic research assessment processes since 1986 [*T1*].
- **Tragedy 2:** The manner in which playing the research assessment ‘game’ has arguably become more important than promoting the vibrancy of scholarship itself [*T2*].
- **Tragedy 3:** The failure of political science to utilize the insights of the discipline in order to challenge the imposition of an assessment model that was infused with neo-liberal values [*T3*].
- **Tragedy 4:** The manner in which political science has (and is) going MAD [*T4*].
- **Tragedy 5:** Disappointment in the sense that political science has not developed a ‘new politics of political science’ in order to turn challenges and problems into positive opportunities for the discipline [*T5*].

It is neither possible nor necessary to examine each of these tragedies in turn apart from noting the manner in which some relate to the broader qualities of higher education in the UK and are not discipline-specific (*T1*, *T2*) whereas others are more disciplinary focused (*T3*, *T4*, *T5*). At the broadest level, there is little doubt that the professional representation of higher education to the government (i.e. its ability to speak to power with one clear and loud voice) has been and remains hampered by the existence of a number of university groupings that, in effect, attempt to protect the interests of their members rather than of the university sector as a whole. This was to some extent acknowledged with the creation of the Council for the Protection of English Universities in 2012 but the pressure politics landscape for higher education remains fragmented and therefore diluted. The second tragedy is hard to substantiate in solid, data-driven form but as C Wright Mills

argued, ‘Scholarship is a choice of how to live as well as a choice of career; whether he knows it or not, the intellectual workman forms his own self as he works toward the perfection of his craft... you must learn to use your life experience in your intellectual work’. With this sentiment echoing in my mind I can say with some confidence that a second tragedy, particularly for political science, is the manner in which it has allowed itself to become – whether it admits it or not – a REF-driven discipline. It is far from unique in this regard but there is something slightly more troubling about a discipline that was born with a commitment to engaged scholarship and contributing to the broader health of democracy being so easily and compliantly trapped within an external assessment process. It might be thought that those full-time students of politics who stake their claims to professional respect and credibility on having a sophisticated grasp of both politics ‘in theory’ and politics ‘in practice’ might have been slightly better equipped to shape, respond and in some cases reject some of the external pressures that have been brought to bear. To some extent that professional collective spirit was being undermined by the introduction of a brand of managerialism in which academics and academic institutions were implicitly incentivized to compete and not to share best practice, research capacities, impact networks, etc. That is not to say that the shared public ethos of British universities was or has been wiped away but it is to make a strong argument that it has been eroded and replaced by an ever more aggressive form of ‘gaming in targetworld’.

Phrased slightly differently it could be argued that one of the tragedies of political science is therefore how it has succumbed to a form of professional MADness. MAD in this sense being an acronym for the phenomenon known as ‘multiple accountabilities disorder’ which hollows-out and undermines institutions or disciplines by ensuring that their time is spent accounting to an ever-growing number of political, professional, regulatory and bureaucratic organisations to the detriment of being able to focus on their core and primary tasks. Failure, frustration and disillusionment are therefore almost the guaranteed symptoms of going MAD. The final tragedy is therefore one that focuses on the adaptive capacity of political science in the sense of developing a ‘new politics of political science’ that is vibrant and sophisticated and recognizes both the opportunities and challenges for the discipline presented by the changing contextual landscape. This ‘new politics of political science’ is something that will be discussed in more detail in Part III (below) but the next section explores some of the unintended consequences of the research assessment framework in the UK.

2. Unintended Consequences

The aim of this article is not to offer a polemical critique of the research assessment process in the UK. There is no doubt that the evolution from RSE to RAE and most recently to REF has delivered some positive outcomes in terms of acting as a driver of research quality, delivering greater public accountability and the opportunity to lever new funding resources through partnerships. Within organisations research assessment has also led to the recalibration of resources in order to maximize the value of funding in an increasingly constrained financial environment. Whether this process is viewed as ‘fine tuning’ or crude short-term intellectual engineering is a matter of intense debate but there seems little doubt that there has been an overall upturn in the quality of the UK research base (e.g. articles in the top 1% of citations up from 11% in 1996 to 16% in 2012). And yet to as-

sume an obvious causal link between the introduction of external research assessment processes and these performances based statistical indicators arguably reflects the nature of the problem – the adoption of an incredibly narrow, technical and arguably self-defeating view of scholarship. This is, if anything, the deeper tragedy that risks polluting each and every discipline due to the almost dampening effect that the assessment process can have on what C Wright Mills would call ‘the sociological imagination’ – that intellectual spirit of curiosity and freedom, the ability to trespass across inter-disciplinary and professional boundaries, a belief in the innate value of knowledge and learning without needing to rationalize each and every module against the demands of the economy. When stripped down to a core and basic conclusion the main unintended consequence of the research assessment process has arguably been the imposition (and academic acceptance) of a brand of academic managerialism that is almost designed to squeeze-out intellectual innovation, creativity and flair in favour of a ‘tick box’ ‘REF-return-first’ mentality.

This is, of course, my own and highly personalised view of the impact of research assessment processes in the UK. It could be completely wrong but I would argue that there is sufficient evidence to underpin my position. Indeed, it would be possible to make an even stronger argument and suggest that ‘the politics of the RAE, REF (or whatever it will be called in the future)’ has never been sufficiently exposed in ways that combine to facilitate a fundamental challenge of the process itself. That is not to say that some form of research assessment is not completely legitimate in light of the public funds committed to university research or that such processes cannot have positive outputs and outcomes. It is, however, to suggest that the experience of the UK provides a salutary tale of a process of bureaucratic creep, accretion and sedimentation to the extent that its impact upon research and universities risks becoming dysfunctional – ‘MAD’. The aim of this section is therefore to shed light on the ‘hidden politics’ of research assessment in the UK but in many ways this is just the precursor to a far larger argument about the changing nature of academic life that is made in the next and final section. What then does the ‘hidden politics’ of research assessment look like? What are its main components? Table 4 provides some answers to these questions and the remainder of this section looks at each of them in turn.

The main issue to understand from Table 4 is that none of the themes are isolated issues, they are interwoven into the fabric of the research assessment process and to some extent they are the natural consequences of the imposition of a crude bureaucratic structure upon higher education. Take, for example, Theme 1 ‘bureaucracy’ – the original architects of the Research Selectivity Exercise had no intention of creating a system of assessment that would by 2014 cost universities around a quarter of a billion pounds (£246m to be precise) to administer. That the 2008 RAE imposed an administrative burden of around £66m on universities provides some sense of the manner in which a reform that was intended to increase organizational efficiency and effectiveness has actually spawned a bureaucratic leviathan. And yet to some extent the research assessment process is actually no longer simply about ‘research’; the league tables and rankings that are generated from the assessment process have actually become more like proxies of overall institutional standing that, in turn, are critical in terms of the recruitment of international students and recruiting the very best academic staff. This is a critical point: the politics of research assessment has expanded far beyond research itself.

Table 4. Unintended Consequences: The Politics of Research Assessment in the UK.

Theme	Explanation
1. Bureaucracy	A framework initiated by a government that was committed to 'banishing bureaucracy' has had the opposite effect. The bureaucracy is immense, never-ending and stifling.
2. Shadows	The research assessment process dominates administrative life and decision-making far beyond matters of research. It has taken on a life of its own.
3. Rejection	Not being returned to the research assessment process can have detrimental impacts on both careers and the mental wellbeing of academics.
4. Excellence	The research framework adopts an incredibly narrow, specialised and technocratic view of 'excellence' that is interpreted in having to write in a particularly impenetrable style and to publish in a limited number of the 'right' journals.
5. Impact	Round pegs do not fit into square holes and yet the dominant model of scholarly 'impact' is derived from the STEM subjects and is therefore imbued with simplistic assumptions about clear, direct, visible causality that rarely arise in the social sciences.
6. Gaming	In order to 'succeed' when evaluated against the research assessment frameworks institutions have developed increasingly sophisticated strategies for gaming the system.
7. Silos	Research assessments frameworks in the UK have never coped with the challenge of assessing inter-disciplinary scholarship that, in turn, places those scholars working within this idiom at a disadvantage and creates major disincentives.
8. Equality	There is a politics of inequality that exists at the heart of research assessment frameworks due to the manner in which, for example, career breaks are risky and impact activities have strong gendered dimensions.
9. Over-steers	Research assessment frameworks and league tables have meant that research intensity has increased dramatically within British universities. The cost, however, has been a lack of focus on other elements of scholarship, notably upon teaching.

This flows into our second themes and the notion of shadows (*T2*, Table 4). To some extent Mrs Thatcher's initial foray into increasing central government's grip on British universities was a classic example of the manner in which governance really does take place in the shadow of hierarchy. But the shadows of the research assessment process are particularly long and distinctive in the sense that not only do departments become almost 'REF-driven' to the extent that all procedures and processes are designed to (implicitly or explicitly) feed into a carefully managed REF planning process but that the rules, expectations and standards of the assessment process are to some extent imbibed by those institutions. Recruitment panels do not appoint the 'best' candidate but the 'safest' candidate when assessed through a REF lens; decisions about the use of new resources or funding are rarely taken on the basis of pure unadulterated intellectual ambition but on the basis of providing an evidence base for claims that were either made in the environment statement (i.e. the 'REF5' document within submissions) of the previous assessment or might be made in the next. The sphere of scholarly thinking has, I would argue, narrowed as a direct result of the research assessment processes that have cast an ever greater and direct shadow over the nature of higher education in the UK. There is, so it is said, a silver-lining to all clouds but when it comes to shadows I am told they are completely dark and in relation to research assessment there is a dark side that has received incredibly little open discussion – the impact of rejection (*T3*, Table 4).

What happens if your research is judged to be of an insufficient quality to form part of an assessment return? The formal position has always been that RAE/REF processes are completely separate to institutional promotion systems but the reality is far more com-

plex. Rejection (i.e. Theme 3, Table 4, above) can have significant career implications. As Tables 1, 2 and 3 each in their own ways demonstrate, different universities and departments have come to very different conclusions about the inevitable quality-quantity trade-off that any exercise like RAE/REF inevitably brings with it. The rational actor model would incentivise a unit 'going tight' and putting in the smallest number of staff with the highest perceived quality rating (i.e. focusing down on the narrow GPA score and ranking); however, an equally rational actor model might consider that the short-term gains of 'going tight' did not outweigh the 'long-term' gains of 'going broad' in terms of potential 'research power' and 'research intensity'. But there is another reason for being inclusive in research assessment planning in the sense that 'cutting off a tail' in the sense of rejecting members of staff from a submission is potentially an incredibly divisive decision. Moreover, those staff who do not 'make the cut' (usually at the 3* border) are inevitably likely to face potentially unfair knock-on consequences from this decision. 'If Professor X was not returned at the last REF why should we want to appoint them?' 'If Dr. Y's research was not viewed as being REF'able then on what basis should they really be considered for promotion?' There are, of course, lots of reasons beyond a scholar's relationship with a fairly arbitrary five-year research assessment process should not prevent them either being promoted or moving institutions but – just as occurred at the wider institutional level – it is possible to suggest that an individual's REF status has assumed a far broader significance as a proxy of overall scholarly status.

The problem with this development is that whether the Research Excellence Framework actually identifies and rewards 'excellence' (i.e. T4, above) is a moot point (but one that rarely finds expression in open academic debates). The research assessment frameworks in the UK prioritise and therefore incentivize a very specific definition of research excellence that is generally a narrowly scientific idiom encased in verbosity and jargon and that speaks to a tiny scholarly audience. To publish in the types of scholarly outlets that are likely to be highly prized in the assessment process is to narrow one's focus to a level of hyper-specialisation or methodological masturbation. Peer review is taken as a sign of quality despite the well-known risk-averse, conservative predilections of such processes and in this context single-author books become (ironically) almost risky, especially if claims for double-weighting are rejected. Contributions to edited collections are the intellectual equivalent of *persona non grata* as are generally articles in special editions of journals (due to concerns about the rigor of review processes around commissioned articles on a specific theme). It is therefore with a mixture of great sadness and regret that I cannot help but agree with the argument made by Michael Billig in his book *Learn to Write Badly: How to Succeed in the Social Sciences* (2013). As Chair of the Political Studies Association of the UK I received several informal complaints and request for advice in the wake of REF2014 from academics whose research had been assessed at achieving the 3* standard that was widely used as an institutional boundary for submission but they had still been left out of their University's submission to the Politics and International Studies Panel due to a perception amongst senior staff that it did not quite 'fit' the profile the institution was trying to offer. The research might not have been published in the 'right' journal or might actually be adopting an unorthodox position in relation to major themes and issues. In some cases universities made strategic decisions to submit eligible political

scientists to cognate assessment panels – such as Area Studies or Business and Management Studies – due to a belief that these were ‘softer’ sub-fields in terms of assessment.

The point being made is simple: although research assessment processes have undoubtedly incentivized a strong focus on research and publication within higher education, the definition of ‘excellence’ is arguably fairly narrow. It deifies a specific type of scholarship to the detriment of other equally valid forms of research (a point discussed in some detail in Part III, below). The impact of this – to come to our fifth theme (T5, Table 4, above) – is that scholars who do not or refuse to work within this fairly narrow idiom of highly technocratic impenetrable scholarship are put at a significant disadvantage. And yet what was unique about REF2014 was the inclusion of an explicit component of assessment based upon the non-academic value, social impact or public value of research. In many ways the introduction of an impact component, as demonstrated through the submission of ‘Impact Case Studies’, was an attempt to re-orientate research back towards having some applied, engaged or real-world relevance. The challenge, however, is that the dominant notion of impact was derived from the STEM subjects (i.e. Science, Technology, Engineering, Mathematics) and embraced a rather simple and linear process of *knowledge production* through to *knowledge application* that can be traced and demonstrated through the creation of new products, patents or medications. Only very rarely does the social sciences have such direct and clear impacts on society and yet the research assessment process is almost forcing scholars to either gravitate their choice of research topics to those where demonstrable impact might arise or to play a more dubious game in the sense of making rather doubtful claims about the links between a specific research project and some claim to socio-political change or legislative amendment.

Once more the argument is not that the requirement to demonstrate the non-academic impact of publicly funded research is a ‘bad’ thing – many of the impact case studies submitted to the Politics and International Studies Panel offered convincing narratives of positive social engagement. Nevertheless there is a need to be aware of the risks of politicizing political science by over-incentivising user-engagement around a fairly narrow definition of impact which itself must be linked to a fairly narrow definition of research ‘excellence’. The creation of perverse incentives is actually likely to stimulate a set of strategic responses that political science, notably within the fields of public administration, governance and public policy, has spent several decades studying and warning against. Extensive ‘gaming’ of the research assessment framework is therefore the sixth (T6, Table 4, above) unintended consequence and takes a number of forms from ‘buying in’ research grants and publications through high-level appointments, by closing down departments or units where staff are viewed as never going to be able to play the ‘REF game’, using rolling-temporary contracts to reduce the number of formally ‘eligible’ staff. Other elements of gaming include basing returns not on assessments of research quality but on assessments of how many viable impact case studies a unit has and working backwards from that to assess the optimum of staff that should be submitted (a classic example of ‘a tail wagging the dog’). The selective submission of only a small proportion of staff is one of the most common gaming strategies as is hiring a number of overseas research ‘superstars’ on fractional contracts in order for them to be able to be returned within the hiring institution’s submission. This is generally a very positive development for the overseas scholar who is effectively ‘double-dipping’ in terms of the utilization of their research

but it is bad news for early career researchers who cannot get tenure or even a first step on the professional career ladder. As the next and final section will highlight, one major element of this gaming is that academics whose research is deemed to be *only* 'recognised internationally' (i.e. 2* or less) may be pressurized into accepting teaching-track positions in order to make them illegible for external assessment processes.

These pressures are particularly problematic for scholars who work at the nexus or intersection of different disciplines. The research assessments in the UK have always adopted traditional disciplinary silos as their main tool of sifting and assessing research and this is a major problem for inter-disciplinary or simply less orthodox scholars who wish to range across intellectual landscapes. This is a particular puzzle given the emphasis of the UK funding councils in research years and their emphasis on encouraging interdisciplinary research because those scholars who do actually respond to the signals, take risks and refuse to be intellectual pigeon-holed then find themselves defined as 'high risk' in assessment terms. Therefore the politics of research assessment contains a whole set of embedded inequalities that almost prevent an open, dynamic and inclusive approach to intellectual diversity – at exactly a point in history when such approaches are badly needed. Moreover these inequalities are not just disciplinary. The research assessment process arguably maintains a set of gender-based and ethnicity orientated inequalities that have not yet been the topic of sustained analyses or discussion. Long-standing concerns about political science in the UK in terms of social representation and diversity were to some extent replicated within REF2014 returns. Women were less likely to be returned than men, as were scholars from black or other ethnic minority backgrounds. Male professors submitted more monographs, female professors more co-authored articles. The fragmentary force of external research assessments upon the discipline and upon higher education is the focus of the next and final section but before proceeding to that topic it is necessary to comment upon the ninth and final theme from Table 4 (above) – 'over-steer'.

One of the most important insights from recent experience in the UK is that research assessment has not evolved as *research evaluation exercise*: it has evolved into a powerful incentive system that sets the 'rules of the game' (the meta-governance) that institutions feel they must play. It is not a survey and evaluation of research outputs but has come to signify a proxy rating of institutional excellence. The language and terminology of REF was particularly significant in the sense that it was a 'framework' (i.e. a permanent incentive structure intended to shape the sector) rather than a more isolated or discrete 'exercise' as was the case with the RAE. Furthermore, it could (and has) been suggested that the dominant interpretation of 'excellence' encourages a scholarship of risk-averse mediocrity rather than a scholarship of discovery that challenges foundational ways of understanding the world. But in many ways the introduction of research assessment processes has certainly succeeded in its core aim of encouraging universities to think about the management and governance of research funding. The unintended consequence, however, was a perception that an 'over-steer' had occurred within the sector whereby research became the focus and teaching became almost a nuisance or a distraction, something to be avoided or undertaken at the lowest common denominator in order to maximize research focus.

Unsurprisingly, whether this 'over-steer' has actually occurred and whether students are actually disappointed in the standard of teaching they have received is a contested

issue. My own personal experience over the past twenty years would definitely lead me to support an argument that suggested research was very much the primary focus within the main established universities in the UK. That does not mean that teaching standards were not upheld or that academics did not attain a huge amount of satisfaction from teaching but it is to admit that the *realpolitik* of university life meant that tenure and promotion were driven by research assessments not teaching evaluations. Teaching did not enjoy equal status with research but was almost a second-class endeavour. The perception of the current Conservative government in the UK is certainly that a significant degree of re-balancing is required and in November 2015 the Universities and Science Minister, Jo Johnson, announced plans for the introduction of a Teaching Excellence Framework (TEF). The aim of this new initiative being to ‘build a culture where teaching has equal status with research, with great teachers enjoying the same professional recognition and opportunities for career and pay progression as great researchers’. Not surprisingly the announcement that in future the REF would be partnered by a parallel teaching-focused assessment called TEF was not met with rejoicing in the lecture theatres or seminar rooms. Even students were unconvinced that there was a major problem with teaching standards that required such potentially drastic action. The government has promised to ensure that the TEF is a ‘light touch’ review process but similar commitments were made about research assessment when it was first introduced in the mid-1980s. Moreover, the politics of TEF has links to broader concerns about the impact of REF in terms of increasing central government control over universities and facilitating a market-based managerialist logic. The introduction of TEF is therefore attached to plans to lower the bar to ‘market entry’ in order to allow new universities to emerge; as well as potentially allowing institutions to increase tuition fees where they are assessed to be delivering a particularly high standard of teaching.

The introduction of research assessment in the UK has therefore led to a range of ‘negative externalities’ that have come as no surprise to scholars of public administration or regulatory governance. The *tragedies of political science* – like so many other disciplines and with the universities as a collective institutional endeavor – revolve around a failure to mount a politically astute strategy that may have framed or managed the imposition of these external pressures in a more appropriate, sensitive or proportionate manner. And yet the final argument of this article is that it is very difficult to understand the impact and implications of research assessment without having some broader grasp of how it forms just one element of a ‘bigger picture’ that highlights a set of issues that when taken together focus attention on a potentially catastrophic tragedy for political science in the future. That is the unbundling or unraveling of the discipline.

3. Gaps and Splinters

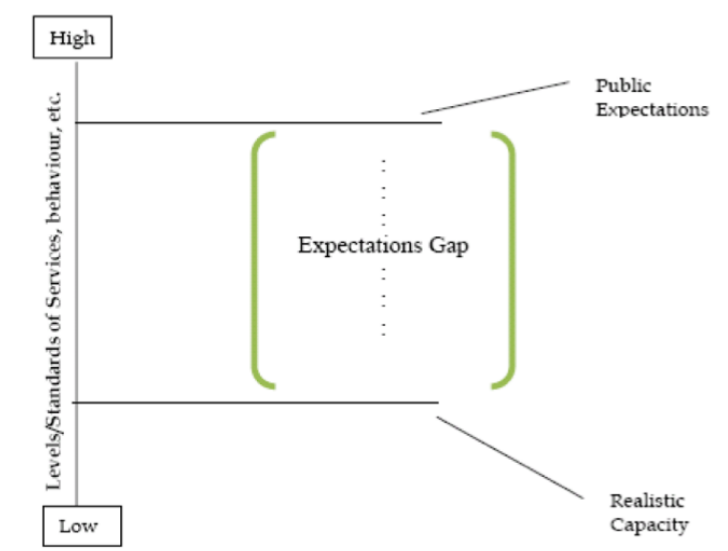
In order to fully understand and expose the politics of research assessment it is necessary to stand back from a focus on tools of research assessment or specific governance frameworks in order to reflect upon how this topic sits within a far broader professional profile. By this I mean the manner in which the nature of scholarship has and is changing and therefore how the notion or meaning of being a ‘University Professor of Politics’ – to use the phrase adopted by Bernard Crick in his ‘Rallying Cry to the University Professors of Politics’ that formed a new part of the second edition of his classic *In Defence of Politics* in

1981 – is also changing. How can scholars understand their role within and beyond academe? Why and in what ways have the professional pressures placed upon academics altered? What can they do to stop themselves going MAD? To make this apparently flip-pant statement about madness is actually to provide a link into an important topic – the rise of mental illness amongst UK academics.

A study of academics discovered that job stresses had increased significantly in recent years and levels of job satisfaction and professional support had declined (see, for example, Kinman and Jones 2010). What this stream of scholarship reveals is that the introduction of research assessments in the UK is just one part of a broader story concerning the gradual imposition of an ever-expanding array of expectations and responsibilities upon university staff. I sometimes find myself envying former colleagues who now enjoy a more leisurely existence as emeritus professors and who held tenured appointments when the pace of academic life was certainly slower.

This ‘slowness’ may well have been exactly what Mrs Thatcher interpreted as a rather over-protected and under-productive charmed scholarly life but my deeper concern rests with the manner in which scholarship is being stretched to breaking-point and one way of understanding and conceptualizing this role expansion is through the notion of an expectations gap (Figure 2 below). As the most eagle-eyed reader will immediately have spotted, this is a rather simple heuristic that is never likely to be judged complex enough for a REF’able piece of work. But the simplest frameworks are often the best and in this regard Figure 2 illustrates how a ‘gap’ might be formed by the variance between the realistic level of capacity given the available resource package (i.e. lower bar) and the public or political expectations placed upon an individual, organisation, community, discipline, etc.

Figure 2. The Expectations Gap



It could be argued that the existence of a small ‘expectations gap’ may well be positive in the sense that it encourages ambition, reflects external confidence, forces institutions to consider innovations and adaptations, etc. And yet the existence of a large expectations gap also risks becoming pathological in the sense that institutional overload and burnout

become real risks. Placed in the context of academe, in general, and political science, in particular, Figure 2 encourages a form of ‘gap analysis’ whereby the demands and pressures placed upon academics and their disciplines (i.e. upper bar) is assessed against some reasonable conception of realistic capacity (i.e. lower bar).

As already mentioned, the breadth of this article in terms of ‘the future of political science’ embraces a broad range of countries, sub-fields and institutions. The pressures on predominantly teaching-only universities or liberal arts colleges, for example, are likely to be very different (but not necessarily *less*) than those facing Ivy League, Group of Eight or Russell Group Universities in the United States, Australia and United Kingdom (respectively). Indeed, the ‘expectations gap’ might be quite different in nature or size in different parts of the world or between different parts of the higher education landscape within a polity. But the simple fact is that from Sheffield to Sydney vice chancellors are increasingly speaking out about the existence of an untenable gap between supply and demand (see, for example, Burnett, 2016).² In this context the options for closing the gap include:

- **Option 1:** Increasing Supply (moving the bottom-bar up);
- **Option 2:** Reducing Demand (moving the top-bar down);
- **Option 3:** A Combination of Options 1 and 2 (closing the gap from above and below)

The argument in relation to the UK is that an ‘expectations gap’ as emerged within British higher education and that this is having a splintering affect upon academic careers that has not been fully acknowledged. The simple position is that over recent decades the upper bar has been pushed upwards without a significant increase in resources. Higher education expansion underlines this claim. In 1950 just 3.3 per cent of young people in the UK went to university; by 1970 the rate was 8.4 per cent; and in 2015 the rate was nearer fifty per cent (over half a million young people taking up a university place). In the 1960s and 1970s small group teaching would generally take place in an academic’s office and involve no more than a handful of students; in the 1990s small groups had expanded to ten or twelve students; and today small groups are often closer to twenty-five or thirty students in number. (The one-to-one tutorial system that has been at the heart of Oxbridge teaching system for centuries is under increasing financial strain.) One early impact of the TEF is that universities have engaged in almost a bidding war to increase levels of teaching contact time for students that will have obvious knock-on consequences for staff research capacity. The research assessment processes therefore form just one element of this gradual process of role accretion or sedimentation. Take, as a starting point, the five main components of an academic position in a British university:

1. **Research:** As displayed through international peer-reviewed publications and significant external research grant income.
2. **Teaching:** Evidence of excellence in teaching as displayed through student feedback and external audit processes.
3. **Administration:** The capacity to undertake significant administrative and managerial responsibilities within and beyond your home department.
4. **Impact:** The ability to demonstrate that your research has achieved a clear, direct and auditable ‘impact’ on non-academic research-users and/or the public.

5. **Citizenship:** A clear contribution to professional ‘good citizenship’ through activities such as journal editing, external examining, pastoral responsibilities, government or parliamentary service, leadership of learned societies, etc.

To undertake world-class 4* research, to demonstrate ‘excellence’ in relation to the teaching of evermore demanding students, to successfully apply for competitive research funding and fellowships while also managing an ever-increasing bureaucratic burden... while also delivering ‘impact case studies’ that could withstand almost forensic analysis as to their veracity *and* showing evidence of professional engagement beyond your own university. The new work demands in higher education are possibly becoming untenable and to some extent the splintering or fragmentary effect that forms the focus of this final section is a fairly obvious consequence. For those readers who think that it is *me* that is over-inflating the contemporary situation in the UK it is worth thinking in a little more detail about the expectations placed upon early career researchers in political science:

- To trespass across disciplinary and professional boundaries while also displaying increased hyper-specialisation;
- To enjoy ‘academic autonomy’ and ‘intellectual freedom’ in an increasingly directive and constrained environment;
- To increasingly engage with quantitative methods and ‘big data’ while also producing nuanced, accessible and fine-grained analyses;
- To manage the temporal misalignment between academe timescales and politics in practice;
- To be able to ‘talk to multiple publics in multiple ways’ while acknowledging a constant pressure to ‘tech-up’ within political science;
- To cope with a system where the incentive structure still pushes scholars towards ‘pure’ scholarship and peer reputation rather than ‘applied’ scholarship or public reputation;
- To navigate the problematic relationship between facts and values, and the prevailing rhetoric of neutrality in research;
- To innovate and share ‘best practice’ while also working in a competitive market environment;
- To deliver world-class research and writing while also providing excellence in teaching;
- To provide a personalized student-centred learning experience in a climate of mass and often digitally refracted access;
- To take risks in what is generally a risk-averse professional environment;
- To balance a traditional focus on ‘problem-focused’ political science with external demands for ‘solution-focused’ political science;
- To ensure that research informs public debate without being ‘dumbed down’ or co-opted by partisan actors;
- To be responsive to ‘students-as-customers’ while upholding academic standards and relationships; and
- To achieve some notion of a personal, private or family life while fulfilling the demands of the role.

Turning back to the focus of Part II (above) what has in reality occurred in the UK in recent years is a REF-driven focus (bordering on obsession) with research as the primary component of an academic role. The TEF is therefore an attempt to rebalance higher education back towards teaching while both TEF and REF will inevitably increase the administrative burden on individuals and departments. It would at this point be possible to locate this shift in the context of Ernest Boyer's 'Taxonomy of Scholarly Endeavour's' but I have done this elsewhere (Flinders, 2017) and the real focus of this section is on professional splintering as both a gaming strategy and a personal coping strategy. What I mean by this splintering is that the notion of an 'all rounder' scholar who undertakes research, teaches and plays a leading role in the administration of either their department or their disciplinary learned society is eroding and is being replaced by an increasingly fragmented community of political scientists – the modern 'specialist scholar'.

Traditionally British universities have maintained a broadly egalitarian approach whereby all staff are expected to undertake at least some element of teaching and administration. The exception to this was generally where staff had secured teaching 'buy outs' through external research grants but in the last two or three years a bifurcation between teaching-only and research-only staff is beginning to emerge. Between and betwixt these two extremes exists an increasingly large academic 'precariat' consisting generally of younger new entrants to the profession who are expected to accept either a succession of temporary (and generally teaching-focused) contracts or to undertake an even more precarious academic existence on the basis of a portfolio of fractional roles undertaken concurrently at several different universities. Escaping 'the precariat' revolves around securing tenure but even here a professional pathology exists in the form of a pressure to 'publish or perish' that inevitably incentivizes a combination of hyper-specialisation and self-plagiarism. This, in turn, does little to nurture intellectual ambition and even less in terms of building confidence amongst non-academic user-groups that political science has the capacity to respond to allegations of irrelevance. The flip-side is that exploring new approaches, developing new theories, demonstrating relevance or public value, investigating the nexus between disciplines, etc. – all of those main activities that funders, researchers and governments around the world prioritise – demand time and the acceptance of positive inefficiencies (e.g. risks that do not pay off, roads to relevance that turn out to be *cul de sacs*, etc.). The contemporary tragedy of political science – to paraphrase Ricci (above) – is double-edged: the young fresh minds with the most to offer are immediately squeezed into a system that could have been designed to squeeze-out ambition and creativity and incentivizes 'playing safe'; while the profession as a whole offers little space for positive inefficiency, no matter how positive the returns might be.

4. Conclusions, and few suggestions

Those with an awareness of very recent shifts within British higher education might respond that my analysis is out-dated. 'Doesn't he know that 'publish or perish' has been replaced by 'quality over quantity'? I hear them cry. This is certainly the new mantra amongst vice chancellors and deans but the reality beneath this rhetoric is a professional sphere in which very few academics are brave enough (or have the intellectual headspace) to step-off the publication production line. And yet at the other end of the spectrum it is possible to identify the recent emergence of a very small cadre of tenured 'high impact'

academics who enjoy a visibility within the practitioner and media spheres. The ‘stretch’ or ‘span’ of an academic career has therefore widened significantly in response largely to the imposition of external audit regimes and higher expectations. The malleability of some institutions has reached breaking point and this is reflected in the manner in which some teaching focused universities have dropped out of the REF process and some research-focused universities are threatening to boycott the forthcoming TEF process (see Havergal, 2016). And yet my sense is that this fragmentation appears to be locking-in rather than challenging a number of pre-existing inequalities within the discipline. For example, the research professors and ‘high-impact’ professors generally make little contribution in the sphere of institutional or academic governance and undertake little (if any) teaching. They are also generally men.

The real tragedy of political science (or a central tragedy for political science) is that it has so far failed to acknowledge the existence of the politics and management of this expectations gap surrounding scholars, or to acknowledge its splintering dynamic which leads me to suggest that the future of political science depends upon the emergence of ‘a new politics of political science’ that seeks to control and manage external pressures – to somehow close the expectations gap – for the collective good of the discipline. This would involve a new professionalism that permeates down from national learned societies, professional associations and funders, through institutional units and to individual scholars. That is a *new politics* that is – quite simply – more aware of the external context in which sciences takes place and that balances internal expertise and external engagement. More specifically the nexus *between* academe and society must form the focus of greater attention and, as a result, the role of an academic is likely to change. As the Brexit debate in the UK illustrated, politicians will always ignore or seek to reinterpret research that does not suit their partisan needs but there is a far wider community of potential research users than the discipline generally recognises. The dominant perception of a clear qualitative distinction between ‘pure’ and ‘applied’ research will have to be re-cast in a more dynamic mode of understanding. More specifically, there will have to be some understanding of the manner in which ‘impact’ can actually underpin, nourish and nurture excellence in terms of both research and teaching. Once again, the ‘new politics’ or ‘new professionalism’ will have to understand the knowledge ecosystem in ways that have largely been forgotten but must now be rediscovered if the discipline is to prosper. The exact nature of this new disciplinary strategy will be for national associations and institutions to decide, but in terms of offering elements of this ‘new politics’ the following ideas are worthy of consideration.

Firstly, political science cannot and should not adopt a victim mentality but a more robust and confident professional persona. In this regard, the role of the main learned societies is vital as the source of external promotional activities and more specifically as the driver of proactive *knowledge-brokerage*, *knowledge-filtering* and *knowledge-framing* activities. Put within the framework of Figure 1, the role of learned societies and professional associations has to support the discipline in terms of raising the lower bar of realistic capacity where possible while paying far more attention to their external/strategic role in actually managing the expectations of the public and policy makers vis-à-vis the upper bar (i.e. Option 3, above). Simply stated, learned societies and professional associations must take the lead in closing the expectations gap from above and below. In this regard, relatively simple steps can yield significant returns. Of particular

significance, for example, given the temporal misalignment between academe timescales and politics in practice is a clear approach to horizon-scanning so that translated packages of research can be prepared and delivered to research users (media, practitioners, etc.) at specific ‘windows of opportunity’ when the demand for such information will be high. Moreover, learned societies, in partnership with funders and research-users, should also take the lead in terms of innovating in relation to both training and bridging activities. Take, for example, the Political Studies Association of the United Kingdom’s annual *Total Exposure* competition.¹ It would be almost impossible to design a more simple initiative: academics receive support, training and guidance on how to ‘pitch’ an idea to broadcasters based around translating their research into a documentary or series of documentaries for television or radio. Academics can submit ideas on their own or in small teams, interdisciplinary ideas are encouraged and the overall emphasis is on creativity and intellectual energy. A panel of senior commissioning executives then sifts applications and selects twelve finalists who are then invited to London to make their pitches in person during a face-to-face sixty second slot in front of the broadcast specialists.

There is no ‘prize’. No broadcaster is ever going to guarantee to commission a project through an open competition. But what *Total Exposure* does achieve is an opportunity for academics to learn new translational skills and to expose themselves – personally and intellectually – to a new professional audience who approach the value of scholarship from a very different perspective. Three things are worth noting about *Total Exposure*. First, it has proved to be an incredible success. Of the twelve pitches shortlisted in 2016 nine received ‘call backs’ to discuss their ideas in more detail with commissioners and one pitch was taken straight into production (Cathy Gormley-Heenan’s documentary on ‘the politics of peace walls’ around the world); in 2017 eight of the nine short-listed pitches received call-backs and several look likely to move into production. Secondly, just like politics a lot of the real work takes place not within the sixty-second pitch or the subsequent discussion but in the coffee breaks and over lunch. The commissioners often have ideas for new programmes and are looking for new faces, new voices and new talents with the capacity to engage, inform and entertain in equal measure. Rejected pitches may well lead to unexpected opportunities at a later date. The final twist of *Total Exposure* takes us back to the issue of equality and diversity and flows into a set of debates concerning demographic change. Younger scholars, women and individuals from black or ethnic minority backgrounds have dominated the list of finalists. As such, the social composition of the short-listed candidates tends to be far more representative of society at large and therefore decidedly unrepresentative of the political science community in the UK. *Total Exposure* therefore not only takes the very best social and political science and translates it for dissemination through mass access broadcasting platforms but it also appears to have somehow short-circuited some of the traditional professional blockages that prevent equality of participation and opportunity.

Put slightly differently, projects such as *Total Exposure*, led by the national learned society, begin to add tone and texture, even substance, to a ‘new politics of political science’ that is founded on an understanding of the manner (1) the discipline has evolved to contain and sustain significant structural inequalities, (2) that these inequalities cannot

¹ See also: <https://www.psa.ac.uk/totalexposure>.

be ignored and that (3) each of Boyer's forms of scholarship are mutually supportive and combine to sustain a rich intellectual ecosystem.

A second element of this 'new politics of political science' might take this more ambitious, coherent and holistic approach one step further through a generational approach to student recruitment that moves the focus down the educational pipeline so that students in schools and colleges appreciate exactly what the study of politics involves and why it matters, its potential in both intellectual and vocational terms and the available professional career paths via higher education. This educational pipeline provides a critical tool through which to understand and address long-standing issues concerning diversity and inequality and – beyond this – to democratise the study of politics to exactly those sectors of society who appear to have become disenchanted. Scholars in the field of political (dis)engagement have for some years outlined a shift in modes of political expression and activity *from* traditional party-based, mass member, formalized, etc. [i.e. 'old' modes] *towards* more individualized, issue-based, direct, digital and informal 'new' modes. But political science has arguably failed to utilize these insights when it comes to proactively promoting or demonstrating the value of their discipline. School 'outreach' events therefore tend to continue to be held in the traditional institutions of politics – the city halls and parliaments – but rarely exhibit the creative dynamism that young people crave by 'reaching-out' within exactly those new political arenas, like music, film or literature festivals, where debates, discussion and recruitment takes place. Even the language of politics needs to be considered within this new politics of political science. 'Outreach' and 'reaching-out' arguably bring with them subtle but subliminal connotations, the former somewhat cold, formal, distant (exactly those characteristics that 'disaffected democrats' level at politics) the latter perhaps far warmer, friendly, engaging.

A third element is highly political and involves the colonisation of the broader research community in terms of places on the boards of research bodies, government advisory bodies, international non-governmental organisations, media organisations, etc. My sense is that other disciplines have been far more professional and ambitious in terms of monitoring when places on influential organisations are advertised and then encouraging (and supporting) members of their discipline to apply. This allows the discipline to be embedded and have tentacles far beyond the university sector and to have ambassadors in key posts. Once again, this regular vacancy monitoring and proactive encouragement is fairly low cost but potentially incredibly important for the external profile and visibility of a discipline. The targeting of professional appointments can also be built into a more ambitious equality and diversity agenda, while also being of value to the individual academic in terms of their 'good citizenship' requirements and the need for impact-related or research-related networks. (This targeted approach to recruitment also works in the opposite direction in the sense that professional associations and learned societies might also usefully include a number of non-academic research users on their boards.) What these three elements really point to is the manner in which the 'scientific' and the 'political' (or the 'academic' and the 'public') components are both mutually inter-dependent – almost positively parasitical in the sense that they feed upon each other – within a modern academic career where the professional responsibilities of academics to the public who fund their work are increasingly explicit. In this regard claims to be delivering more research of a higher quality will carry little weight if that research does not percolate through

into the public sphere in accessible and purposeful ways. Without this ‘new politics’ political science will be politically disadvantaged (and therefore structurally disadvantaged in resource terms) *vis-à-vis* other disciplines in a climate of already shrinking resources. That really would be a tragedy.

References

- Almond, G. (1990) *A Discipline Divided*. New York: Sage.
- Barrow, C. (2008) ‘The Intellectual Origins of New Political Science’, *New Political Science*, 30:2, 215-244.
- Burnett, K. (2016) ‘Cast starved campuses must raise fees or drop standards’, *Times Higher Education*, 1 September.
- Capano, G and Verzichelli, L. (2016) ‘Looking for Eclecticism? Structural and Contextual Factors Underlying Political Science’s Relevance Gap’, *European Political Science*, 1-22.
- Capano, G and Verzichelli, L. (2010). ‘Good but not good enough: recent developments in political science in Italy’, *European Political Science*, 9:1, 102-107.
- Easton, D (1969) ‘The New Revolution in Political Science’, *APSR*, 63:4, 1051-61.
- Farr, J. (1988) ‘The History of Political Science’, *American Journal of Political Science*, 32:4, 1175-1195.
- Flinders, M and Kelso, A. (2011) ‘Mind the Gap’, *British Journal of Politics and International Relations*, 13(2), 249-268.
- Flinders, M. (2013) ‘The Tyranny of Relevance and the Art of Translation’, *Political Studies Review*, 11(2), 149-167.
- Flinders, M and Dommett, K. (2013). ‘The Politics and Management of Public Expectations’, *British Politics*, 9:1, 29-50.
- Flinders, M. (2013) ‘The Politics of Engaged Scholarship’, *Policy & Politics*, 41:4, 621-642.
- Flinders, M Savigny, & K Awesti, A. (2016) ‘Pursuing the Diversity and Inclusion Agenda’, *European Political Science*.
- Ginsberg, R. (1999) ‘Conceptualizing the European Union as an International Actor’, *JCMS*, 37:3, 429-454.
- Hay, C. (2009) ‘Academic Political Science’, *Political Quarterly*, 80:4, 587.
- Havergal, G. ‘Some Russell Group Universities ‘Could Opt Out of TEF’, *Times Higher Education*, 1 September.
- Kinman, G and Jones, F. 2010. ‘‘Running Up the Down Escalator’ stressors and strains in UK academics’, *Quality in Higher Education* 9(1), 21-38.
- King, G Schlozman, K and Nye, N. eds. (2009) *The Future of Political Science*, London: Routledge.
- Pielke, R. (2003) *The Honest Broker: Making Sense of Science in Policy and Politics*, Cambridge: Cambridge University Press.
- Schmitter, P. (2002) ‘Seven (disputable) theses concerning the future of ‘transatlanticised’ or ‘globalised’ political science’, *European Political Science*, 1:2, 23-39.
- Toje, A. (2008) ‘The Consensus-Expectations Gap’, *Security Dialogue*, 39:1, 121-141.
- Trent, J. (2011) ‘Should Political Science be more Relevant?’ *European Political Science*, 10, 191-209.

Two Decades of Political Science Research Assessment: the Dutch Experience

Rudy B. Andeweg

LEIDEN UNIVERSITY

A High Level of Acceptance

Today, there is a general acceptance of, or at least resignation about, Dutch Research Assessments, whether in political science or in other disciplines. Research assessment exercises started in the Netherlands in 1993, and are held every six years. To a large extent, research assessments are a non-issue. In comparison to the experience of political scientists in many other countries, this may seem surprising, but several factors help account for this counter-intuitively high level of satisfaction.

The single most important factor underlying the acceptance undoubtedly is the simple fact that the role of the government in organizing, administering and supervising the assessments is marginal. A recent report by an independent think tank concluded that nowhere in Europe is the involvement of the government or other state actors as minimal as it is in the Netherlands (Van Drooge et al. 2013). The universities alone are responsible for the assessments. The *Standard Evaluation Protocol* (SEP), which outlines the aims and procedures of the research assessments, has been developed by the Dutch Association of Universities (VSNU) together with the Dutch Science Foundation (NWO) and the Royal Academy of Sciences (KNAW), organizations that are beyond the direct control of the government. The introduction to the latest edition of the Standard Evaluation Protocol (2014) mentions that it was presented to the Minister of Education, but merely out of politeness. Neither the Minister nor her civil servants had been involved in setting the evaluation criteria, and even the obligation to send a copy of each completed assessment exercise to the Education Minister has been dropped several years ago.

The universities define the research units that are to be subjected to an assessment exercise; each university decides whether its research units will be assessed in a stand-alone exercise, or whether they will be part of a nation-wide comparative assessment of research in that particular discipline. The most recent Political Science Research Review (Verdun et al 2014), for example, did not include the Department of Political Science at Radboud University Nijmegen, because that university had opted for a stand-alone assessment of its political science research programme. The universities decide on the composition of the peer review committee that will conduct the assessment, as long as it is an international committee and its members have no conflict of interest with any of the departments, and often the university executives will delegate the search for committee

members to representatives of the departments concerned. The universities also provide logistic and administrative support to the assessment committees, and through the Dutch Association of Universities they have set up an independent agency QANU (Quality Assessment of Netherlands' Universities) which specializes in offering such support. It is fair to conclude that the Dutch Research Assessments are free from governmental interference.

A second reason for the general acceptance of the research assessments is that they hardly have any direct consequences for the scholars whose work is evaluated. To some extent this is related to the lack of government interference. The government could still use the reports, which are made public, to shape its funding decisions, but it does not. Even the universities do not attach direct consequences to the assessment outcomes. Doing so would contravene the twin aims of the assessment exercises: accountability for the use of taxpayer money, and improvement of the research units involved. These aims are explicitly stated by the universities themselves, which limits their ability to punish a research unit for poor assessment results by reducing funding or closing down departments. The only direct consequence that I have been able to find is for the accreditation of Research Master Programmes. In the Netherlands, Master programmes in all but a few disciplines are one year programmes. Ministerial permission is required for the start of a two-year Research Master catering to selected talented students, primarily potential PhD candidates. The Minister bases such decisions on the recommendation of (re-)accreditation panels, and one of the criteria used is having obtained high scores in the most recent research assessment exercise.

There are more indirect consequences. Departments take the research assessments very seriously because they affect their reputation. Getting a bad evaluation, or even a good evaluation that is significantly below the evaluations of other departments in the same discipline, has a negative effect on the department's reputation, which is feared to weaken a department's potential to recruit good PhD candidates and faculty, and to weaken its potential to receive research grants from the science foundation. Still, it would seem that the absence of direct sanctions helps explain the relative satisfaction.

In the Netherlands, there is a parallel scheme for the assessment of teaching quality, and there seems to be more concern about the nature and aims of those reviews. In any given six-year cycle, most departments will be evaluated twice, once on the quality of their research, and once on the quality of their teaching. Although the teaching quality assessments are also organized by the universities themselves, the reports are used by the Minister of Education and her Inspectorate. In 1994-1995, such an assessment report was used by the Inspectorate and the Minister to threaten to withdraw the accreditation of the Bachelor programme in political science at Radboud University Nijmegen – a threat that was lifted only after the University promised major reforms. Moreover, the outcomes of the teaching quality assessments are used by others, including commercial publishers, who draw up rankings of Bachelor programmes to aid prospective students in choosing which university to go to. As the funding of universities, and of departments within universities, is largely determined by student numbers, a poor teaching quality assessment may have immediate effects on the intake of students, and thus on the funding, of departments. So the immediate consequences of the teaching quality assessments are much more important than those of the research quality assessments.

Criticisms

The fact that research assessments are hardly controversial in Dutch academia does not mean that there are no criticisms of aspects of the assessment exercises. Some of the criticisms have led to adaptations in the regularly updated Standard Evaluation Protocol, but on others the process has been less responsive.

Administrative burden

A major complaint refers to the administrative burden. For each assessment, a department has to hand in a self-evaluation report. Such a report should contain quantitative information on the research input and output, conforming to very specific standardized criteria. Occasionally this requires collecting new data or transforming existing data to meet the Standard Evaluation Protocol's criteria – for example when a university employs different definitions of peer-reviewed/non-peer-reviewed publications, or national/international publications for its internal use. In addition, the self-evaluation report should contain a qualitative reflection by the department of its own research policy, publication strategy, etc. This should be presented in the form of a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis, and often prompts lengthy deliberation on finding the right balance between being honest and being strategic: being very honest makes it easy for the assessment committee to expose the department's weaknesses; being too strategic may prompt the committee to distrust the self-evaluation and to dig deeper itself.

An often used strategy to deal with this dilemma is to be quite honest about the department's weaknesses, but to start reforms to address these weaknesses just before the committee arrives for its site visit. The site visit itself is disruptive for a department, but it is brief. More work awaits the department after the assessment report has been published, as most university administrations will request a follow-up report from the department to show what will be done with the committee's recommendations. Moreover, most universities fear the effects of a negative assessment on their reputation, and require departments to organize a midterm assessment themselves in order to be able to address any vulnerabilities before the real assessment takes place. Although this is less burdensome than the real research assessment, it still requires compiling a self-evaluation document and discussing it with an external assessor, usually a trusted colleague from a university outside of the country.

Given the fact that assessments of research and teaching quality follow quite similar procedures, most departments have to write two self-evaluations, two follow up reports, and organize two midterm assessments in any given six-year cycle. Nothing has been done to alleviate this administrative burden.

The Improvement/Accountability Dilemma

As mentioned above, the stated aims of the Dutch research assessments are accountability and improvement. These aims are not contested, but in practice they are difficult to reconcile. In terms of accountability it is necessary that the assessment reports are given wide publicity, and include the evaluations of all research units in a given discipline. This makes it easy for the interested taxpayers to see what was done with their money. But such public and comparative reports may lead to posturing by departments rather than to frank SWOT analyses in their self-evaluation reports. Such reticent self-evaluations will hamper

assessment committees in identifying weaknesses and developing useful suggestions for improvement. Since 2003, universities are allowed to organize stand-alone research assessments, and an example of such an assessment exercise in political science was mentioned above. Even if such non-comparative assessments are made public, they do not attract the same amount of attention that the comparative reports attract. It could well be argued that stand-alone reports are preferable in terms of searching for improvement of research quality as there is less need for a department to act strategically. However, this comes at a cost in terms of accountability. Moreover, withdrawing from the national and comparative research assessment exercise is generally interpreted as an admission of weakness by one’s colleagues. Nevertheless, the number of stand-alone research assessments has increased considerably. Across all disciplines 222 research assessments have taken place between 1994 and 2012, 136 of which were confined to just one university or research unit (Van Drooge 2013: 7). In political science, with the exception mentioned, comparative assessment exercises are still the norm.

One size fits all?

Originally, the assessment protocols made no allowance for differences between disciplines. The assessment criteria were largely based on what was customary in the technical and natural sciences. Research assessments were not alone in having this bias towards a publication culture that favours journal articles over books, English-language over Dutch-language publications, and multi-authored over single-authored publications. This bias has had a marked impact on the publication culture within political science. Gradually, however, the protocols allow for greater variety and fine-tuning to the needs of the discipline being evaluated. In the most recent political science research assessment, for example, it was decided to use bibliometric data from Google Scholar rather than Web of Science, as the first has a better coverage of political science publications than the latter.

A recent report of the Royal Netherlands Academy of Arts and Sciences advocates to find a balance between uniform assessment criteria and taking into account the variety within the social sciences, by adopting a simple 2×3 table of assessment categories, and leaving it to each discipline to fill those categories with indicators that are relevant to that discipline (Bensing et al. 2013).

		Quality domains	
		<i>Scientific quality</i>	<i>Societal relevance</i>
Assessment Dimensions	<i>Output</i>	Output regarded by peers as of outstanding quality	Output for external target groups
	<i>Utilization</i>	Utilization by peers of researcher’s output	Utilization by external target groups
	<i>Recognition</i>	Recognition by peers Researcher’s output	Recognition by external target groups

It is too early to say whether this recommendation will be implemented and assessment criteria will be furthered tailored to the publication culture and the specific needs of political science and the other social sciences.

The Problem of Proxies

Research quality is a largely subjective concept for which no clear and generally accepted indicators are available. As a consequence, all indicators that are used in assessment exercises are proxies, and usually proxies of a quantitative nature: the number of publications, citation scores, the amount of external research funding, etc. There is increasing dissatisfaction with such quantitative criteria that almost by definition imply that ‘more is better’. The concern is that it will lead to strategic behavior: mutually adding colleagues as coauthors so that all members of the department have more publications. In at least one Dutch political science department it has become the rule that the PhD supervisor is automatically listed as a coauthor of all publications of the PhD candidate. This led the most recent Assessment committee to conclude that ‘there are questions for each of these Institutes about whether PhD candidates in their Programmes should publish together with their supervisors (and if so whether those publications should form part of their dissertation work)’ (Verdun et al. 2013: 13).

Here too there has been some responsiveness to those concerns. Research units are asked to list what it considers its five best publications over the past six years, and assessment committees are expected to read them, although it is not always clear from the report that the committee actually did so. Of the four quality indicators used so far: (scientific quality, scientific productivity, societal relevance, and viability), the most quantitative indicator – productivity – has been dropped, and research integrity has been added.

Outcome inflation?

Although the research units that are assessed do not complain, it is perceived by policy-makers as a problem that the average scores that are used to summarize a department’s research quality have gone up over the years, leaving very little variation between the research units that have been assessed. So far, the scores have been expressed on a scale of 1 to 5. On the indicator of quality, for example, the average score went up from 3.65 in the first assessment cycle in the 1990s to 4.39 in the most recent 2009-2015 cycle (Van Drooge 2013: 10). Cynics might surmise that this increase is correlated to the increase in stand-alone assessments, but a comparison between the average scores used in comparative and in stand-alone assessments shows that this is not the case.

In the most recent Political Science research assessment (Verdun et al. 2014) the variation in scores across departments is indeed small:

	Quality	Productivity	Relevance	Viability
Leiden University	4.5	5	4	4.5
Amsterdam University	4.5	4	4	4
Free University Amsterdam	4.5	4.5	4.5	4.5
Twente University	4	4	5	3.5

In response to what is perceived as ‘score inflation’, the scale has been redefined several times. From 1=poor, 2=unsatisfactory, 3=average, 4=good and 5=excellent, to 1=unsatisfactory, 2=satisfactory, 3=good, 4= very good and 5=excellent. In the next round the scale will be reversed and range from 4=unsatisfactory, 3=good, 2=very good, to 1=

world leading. It is hoped that such changes will also produce more variation in the scores awarded to various research units.

However, it is not clear whether the higher and more homogeneous scores indeed reflect score inflation. After all, it is one of the explicit aims of the research assessments to help improve the research quality at Dutch universities. If, after over twenty years of research quality assessments, quality would not have improved, this would not reflect well on the utility of the whole exercise. Similarly, as the room for improvement was greater for departments that started out with relatively low scores, it should not come as a surprise that there is less variation two decades later.

Impact

As they hardly have any direct consequences, it is not possible to measure the impact of the research assessments. Moreover, the introduction of research assessments in the early 1990s was but one element in the general professionalization of political science in the Netherlands. This professionalization was not only imposed from above by research assessments, by reducing the income that universities receive from the state directly, making them more independent on the competition for external research funding, etc., but it has also been initiated from below, by political scientists who sought to maintain or strengthen their reputation in an increasingly international environment. A recent overview of the development of Dutch political science is entitled ‘from politicization to professionalization’ (Andeweg & Vis 2015), and describes how professionalization has also been a reaction to political scientists growing tired of the ideological conflicts that plagued some of their departments (the two universities in Amsterdam and Nijmegen university in particular) from the 1960s to the 1980s. In that light, the undoubtedly positive outcome of professionalization and internationalization can only in part be attributed to the research assessments.

The other side of the coin is that the downside of professionalization and internationalization can also be blamed only partially on the research assessments. One of these negative side effects is the shift in the publication culture towards co-authored English-language articles in peer-reviewed journals. There are no intrinsic reasons for this shift from books to journals and for the increase in the average number of coauthors. It has less to do with increasing quality than with succumbing to the temptation to measure research quality by readily available bibliometric indicators. We have allowed ourselves to be taken hostage by a commercial firm: Thomson Reuters and its Social Science Citation Index!

The trend to publish more internationally, i.e. in English, does not have only negative consequences. After all, an English language publication is accessible to a much wider readership than a publication in Dutch, which brings a higher level of scrutiny and debate. This can only have beneficial consequences in terms of research quality. However, the shift in publishing from Dutch to English, and the higher threshold to readers because of the more sophisticated methodology used, has also meant that political science plays a significantly less prominent role in public debate in the Netherlands: science for science, rather than science for society. In the media, we see that historians and constitutional lawyers increasingly replace political scientists when journalists need expertise to explain current events.

The changes that already have been made to the Standard Evaluation Protocol, and the further changes that have been advocated, can be seen as efforts to address the negative effects of professionalization and internationalization: less emphasis on productivity and more attention to research integrity may help stop some of the strategic publishing choices that have emerged, and more attention to societal relevance may induce political scientists to invest in contributing to the domestic public debate by – also – writing in Dutch and for a wider public. We shall see: the next assessment of research quality in political science is scheduled for 2019.

References

- Andeweg, R.B. and B. Vis (Eds) (2015), *Van Politisering naar Professionalisering; Politicologie in Nederland*, Oudewater: NKWP, 126 pp.
- Bensing, J. , R. Andeweg, Ph. Franses, B. Meyer, C. Prins, K. Schuyt (2013), *Towards a Framework for the Quality Assessment of Social Science Research*, Amsterdam: Royal Netherlands Academy of Arts and Sciences, 43 pp.
- Standard Evaluation Protocol 2015-2021; Protocol for Research Assessments in the Netherlands* (2014) Association of Universities in the Netherlands (VSNU), Royal Netherlands Academy of Arts and Sciences (KNAW), Netherlands Organization for Scientific Research (NWO), 32 pp.
- Van Drooge, L., S. de Jong, M. Faber and D. Westerheijden (2013), *Twintig Jaar Onderzoeksevaluatie; feiten & cijfers*, Rathenau Instituut (www.rathenau.nl), 19 pp.
- Verdun, A., D. Farrell, O. Gabriel, C. Hay, H. Heinelt, K.E. Jørgensen, M. Kenny (2014), *Research Review Political Science 2007-2012*, Utrecht: Qanu, 69 pp.

The French HER system and the issue of evaluation

Catherine Paradeise

UNIVERSITÉ PARIS EST-MARNE-LA-VALLÉE

1. A historical overview

The French higher education and research system is based on a double divide of missions, statuses and recruitment procedures. On one side, over 30 public research organizations – whose permanent scientific staff varies from about 300 to over 12.000 – differentiate themselves from higher education (HE) institutions. On the other, the HE sector itself is made of about 80 universities and about 200 *grandes écoles*.¹ The structure of the system results from the French history of universities, which were suppressed as territorial entities then recreated as a collection of disciplinary ‘faculties’ co-managed from Paris and settled in about 13 big cities during the Napoleonic times. *Grandes écoles* (named *schools* below) were created at the turn of the nineteenth century to educate state engineers on a very selective basis. A school of public administration was added to the list in 1945. Business schools were added in the seventies. Research organizations were built up after the Second World War to face the weaknesses of both types higher education institutions in research. Since the 1990s, various reforms have first incrementally contributed to integrate education and research in both schools and universities, then radically pushed towards building consortia or even entering mergers between higher education institutions.

Although interrelations between these poles developed over time and job contents to a certain extent became more similar, this historical divide remains. Radical reforms occurred at the turn of the twentieth century, which drew these institutions and their evaluation systems nearer to each other but, although they experienced some convergence, they still remain institutionally distinct.

Regarding research units located in higher education institutions, no formal assessment took place outside the “associated research centers” (URA) between universities and CNRS, which emerged in 1965 and largely muted to stronger partnership in “joint research centers” (UMR) between *Centre national de la recherche scientifique* (CNRS) and HE institutions (UMR account for about 90% of all CNRS research centers at the turn of the 2000s). Such units were subjected to a four-year assessment according to CNRS procedures. No other assessment was required, except by ad hoc committees for individual

¹ This number refers to the *grandes écoles* that are accredited by the *Conférence des grandes écoles* among over 400 which deliver post-baccalaureate education.

scholars applying for recruitment and promotion, or public research call for tenders. This state of affairs changed with the radical reforms of the 2000s.²

For its part, not much has changed in the evaluation of teaching. Although rules changed several times since the sixties, recruitment and promotion in universities are basically handled at two levels. First, a national body, the *Conseil National des Universités* (CNU) subdivided into partly-elected, partly appointed disciplinary committees, is in charge of awarding “*qualifications*” to candidates on top of their doctorate (for associate professors) or habilitation (for full professors). Second, local disciplinary ad hoc committees are in charge of recruiting single faculty members (Paradeise ESF 2010). Most *schools* employ a small permanent faculty and a large number of high-level adjuncts who are hired on short-term contracts while being permanent members of universities, research centers, administration, business and industry. As in universities, the academic staff is usually not assessed after recruitment, except for promotion. Evaluation occurs at recruitment and promotion only. Should they not fit the needs of the school and students’ evaluations, their contracts would not be renewed. Research organizations are the only institutions, which carry out a periodical formal assessment since their foundation. They evaluate both their research centers and their full-time researchers, usually on a four-year basis, based on partly peer-elected, partly appointed committees in each large disciplinary field.

As elsewhere in Europe and in many countries worldwide, radical reforms followed, starting 2006, the incremental phase of the 1980-90s. The issue of evaluation had already been considered in the 1970s and more and more in the 1980s. Yet the need to assess all academics, universities and research centers only became consensual in the 1990s (Merindol 2008). Consensus developed as a counterpart of a rising awareness that more autonomy would benefit all stakeholders of universities, which were still highly dependent from state authorities. In 1983, four-year contracts on research and teaching were first introduced between the state and each university, bringing the latter to be identified as an assessable organizational body of its own, able to strategize and plan its future and to argue for its funding application. The *Comité National d’Évaluation* (CNE) was set up in 1985 – with little resources and major ambitions – to improve transparency on HE institutions performance. An *Observatory of Sciences and Techniques* (OST) was founded in 1990 to forge indicators of performance, with the purpose to progressively better support allocation decisions.

The 2006,³ 2007,⁴ and 2013⁵ legislative acts ruled on the autonomy and accountability of universities. Although it was at first perceived as a major break-through, it only granted a limited autonomy as compared to other European countries (EUA 2011), but it did impel the foundation of systematic tools and methods backing assessment on performance, and allocation on assessment.

² CNRS is the largest basic research organization. It is followed, in term of scientific staff numbers, by two targeted research centers, the *Institut national de la recherche agronomique* and the *Institut national de la recherche médicale* (INSERM).

³ *Loi de programme* No. 2006-450 du 18 avril 2006 pour la recherche.

⁴ *Loi* No. 2007-1199 du 10 août 2007 relative aux libertés et responsabilités des universités (so-called LRU).

⁵ *Loi* n° 2013-660 du 22 juillet 2013 relative à l’enseignement supérieur et à la recherche (so-called loi ESR).

On the one side, a national funding agency, *Agence nationale de la recherche* (ANR) was set up in 2006 with the purpose to increase the share of competitive public funding of research, partially based on research programs and partly on open programs. Moreover, the creation of a *Commissariat aux investissements d'avenir* (CIA, General Commission for future investments) in 2012, stressed the importance of large competitive funding of many institutional and operational levels of excellence consortia of universities (IdEx), laboratories of excellence (LabEx), excellence facilities (EquipEx), etc. This investment program, which starts its third round in 2017, has already dedicated 22 billion euros to “excellence initiatives” in higher education and research across the country.

On the other side, public authorities set up in 2007 an evaluation agency, *Agence d'évaluation de l'enseignement supérieur et de la recherche* (AERES) renamed in 2013 *Haut comité à l'enseignement supérieur et de la recherche* (HCERES). Its main duty is to assess all components of public higher education in all fields and at all levels outside individual scholars (HE institutions, research organizations, research centers and degrees). A (poorly) called “SYMPA formula” was elaborated at the same time by the ministry bureaus to translate performance in terms of funding allocation. It was intended to base a proportion of universities block grants (about 20%) on their outputs in teaching and research, in addition to the 80% based on their inputs (number of students in various fields, etc.). The internal allocation of university block grants was to be handled by the single universities themselves.

2. Reception and impact of assessment

The accountability turn accelerated the production of tools as a basis for resource allocation. In addition to the introduction of cost accounting, autonomous universities created their own management dashboards to back their strategic decision-making. The evaluation agency disseminated its own lists of detailed indicators on universities, research centers and curricula,⁶ in order to assess their organization, governance, funding and performance.⁷ Indeed such indicators inform the assessment, which is yet based on the evaluation drawn up every four years by ad hoc visiting committees using basically few metrics. As far as universities are concerned, such committees are required to synthesize their evaluation under a set of dimensions, covering governance, leadership and management, and strategy in research, knowledge transfer, teaching, student life, external and international relationships and communication. Curricula are assessed according to their goals, context, organization and results. Research centers are assessed on six dimensions: scientific production and quality, reach and attractiveness, interaction with the social, economic and cultural environments, internal organization and life, involvement in teaching research, and strategy and five-years programme.

⁶ More units are assessed, where for instance federal structures or consortia of universities (and possibly *schools*) have been set up.

⁷ Such as number and status of academic and management staff, organizational chart, decision-making procedures, etc.; attractiveness and placement of curricula; grants captured, patents and publications (based on a list of ranked refereed journals built up by disciplinary ad hoc committees), etc.

2.1. Reception of assessment and its consequences on assessment tools and processes

IS THE ASSESSMENT AGENCY LEGITIMATE?

A heated debate first developed about the issue of accountability. Each discipline used its own channels of influence to seek better arrangements for itself. Strong and politically threatening lobbies such as *Sauvons la recherche* brought together individual and collective discontent, based on shared protest against the so-called neo-managerial or neo-liberal turn in higher education. Such pressure groups resisted the use of assessment tools as a basis for funding, and they claimed the need to adjust evaluation criteria to the specificities of disciplines. The most radical individuals and groups disseminated the recommendation to boycott invitations to join ad hoc committees or to resist English speaking in such committees. At the end of the day, they organized political pressure to get rid of the agency and its tools.

Collectively, the attractiveness of protest varied depending on how much assessment was individually and collectively considered as a threat. Collectively, the less accessible the fields to international journal rankings – either because they belong to “non-nomothetic fields” such as the social sciences and humanities (Passeron 1991) or because they deal with professional knowledge as in law studies, accounting and engineering prone to pragmatic knowledge rather than academic publications – the more threatened they felt. Individually, the more scholars cumulated disadvantages such as being low-publishers and having lower academic statuses, the more they were likely to reject the assessment-driven model. The more disturbing reforms for the social exchange model they had built up with their own university, the more they were likely to raise their voice (table 1).

Table 1. Contributions of scholars, expected returns and reception of assessment.

Exchange model between scholar and...	Individual commitment	Expectations towards one's university	Individual behavior
...his/her university	Involvement in organizing and managing	Local promotion should reward organizational commitment	Withdrawal if HRM does not deliver promotion. Hostility towards the notion and measure of scientific excellence and voice against assessment and performance-based resource allocation
...his/her discipline	Personal conquest of excellence	The university should provide resources for personal accomplishment	Exit if the university does not comply with expectations, based on demonstration of an excellent performance. Positive behavior towards the notion and measure of scientific excellence and performance-based resource allocation
...both	Both are considered complementary	Local promotion should reward scholars committed as academics and as good citizens of the university	Local withdrawal or exit if the university does not meet the expectations. Voice in favor of pushing forward the reforms in such a way that universities really become complete organizations

The assessment frame evolved over time under such pressures, but the general rationale of the reform remained untouched. Indeed the assessment agency (AERES) was theoretically discontinued to comply with lobbies during the 2012 presidential campaign, but was practically immediately reopened under the 2013 ESR act with a new name (HCERES) and with a slightly renewed legal status, which kept though almost exactly the same jurisdiction as before.

WHAT SHOULD BE ASSESSED AT THE NATIONAL LEVEL?

One first struggle about assessment concerned the selections of entities to be assessed by AERES/HCERES. Each institutional actor – research organizations, the university system and *schools* – argued that it already implemented its own individual assessment. Their coalition, backed by individual scholars, led to finally leave them out of reach. The same fight occurred on research centres, but it was unsuccessful. It was indeed a major stake for the government to arrange uniform tools allowing for the development of research incentives, whether or not associated with any public research organizations. By linking part of block grants to the university outputs in research as measured by formal indicators such as number of patents and rates of publications in so-called best international journals, it was supposed to encourage individual scholars to improve their own performance and good research centers and universities to select best-performing scholars.

The development of research performance-based funding, however limited it actually remained, disrupted the traditional allocation scheme, by making it clear which centers and universities concentrated good publishers and which did not, thus reconsidering their bases of reputation (Paradeise and Thoenig 2015). Many scholars – individual academics or subgroups – felt they might be stigmatized as low research-performers. They also feared that such uniform indicators, which might not fit their ways of publicizing their research, would erode their disciplinary and cultural specificities and ostracize their field, by for instance favoring journals against books or memoirs, and English-written publications against their native language. Such reactions mostly took place among low- or non-publishers, and among humanities and social scientists.

SHOULD ASSESSMENT BY MADE PUBLIC AND HOW?

Making assessment reports and ratings public was a major change introduced with the creation of AERES. Used to shame or, rather, legitimize policies and funding, no one in the same field or in the same institution could ignore the comparative performance of research units or institutions. This would allegedly help the state decide upon resource allocation across universities, universities decide upon resource allocation between its sub-units, research groups and departments strategize⁸ in order to try get rid of non-publishers or improve their scores through better recruitments, etc. Academics feared a mechanical implementation of scoring on decision-making, while management dashboards could be used in many other ways, for instance to reinforce a poorly performing discipline which, for some reason, was considered important by a given university. Protest denounced the illegitimacy of such publicity. For these reasons, they first opposed synthetic scores on a scale of five, (from A to E) which the visiting committees were required to deliver. Soon after the first round of assessment, the agency decided to buckle under this pressure and reconsidered ratings as a list of itemized non-additive scores on each of the dimensions under evaluation. Finally, it was invited to totally renounce and frankly discontinue any form of scoring.

⁸ We have shown elsewhere how much the strategizing capacity varies from one place to the other within a single country (Thoenig and Paradeise 2016).

HOW SHOULD JOURNALS BE ASSESSED?

One important issue emerged that was about the ranking of journals in certain fields, which feared inadequacy of the uniform criteria applied in assessing publication performance (table 2). “There is a classification of journals in natural and life sciences... (which is) mostly based on English-written journals. Such a classification does not exist in the social sciences and humanities. And it seems to poorly fit academic outputs in these fields, largely French-written” (Glaudes 2014). As a result, the behavior of disciplinary communities varied widely about the injunction to list reference journals in their field and even rate them on a three steps scale.

Table 2. Disciplinary publication cultures.

	Math	History	Political Science, Sociology	Chemistry, Biology
Privileged medium	Journals, Annals	Books, Conferences	Journals, Books	Journals
Agreement on media hierarchy in the community	Strong	Average +	Average -	Strong
Production time	Slow	Slow	Medium	Rather fast
Citation accumulation	Very slow	Medium	Medium	Fast
Citations obsolescence	Slow	Slow	Slow	Rather fast

Certain disciplines, such as economy or management, simply replicated lists built elsewhere, for instance by public research organizations or international newspapers such as the *Financial Times*. Others, such as law, built up their own list, based on the empirical signals of reputation established by their representative authorities. Others used their own indicators without any effort to link them with those of others. Philosophers defined for instance their own four specific criteria (requirement to principally publish articles in philosophy, existence of a scientific committee and an editorial board, including non-French members, double-blind evaluation, selectiveness). On top of similar criteria, communication scientists added up a list of other items such as regularity of publication and size of articles, restriction of auto-publication, institutional links with the discipline and indexation in international databases.⁹

Methodological diversity added up to differences in the established lists. Several disciplinary committees in the field of arts, social sciences and humanities (for instance in sociology, political science, theology, philosophy, anthropology, geography and urban planning, history, arts and law studies) simply refused to set up journal rankings. Some contributed by listing journals that belonged to their scientific perimeter. They promised to, and did progressively develop their own rankings (communication studies, psychology), each with its own scale. Finally, a series of disciplines (concentrating in languages, literature and civilizations) totally rejected the very notion of a list, arguing that they were irrelevant in their field.

The AERES finally took notice of this resistance and, since 2010, started rebuilding the lists, including other items based on a more cautious typology of media, such as scien-

⁹ See online <http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-de-l-AERES>.

tific books (based on publishers, signatures, purposes and editorial work). It made it clear that it did not favor quantitative evaluation but followed the moderate recommendations of the French academy of science¹⁰ by more generally restricting the use of bibliometrics to the assessment of entities which size exceeds 30 scholars, and systematically referring bibliometric results to their average values and to the 10% top values in a given field. In addition, experts – who are always peers in the fields under assessment – were invited to be cautious about possible biases of such results. Thus AERES fostered a non-mechanic use of bibliometric indicators and insisted that they should be contextualized and interpreted, and should not replace the reading of papers in order to assess their actual scientific interest.

WHAT SHOULD BE THE PERFORMANCE OF A “PUBLISHER”?

Lobbies paid attention to the norm set up to define what to be a “publisher” means. They also insisted that this norm should vary according to the publication tradition of each field. They worked at lowering the threshold and finally ended up in the social sciences and humanities accepting a (very) light norm of 4 articles for a full-time researcher or 2 articles for a professor in a four-year period, with indeed very little variation from one field to the other.

OVERALL STRENGTH OF ASSESSMENT IN FRANCE

To tell the truth, protesters over-emphasized the threats of assessment, at least during the current stage of reforms. On the one hand, a recent EUA survey shows that the impact of evaluation of teaching is comparatively very low in France (table 3). On the other hand, research outputs are regularly assessed, have gained influence on recruitment and promotion, but have no impact whatsoever on tenure and salaries of academics who remain civil servants paid according to a fixed national grid of statuses.

Table 3. How much is assessment actually developed in France? (source: EAU)

	Europe	France
Students satisfaction survey are used in the assessment of teachers	93%	39%
Courses are regularly assessed	89%	33%
Research outputs is regularly assessed	84%	100%
There are ways to face situations where teaching performance of someone remains weak	72%	18%
Heads of departments/deans of faculties discuss regularly the quality of teaching with all teachers	67%	24%

In addition, the impact of evaluation by the assessment agency on resource allocation remains limited. Yet, other forms of evaluation play a major role in differentiating individuals, research centers and universities in the competition for grants.

¹⁰ See online <http://www.academiesciences.fr/activite/rapport/avis170111.pdf>.

2.2. The impact of assessment on funding

THE IMPACT OF AERES/HCERES

Performance evaluation by assessment agencies was developed as a tool of accountability for a better governance of the new autonomous universities. It justified building all sorts of indicators intended to inform the SYMPA formulae mentioned above. This top-down tool however remained ineffective until 2017. First, it took time to set up the required databases. Second, many universities considered that a top-down approach to the building of indicators would not be able to capture their actual performance. Third, in a context of stability and even reduction of higher education public budgets, the government did not dare bypassing the established distribution of block grants. In other words, the top-down approach, which prevailed in the SYMPA formulae as a link between performance and allocation, was discarded almost as soon as it was created. Starting in 2016, a new bottom-up approach by sector, developed with the collaboration of HER institutions, is supposed to design a new formula that should better fit the specificities of each field.

To put it in a nutshell, the output-based assessment by AERES/HCERES has since its foundation proved rather ineffective operationally. Nevertheless, its symbolic impact has been enormous by making publicly known the strengths and weaknesses of units and sub-units in the system of HE and research, by fostering strategic moves at each level, by setting up the issue of publication and by insisting on its contribution to the missions of academics, etc.

THE IMPACT OF ANR AND CIA

On the contrary, the development of project-based grants over the last ten years has had major operational impacts on the dynamics of universities, research centers and to a certain extent individual careers, first with the funding of research projects by ANR programs and increasingly with the substantial sums supplied by the CIA institutional excellence initiatives. Three waves of funding have been set up since 2012, covering a variety of large projects involving not only research but also the founding of new institutional bodies – laboratories of excellence (LabEx), excellence facilities (EquipEx), excellence institutions (IdEx). The CIA program progressively diversified its funding streams, which now include innovation in teaching as well as incentives targeting the development of specific niches of excellence within universities with the *Initiatives Science-Innovation-Territoires-Economie* (I-SITE). The international high-level evaluation committee pays much attention to the relevance of projects in scientific and operational terms but also to their feasibility in terms of governance. The important resources procured by such programs operate as very strong incentives that also encourage the development of project-based consortia and even mergers between research centers, departments, universities and schools. The CIA programs have thus come to play a key role in the current on-going stratification of French higher education and the restructuring of the national landscape, both at the institutional level and between and within disciplines. By concentrating important resources on specific territories, these programs favor the visibility and attractiveness of certain universities or certain niches within universities.

2.3. The impact of assessment on the profession

The development of assessment has provided a rationale for the redistribution of resources between universities and between disciplines. It may not have had much impact upon individual salaries and national careers of academics, who remain civil servants, but it has positively changed working conditions of the best-performing units, whatever the discipline. Altogether, humanities and social sciences have received less budgetary resources than hard sciences, partly because they display lower needs than experimental sciences, partly because assessment tools too often have difficulties grasping their specificities. But, as shown by the relative growth of their memberships, being in line with the massification of higher education in a non-selective system, they have not been ostracized as such (table 4).

Table 4. Size of permanent academic staff by discipline. 1992-2013 (source: DGRH, French Ministry of Higher Education and Research), mentioning the position of political science and the highest and smallest growth in each large field.

Field	2013	% Growth 1992-2013
Social sciences and humanities	14,999	+57.2%
<i>Sociology</i>	877	+98.9%
<i>Literature and languages</i>	6,132	+24.0%
<i>Political science</i>	360	+59.3%
Hard sciences	24,907	+40.7%
<i>Engineering</i>	6,845	+71.4%
<i>Math and informatics</i>	6,601	+63.9%
<i>Physics</i>	2,429	+2.0%
Medicine and pharmacy	8,050	+2.4%
Grand Total	55,515	+40.4%
Students population	2,471,000	+43.9%

Yet, political scientists have been impacted, as have other scholars especially in the social sciences and humanities. The development of performance-based assessment, however limited if compared to some other European countries, has revealed a more visible hierarchy between scholars. Reputations and statuses have been tested by performance as measured by “excellence” metrics (Paradeise and Thoenig 2015). The worldwide generalization of accountability is segmenting the academic population, building up a pecking order between first-class and second-class scholars, publishers and non-publishers, members of top, second- and third-tier institutions. As stratification between universities increases, one may expect that the best-rated departments and/or universities will increasingly attract first-class scholars, who are also chased on the international market and whose salaries may become much more flexible and substantial. Two labor markets are thus being created. Roughly speaking and with several exceptions, the international one increasingly takes care of the “stars” while the national one takes care of the others. Since French civil servants’ salaries are all but competitive, institutional reputation will not be enough in the future to prevent more academics to leave the country, a tendency already confirmed by a still limited but increasing trend among younger scholars.

3. Conclusion: pros and cons of research assessment

When considered at the systemic level, the obsession of HE policies to make French universities “visible from Shanghai” could endanger universities and departments which have no hope of accessing the Walhalla of world excellence, but place a major emphasis on the higher education of large segments of the young population. Thus, France should remain cautious not to concentrate evaluation solely on cutting-edge research. As other European countries, it should take care to preserve and encourage the many and varied “excellences” that are needed to face the various missions of universities.

For all these reasons, it is difficult to sum up and provide a uniform overall assessment of research assessment in France. The reception of assessment mostly co-varies with the opportunities it provides and the threats it involves for universities, faculties, research centers and individuals. The analysis of such opportunities and threats does not identify disciplines and scholars that would uniformly be the losers or the winners of the new rules of the academic game but rather cuts across all of them (table 5).

Table 5. Reception and impact of assessment in France. A synthesis.

	University level		Professional level		Individual level	
	Pro	Con	Pro	Con	Pro	Con
Reception	Backs strategic perspectives	Fear that strategic decisions be taken in a neo-managerial perspective and hurt collegial regulations and the very nature of universities	Resource for challengers as opposed to oligarchic cliques within the profession	Appropriateness of metrics varies across disciplines Finalized competitive grants may negatively impact blue sky research	Fragmentation between publishers and non-publishers	Fragmentation between publishers and non-publishers Publications and patents do not account for all types of contribution as a scholar
Impact	Contributes the rising stratification of universities by differentiating resources: recruitment, budgets and facilities. Strategizing involves more conversation between departments to decide on and legitimize the allocation of resources.	Contributes the rising stratification of universities. Risk to try get rid of disciplines that do not score in rankings and bring costs rather than budgetary resources. Possible emergence of internal political fights based on divergences between departments on what must be valued in the academic work	Criteria of recruitment and promotion necessarily involve research activities Contributes the rising stratification between departments in a given field by differentiating resources: recruitment, budgets and facilities	Contributes the rising stratification between departments Risk of constraining recruitment and promotion choices based on performance as defined by metrics	Rewards who does the academic job (teaching, research, administration). Encouragement to target better vectors and upgrade publications	Disqualification of non-publishers, even when they do a good job in teaching

References

- Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs, Académie des sciences, janvier 2011.
- Glaudes, P. 2014. L'évaluation de la production scientifique en France par l'Agence d'évaluation de la recherche et de l'enseignement supérieur?, *Mélanges de la Casa de Velázquez*, 2, Tome 44, 293-300.
- Merindol 2008 Mérimond J.-Y. 2008. Comment l'évaluation est arrivée dans les universités françaises, *Revue d'histoire moderne et contemporaine*, 55-4bis, 7-27.
- Passeron J.-Cl. 1991. *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*, Paris, Nathan Essais et Recherches.
- Paradeise C. 2011. 2011. Higher education careers in the French public sector, Permanence and change, in Avedduto S. (ed.), *Convergence or differentiation. Human resources for research in a changing European scenario*, ScriptWeb: Napoli, 159-184.
- Paradeise and Thoenig 2015. *In search of academic quality*, London: Palgrave-Macmillan.
- Thoenig J.-C. and Paradeise C. 2016. The strategic capacity of academic institutions, *Minerva*, 54(3), 293-324.

The Italian Research Assessment Exercises

Daniele Checchi
UNIVERSITY OF MILAN

The Italian experience

Italian universities have so far experienced three assessment exercises (2001-3, 2004-10 and 2011-14), which are described in details in Table 1. The fiscal law approved in December 2016 dictates that from now onwards the reference periods will be quinquennial, reducing the discretionary power so far exercised by the Ministry of Education in designing the exercise.

Table 1. The three research assessment exercises

Acronym	VTR 2001-2003	VQR 2004-2010	VQR 2011-2014
Evaluator Agency	CIVR	ANVUR	ANVUR
Output to be evaluated	Research outputs, proportional to the number of researchers – no constraints on the research areas to be covered	A fixed number of research outputs for each researcher (3 for university professors, 5 for researchers in research agencies)	A fixed number of research outputs for each researcher (2 for university professors, 3 for researchers in research agencies)
Period of reference	2001-2003	2004-2010	2011-2014
Subjects to be assessed	Compulsory for universities, voluntary for research agencies – finer assessment down to departments	Compulsory for universities and research agencies under the monitoring of Ministry of Education – voluntary for any other research agency – finer assessment down to departments	Compulsory for universities and research agencies under the monitoring of Ministry of Education – voluntary for any other research agency – finer assessment down to departments
No. of research outputs	18.000	185.000	105.000
Assessment scale	4 merit classes	4 merit classes	5 merit classes
Assessment method	peer review	bibliometric methods for hard sciences - peer review for soft sciences	bibliometric methods for hard sciences - peer review for soft sciences
Other indicators for assessment	<ul style="list-style-type: none"> ownership of copyright and royalties number of PhDs and Postdocs degree of internationalisation (researcher mobility) attraction of funding ability to self-financing research projects 	<ul style="list-style-type: none"> attraction of funding ability to self-financing research projects scientific merit of newly hired/promoted personnel degree of internationalisation (researcher mobility, co-authorship with foreign-based researchers) improvement over the positioning during previous VTR 2001-2003 	<ul style="list-style-type: none"> attraction of funding scientific merit of newly hired/promoted personnel improvement over the positioning during previous VQR 2004-2010
External reviewers	yes	yes	yes
Panels	20 groups of evaluators	14 groups of evaluators	20 groups of evaluators
Impacts on public funding	no	yes	Yes

After an initial trial-and-error approach, the second and third exercises have been rather similar, thus consolidating a standard of evaluation, whose principles are the following:

- each assessment is intended to evaluate groups (universities, research agencies, down to departments and institutes) and not individuals (individual assessments are revealed to each researcher, but not to heads of departments, deans or chancellors);

- the assessment considers a fixed number of products per capita/year, which should capture the best production: as such, it is closer to a monitoring exercise than to a quality assessment, revealing the excellences in a given research field;
- using current standards (1/2 product per year per university professor – currently around 52,000 – and 1 product per researcher working in a research agency – currently around 10,000) implies approximately 35,000 products per year; over a 5-year interval it sums up to 175,000 products, making some sort of automatic (bibliometric) assessment unavoidable;
- the process has been managed by groups of experts, defined according to predefined research areas (since Italian professors are pigeon-holed into 371 research fields, then grouped into 14 research areas, known as Aree CUN). Each group was composed by a variable number of experts (from 20 to 60, depending on the expected number of products – the experts were selected by ANVUR from list of applicants according to their publication records and their area of expertise). In turn, these experts relied onto 14,500 external peer reviewers, working in domestic and foreign institutions;
- in the last two exercises, the evaluating agency (ANVUR) requested to the experts a preassigned distribution of journals, according to the world distribution of impact. As a consequence, the top list of journals should correspond to the best 10% of the world production; nevertheless, more than 30% of the submitted products to the last exercise ended up in this category (because the exercise considers only the best products);
- depending on the research area, two assessment procedures have been followed:
 - *bibliometric assessment* consisted of combining the ranking of the journal according to the Impact Factor and the citations obtained by a specific article – articles in highly ranked journal with limited citations and/or highly quoted articles published in low ranked journals were peer reviewed;
 - *peer review assessment* consisted of a product being separately assigned to two experts, who independently selected an external peer reviewer; once the reviews were returned, a consensus report was drafted by the experts. In case of significant disagreement, a third reviewer was introduced, and the final assessment has to be approved by coordinator of the group of experts.

In both cases the submission to experts were non-blind, and the evaluators may have formed their opinion looking at the place of publication, in what has been called as “informed peer review”.

2. The impact of the research assessment

The evaluation of the product is normalised according to the means in each research area, leading to an indicator which combines quality and quantity assessment of a research field in a university.¹ This indicator counts for three-fourths of the funds allocation, and is then

¹ From a technical point of view, the indicator consists of the share of scores attained by a single university/department over the total scores achieved at the national level by all institutions. That share is then applied to the distribution of funds. If a university/department performs above the average, it will obtain a funding share which exceeds the corresponding share computed on the personnel heads.

complemented with other indicators (PhDs, foreign students, external funding) in order to achieve the summary indicator to be applied to a funding scheme for universities. The most recent exercise led to the distribution of ½ of total funding to public universities in Italy (1.4 billions of euro for 2016). Approximately 15% of total funding relies on the proper evaluation of research products.²

As such Italy belongs to *evaluation-based* systems (with the UK, Australia, New Zealand), to be contrasted with *indicator-based* systems (Norway, Denmark, Czech Republic). However, the 5-year interval is long enough to call for alternative methods of evaluation in the intermediate years. In addition, the results of the evaluation have trickled-down, directly or indirectly, to many other dimensions of the life of university departments. Many universities have used the scores obtained by their departments in the internal allocation of funds and promotions; the current accreditation of PhD programs is based on the research assessment of the teaching staff; newspapers articles have widely disseminated the results of the research assessment with reference to local universities, in order to drive the choices of students and their families.

Even if they are formally independent, the process of selecting new academics has been significantly influenced by the research assessment exercises. Selection in hard science research fields makes large use of bibliometric methods, while in soft science journal rankings have been adopted. Though I would not dare claiming that the introduction of assessment exercises has raised the standards of hiring in most disciplines, as a matter of fact in the most recent VQR the average score of newly hired/promoted researchers is higher than the average of permanent members (the indicator called IRAS2). This implies that new entrants in the academia have introjected the assessment approach in shaping the way in which they publish their research outputs.

While the VQR asks for the assessment of “originality, relevance, exposure to international debate”, what is more perceivable (and perceived) is the internationalisation of the domestic production. Publishing in a foreign language (notably in English) has become the dominant strategy in several fields. As a consequence, many Italian journals which used to publish in Italian opted for the English language. A related issue is the multiplication of the number of papers via the diffusion of co-authorship. Since the VQR rules allow for the same product being submitted by more than one author (as long as they belong to different research entities), many authors have followed a strategy of risk diversification, by developing joint research projects in the expectation that at least one of them would obtain publication in a highly ranked journal.

3. The recent VQR (2011-14)

The most recent research assessment exercise ended in February 2017, with the official presentation of global report on the Italian research activity accompanied by specific reports for each research areas and for the social impact activity. 96 Universities participated to the exercise, together with 12 PRO's (Public Research Organisations) and 26 other institutions on a voluntary basis. The distribution of 118,036 products received for evalua-

² To be honest, the impact on funding is less dramatic in the short run, because of high persistence on historical values: each university cannot receive $\pm 2\%$ of what it has received the previous year, thus strongly attenuating whatever result could obtain from the research assessment.

tion is reported in Table 2, where one can easily detect few regularities. Compliance rates vary across research areas, oscillating between 90% and 97%.³ Journal articles represent the dominant submission for hard sciences (reaching 98% in Biology and Medicine), while collected papers (edited volumes) prevail in the social sciences and humanities. Books have almost not been submitted in bibliometric areas, while they represent one fourth of all submissions in some non-bibliometric areas. The residual category [including musical compositions, designs, projects (architecture), performances, exhibitions, arts objects, databases and software] account for a small fraction of the total output submitted to the assessment. This does not produce a representative snapshot of the research activity of universities (PROs have similar composition), because the limit imposed to two products per researcher. Rather it allows monitoring of what can be considered as relevant scientific productivity of the entire research community.⁴

Table 2. Distribution of products by research area and type of output – Italy VQR 2011-14

Research area	Number of products expected	Number of products submitted	% missing	Journal articles	% on total by area	Books	% on total by area	Articles in collected paper volumes	% on total by area	Patents	% on total by area	Other products	% on total by area
1	6,680	6,007	10.1	5,278	87.9	66	1.3	632	10.5	1	0.0	17	0.3
2	10,923	10,562	3.3	10,244	97.0	22	0.2	154	1.5	18	0.2	121	1.1
3	7,232	7,033	2.8	6,916	98.3	18	0.3	75	1.1	23	0.3	1	0.0
4	4,638	4,398	5.2	4,052	92.1	38	0.9	276	6.3	2	0.0	28	0.6
5	11,706	11,033	5.7	10,831	98.2	18	0.2	154	1.4	13	0.1	15	0.1
6	18,148	16,936	6.7	16,612	98.1	54	0.4	256	1.5	7	0.0	1	0.0
7	7,849	7,470	4.8	6,848	91.7	59	0.9	535	7.2	14	0.2	9	0.1
8a	3,659	3,468	5.2	924	26.6	758	23.4	1,633	47.1	11	0.3	87	2.5
8b	3,010	2,818	6.4	2,497	88.6	12	0.6	299	10.6	5	0.2	1	0.0
9	12,074	11,339	6.1	10,055	88.7	67	0.6	1,146	10.1	45	0.4	21	0.2
10	9,363	8,761	6.4	2,822	32.2	1,944	23.9	3,819	43.6	0	0.0	30	0.3
11a	6,476	6,122	5.5	2,098	34.3	1,661	28.6	2,265	37.0	0	0.0	6	0.1
11b	2,385	2,292	3.9	2,026	88.4	80	3.8	175	7.6	0	0.0	4	0.2
12	8,973	8,502	5.2	3,373	39.7	1,977	26.3	2,893	34.0	0	0.0	2	0.0
13	9,039	8,300	8.2	6,041	72.8	638	8.7	1,490	18.0	2	0.0	47	0.6
14	3,242	2,995	7.6	1,232	41.1	699	24.6	1,024	34.2	0	0.0	2	0.1
Total	125,397	118,036	5.9	91,849	77.8	8,111	7.5	16,826	14.3	141	0.1	392	0.3

Note: The evaluation is organised in the following research areas: Mathematics (Area 1) - Physics (Area 2) - Chemistry (Area 3) - Earth Sciences (Area 4) - Biology (Area 5) - Medicine (Area 6) - Agricultural and Veterinarian Sciences (Area 7) - Civic Engineering (Area 8a) - Architecture (Area 8b) - Industrial and communication Engineering (Area 9) - Humanities (Area 10) - History and Philosophy (Area 11a) - Psychology (Area 11b) - Law (Area 12) - Economics and Statistics (Area 13) - Social Sciences (Area 14).

The assessment of each product was conducted according to three criteria:

1. *Originality*, to be intended as the degree according to which the publication is able to introduce a new way of thinking about the object of the research;
2. *Methodological accuracy*, to be intended as the degree according to which the publication adopts an appropriate methodology and is able to present its results to peers;
3. *Actual or potential impact*, to be intended as the level of influence – current or potential – that the research exerts on the relevant scientific community.

³ It is important to recall that a protest organised in some universities led a fraction of university professors to refuse to submit their required output. However, in the first VQR, the submission rate for universities was 95.09% of the expected output, while it went down to 933.82 during the second one.

⁴ The rules prevented the submission of textbooks, working papers and self-publications.

Each publication was attributed a quality profile:

- **Excellent** (weight 1) if it falls in the top decile of the world distribution of publications in the research area;
- **Good** (weight 0.7) if it falls in the 70-90% segment of the distribution;
- **Fair** (weight 0.4) if it falls in the 50-70% segment of the international distribution;
- **Acceptable** (weight 0.1) if it falls in the 20-50% segment of the distribution;
- **Limited** (weight 0) if it belongs to the 0-20% lowest segment of the distribution;
- **Impossible to evaluate** (weight 0) was assigned to missing publications or publications that were impossible to evaluate.

As one can easily expect, any evaluation of a product following the above-mentioned criteria contains some degree of arbitrariness. One can initially consider the language of publication as a proxy for the exposure to the international debate. An inspection to Table 3 seems to suggest that what are considered as bibliometric sectors (in light grey) are largely open to the international debate. From this perspective, the research area 13 (Economics and statistics) could be considered equally open to internationalisation. These areas have mostly relied on automatic assignment of products to the evaluation categories, using the principle that journal with high impact factors are generally speaking more selective in acceptance, and therefore impose higher standards of quality. This principle is complemented with the use of papers' citations, which should capture the relevance of the contents for the scientific debate.

The evaluation in non-bibliometric areas relied on peer review (with the exception of the research area 13, which adopted a ranking of the journals based on the impact factors). If the replacement of an algorithm with human reviewers may be welcome in terms of adherence to the suggested evaluation principles, it introduces the problem of potential disagreement among the reviewers, which is likely to motivate the lower fraction of “excellent” and “good” evaluation recorded in the non-bibliometric areas (see Table 4).

Table 3. Language of the products submitted to VQR 2011-14

Research area	Italian	English	Any other foreign language	Information n/a	Total
1	142	5,907	12	1	6,062
2	69	10,514	3	2	10,588
3	41	6,850	6	0	6,897
4	157	4,265	8	0	4,430
5	107	10,858	21	0	10,986
6	506	16,145	42	0	16,693
7	534	6,998	9	0	7,541
8a	2,239	1,171	46	0	3,456
8b	137	2,692	3	0	2,832
9	154	11,401	9	0	11,564
10	5,574	2,051	1,119	0	8,744
11a	4,295	1,454	374	0	6,123
11b	301	1,962	13	0	2,276
12	7,671	678	139	0	8,488
13	1,897	6,451	35	2	8,385
14	1,866	985	120	0	2,971
Total	25,690	90,382	1,959	5	118,036

Table 4. Distribution of products by research area and received evaluation – VQR 2011-14

Research Area	A (excellent)	B (good)	A+B	C (fair)	D (acceptable)	E (limited)	F (non classified)	Total
1	38.4	28.0	66.4	18.2	10.8	4.2	0.4	100
2	62.2	21.6	83.8	10.4	4.7	0.9	0.1	100
3	49.2	32.0	81.1	12.9	4.6	0.8	0.6	100
4	27.9	29.7	57.5	21.6	14.2	5.5	1.2	100
5	37.3	31.5	68.7	19.0	9.3	1.8	1.2	100
6	39.5	25.8	65.3	17.8	11.9	3.6	1.4	100
7	28.4	31.5	59.8	19.4	14.9	5.4	0.5	100
8a	8.6	34.2	42.8	35.9	16.0	4.8	0.5	100
8b	37.6	29.3	66.9	17.7	12.6	2.5	0.2	100
9	38.6	27.6	66.2	18.2	12.3	2.7	0.7	100
10	18.1	46.2	64.3	25.4	8.7	1.4	0.2	100
11a	16.1	42.4	58.5	29.2	10.2	1.8	0.3	100
11b	30.8	23.4	54.2	19.1	18.7	6.8	1.2	100
12	7.8	41.2	49.0	35.9	12.2	2.2	0.7	100
13	24.6	22.9	47.5	17.9	19.5	12.7	2.3	100
14	8.3	32.5	40.9	34.5	20.0	4.4	0.3	100
Total	32.6	30.8	63.4	20.7	11.6	3.5	0.8	100

4. The receipt of the research assessment exercises in the academic community

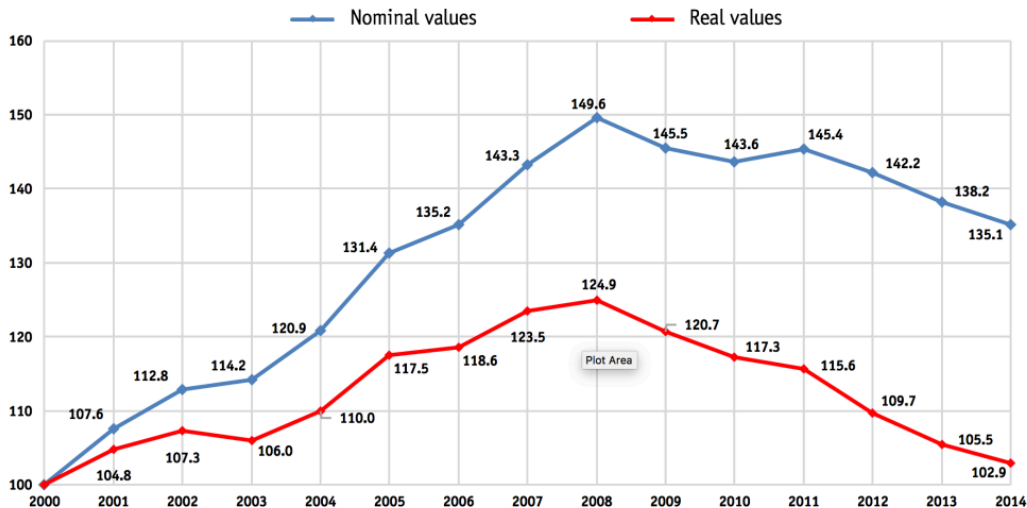
These exercises have generated enthusiasm and collaboration as well as suspicion and resistance. A large fraction of academics definitively cooperated with the exercise, organising the submission within each department and accepting to review the product. A smaller fraction opposed it, on the arguments that these exercises were misleading the Italian research towards irrelevant topics, were promoting harmful competition among research agencies and were destroying the weakest segment of the academia (very often located in Southern universities).⁵

My impression is that the main argument against the research assessment exercise runs as follows: “the assessment legitimizes budget cuts, especially against southern universities. If we want to save the equal opportunity in accessing universities, we should oppose any assessment which associate funding and results”. This argument has some plausibility, especially when looking at Figure 1, which shows the trends in state funding to Italian universities in nominal and real (i.e. deflated by the price variation) terms. Remember that the first exercise with impact on funding was launched in 2011, when the decline in resources became more pronounced. Although the actual impact was not disruptive (due to safeguard clauses – see above), the linkage of resources to assessment opened the risk of “poverty traps”: a poorly performing university received fewer resources and was therefore less likely to improve its performance in the next round of assessment. Budget cuts curtailed hiring possibilities, which were only later released in

⁵ Perhaps the most representative instances of this aversion towards evaluation performed by ANVUR can be traced in the following websites (unfortunately, all in Italian): www.roars.it (usually covering topics related to assessment methods); <http://www.flcgil.it/universita/> (the website of the main union of university workers); <http://firmiamodimissionianvur.org/> (more than 2,000 researchers signed a petition asking for the dismissal of the board of ANVUR, the evaluation agency).

correlation with performance. Thus poorly performing universities were supposedly prevented from hiring better researchers in order to revert their rank position.

Figure 1. Total public revenues accrued to Italian public universities (2000=100)



Despite its simplicity, this line of argument is substantially flawed. During the first decade of the present century, the hiring procedures of Italian universities were reformed, moving from a format of centralised competition to one of local competitions. Each department was left almost free to hire or promote whoever they deemed worthy to be hired. The first exercise (VTR) did not provide a clear picture of the average performance, because it was designed to assess excellence within each university, without considering who wrote what. The second exercise (VQR 2004-10) for the first time revealed that a non-negligible fraction of researchers was unable to submit any research product at all. The third exercise (VQR 2011-14) provided evidence of some convergence of universities towards the mean, thanks to the change in the grading procedure (missing submissions were no longer penalised with a negative grade) but also to the injection of new resources that made possible to all universities the hiring of new scholars.

5. Open issues for future assessment exercises

In the immediate aftermath of the publication of the results of the third exercise, several suggestions have emerged in the press as well as in official forums. Some of them were mainly technical, some other more philosophical. In the following I will review them in brief.

The first concerns the potential bias contained in the evaluation. Given existing rules, co-authored papers to be submitted to foreign journals have the highest probability to receive a high grade. This implicitly “delegates” to foreign editors (and publishers) the choice of what is to be considered relevant for the international debates. Topics that are outside the mainstream, or that are simply concerned with national debates, are likely to appear at best in local journals, which then receive lower evaluation even by referees. Still, most of Italian journals do not yet have standard double blind reviewing procedures, inducing the suspicion that the quality of their articles may be lower.

The absence of domestic databases on publications and citations makes it impossible to introduce a dual layer system, where articles and books in Italian could gain more visibility. The use of peer reviewers is not a panacea, for various reasons. Especially in the social sciences, where the ideological content of the arguments is important, the judgment of the reviewer may be biased by strategical concerns (by attributing a lower score to an author, one may be tempted to alter the competition among different schools of thought). In addition, peer review of papers that have already undergone a real blind review process represent in inherent contradiction: suppose that the final reviewer spots an evident error; who has to be blamed, the author, the journal referees, or the editor of the journal? Finally, the peer review is expensive. Consider the following back of the envelope calculation: in the most recent exercise 52,060 products (corresponding to 44.1% of total production) underwent a double review; each reviewer received 30 euro per review, leading to a total cost above 3 million euro, which is a cost that cannot be frequently afforded.

The second aspect concerns the different publication strategies of different research communities. On average applied physics scholars publish more than 30 papers per year, because the number of co-authors can easily exceed one hundred. The corresponding figure for a theorist in mathematics may not reach one paper per year. To partially account for these differences the scores are normalised by research area, but this does not reduce the evident advantage of sectors where the scholar may select their best production from a larger set of papers.

A related issue deals with the weighing of different products. The most recent exercise introduced for the first time a different weighing for books vis a vis journal articles: under specific request of the author, a book could have been considered as equivalent to two articles, thus satisfying the requirement of submission. But the principle could be extended to other categories of products, because an article collected in a book is probably subject to less scrutiny than an article in a journal. Articles and/or books could be weighed by the number of co-authors. And so on.

A further issue that has been raised deals with the boundaries of research areas. So far the assessment exercises have considered aggregation of research fields (*settori scientifico-disciplinari*) under which academics have been hired to teach. This does not have any correspondence to other classification criteria (like ERC) and tend to penalise cross-disciplinary research. In principle, nothing prevents redesigning of the evaluation areas, but this interferes with the academic careers, which represents the strongest incentive to publish (at least for academics). Thus, a net separation between research assessment and promotion criteria would be required before addressing this problem.

A final point deals with the potential trade-off between teaching and research. The assessment is conducted without any reference to the resources available/invested in research, including the time absorbed by teaching. Most universities in peripheral areas lament the excess burden of teaching created by the chronic lack of staff. From an intuitive point of view, a proper assessment should correct for differences in the starting conditions. Otherwise stressing research results as unique measure for scholars' quality is detrimental to the effectiveness of teaching, because scholars will devote their best energies to article writing. There are possible solutions to avoid this trade-off: if each academic could choose over a menu of different combinations of teaching loads and commitment to publications, we could observe a possible sorting of scholars according to their preferences

and abilities. This would require a revision of the procedure of assessment, because scholars should then be weighed or converted into full-time equivalents.

Overall, the unsolved issue for the Italian research assessment exercises seems to be whether the results should be interpreted as *monitoring the system* (in order to ensure accountability vis-à-vis the tax-payers) or rather a *research quality assessment* (intended to promote excellence). The Ministry of Education oscillates among these two interpretations, which however lead to alternative policy suggestions. According to the former perspective, uniformity of performance is a goal, and the weakest universities should be sustained in order to grant a common standard of tertiary education across the country. According to the latter, the best universities/departments should obtain even greater resources, given their good evaluations obtained in the assessment.

Book Reviews

Section edited by Carla Monteleone and Stefania Panebianco

MANUELA CAIANI AND SIMONA GUERRA (EDS.), *Euroscepticism, Democracy and the Media. Communicating Europe, Contesting Europe* (Palgrave Macmillan, 2017). 284 pp., £52.99 (e-book), ISBN: 9781137596437

Weakening financial and economic equilibrium, rising political relevance of the immigration issue, the resonating anti-EU rhetoric of populist parties, and more recently the outcome of the UK referendum shed light on the ‘crises’ that hit European democracies and, specifically, European Union institutions from different angles. As a result, at the national level discontent towards European integration foments political conflict while gaining more and more salience. Within this context, old (i.e. newspapers) and new tools of communication (i.e. websites and social networks) bridge political actors’ positions towards the EU and shape the public debate over the EU legitimacy.

This book, edited by Manuela Caiani and Simona Guerra, offers an in-depth evaluation of the multi-layered concept of Euroscepticism considering citizens’ changing attitudes (both pro and con) towards EU politics and the role played by traditional and digital media in framing EU polity, politics, and policies. The all-encompassing approach adopted in this book, as stated in the introduction, aims to investigate political parties’ and civil society’s *contingent and qualified* or *outright and unqualified* contestation of the European democracy.

The volume consists of 12 original contributions covering the existing literature on Euroscepticism, democracy, and the media. These contributions analyse the extent to which mass media portray the EU in the political and public debate of different member states, such as the UK, France, Italy, Denmark, Greece, Poland, Germany, Spain, and the Netherlands. The book is organised in three parts. In the first section, the authors review the current academic debate on the themes and offer new theoretical suggestions. The second part focuses on the role played by traditional media in shaping and fuelling the so-called ‘spiral of Euroscepticism’ in comparison to online platforms, such as the internet and social networks. The third part analyses the extent to which new media channel Eurosceptic political conflict to civil society. In the introduction, Manuela Caiani and Simona Guerra provide a well-articulated review of studies on Euroscepticism and the impact that media can

have in influencing public opinions and attitudes. Contestation towards the EU may occur differently across ‘actors, tactics, and forms’.

There is no unique definition of Euroscepticism, rather, it is a nuanced phenomenon that originates from domestic political conflict. Although scholars tend to focus mainly on party-based Euroscepticism, citizens’ emotions play a crucial role in affecting the process of EU integration, as the result of the Brexit referendum demonstrated (Simona Guerra, chapter two). Under these theoretical premises, media can be seen as an *obstacle* or a *facilitator* to EU integration and European democracy. However, the dividing line cannot be reduced to a mere dichotomy between new and old means of communication. Media do not represent only an important political channel to inform and shape public opinions. They can, in fact, be dynamic actors in negatively framing the EU, and their bias produces direct effects in terms of public discontent (Galpin and Trenz chapter three). Significant events, such as the Eurozone crisis, increased the degree of Euroscepticism that has also become mainstream in quality newspaper (Bijsmans, chapter four).

On the demand side, in the emblematic cases of the UK and the Netherlands, evidence has also shown that newspaper readership is related to the common perception of journalists’ political bias on the left-right spectrum, as well as to the position they adopt in favour or against the EU (Leruth, Kutiyiski, Krouwel and Startin, chapter five).

Intensity in the use of news or social media also affects public preferences towards the EU. New media tend to capture more attention from young people while emphasising Euroscepticism, while traditional media tend to frame the EU in positive terms. Consequently, media framing effects can be seen in citizens’ Eurosceptic or supportive attitudes towards the EU (Conti and Memoli, chapter six). Similarly, despite context-related differences, voters’ *news diets* and party preferences relate to their positions towards the EU. Again, social media represent the main facilitators of EU discontent (Mosca and Quaranta, chapter seven). The internet in particular is the arena where extreme-right’s anti-EU rhetoric spreads cross-nationally while fuelling political discontent (Pavan and Caiani, chapter eight). Twitter networks show a clear distinction between Europhile and Eurosceptic camps. Social network analysis highlights that Europhile networks interact more transnationally than Eurosceptic ones (Heft, Wittwer, Pfetschnineth, chapter nine). In contrast to ‘hard’ Euroscepticism, the austerity policies’ effect produces a ‘soft’ EU discontent. This can be seen in the claims of movement parties such as *Syriza* and *Podemos* that support the idea of ‘another Europe’ rather than being completely against it (della Porta, Kouki, Fernández, chapter ten). Moreover, technologies may be used to develop a new model of citizenship and political representation that transcends national borders. A long-term and sophisticated EU ‘u-government’ model would be shaped by a mixed reality technology (Fanoulis and Peña-Ríos, chapter eleven). As further argued in the conclusion, moving beyond Eurosceptic parties’ strategies is an essential starting point to better understanding the different shades in which EU discontent manifests itself (chapter twelve). Empirical evidence shows a nuanced Euroscepticism and provides substantive arguments for further investigating this highly-contested phenomenon through a bottom-up approach.

Finally, in so-called “times of crisis” this book outlines the state of the art on the theoretical and empirical implications that sit behind different Eurosceptic labels. A positive connotation of EU contestation is adopted and, often, shared as a dominant frame across countries and actors like citizens, social or political movements, and the extreme right. The book proposes new and interesting stimuli to the study of dissent towards the process of European integration. It also highlights the double-sided role media play as agents and arena for political conflict. By so doing, it represents a valuable starting point for further studies on European politics and political communication.

Ornella Urso, *Scuola Normale Superiore, Pisa*

* * *

PAOLO CHIOCCHETTI, *The Radical Left Party Family in Western Europe, 1989-2015* (London, New York: Routledge, 2017). 244 pp., £88.00 (hardback), ISBN: 9781138656185

The book is based on an analysis of quantitative data, historical records and public statements that characterized the parties of the radical left in seventeen countries in western Europe, from 1989 to 2015. The analysis of the individual parties follows four lines of research: the electoral strength, organizational characteristics and the political strategy. The strength of the parties is measured both in absolute terms and in the systemic context, with the number of votes, of members, of parliamentary seats won and the members’ ability to influence the government, in particular the ability to increase public expenditure in terms of the GDP. The book proposes a “new holistic approach” to conceptualizing and analyzing the party family of the radical left and is developed along three dimensions: the family of radical parties, the individual parties, and the most important fields of investigation. The family party is characterized as opposed to the dominant tradition of social democracy.

According to the author, firstly relating the specific identity of the radical left in terms of class clearly distinguishes it from other family parties, maintaining at the same time its internal pluralism; as well as enabling its changes in space and time to be understood whilst maintaining a cognitive compass.

In the first chapter the author outlines the theoretical and methodology framework of the book. Chiocchetti defines the new European left as the family which responds to the class and communist left, which is separate and distinct from the dominant tradition of social democracy; and acknowledges its constitutive pluralism and historicity. The second chapter reconstructs the parable of the radical left in Europe from 1914 to 1988. It originated as a radically anti-capitalistic branch of labor socialism: it was divided between the defense of the soviet model and “real existing” socialism and the acceptance of a reformist model centered on the

redistribution of wealth and the expansion of welfare. The third chapter reconstructs the panorama of the radical left in Western Europe after the historic breakup in 1989, until 2015. In fact 2015 represented the turning point in the history of the radical left, which in three countries gained exceptional electoral success (45.0% in Greece with SYRIZA, 25.8% in Spain with PODEMOS and 21.5% in Portugal with BE and PCP – p. 66). Instead, in the three major countries of the Eurozone (Italy, France, Germany) its electoral proposal did not meet in the same way with the favors of the electorate, who preferred other center and right policies. Chapters 4 to 6 deal with three special cases of the development of the European left in Germany (“A success story”), Italy (“History of failure”), and France (Failure or success?). The final chapter attempts a comprehensive reading of the political trajectory of the left in Western Europe after 1989 (“Filling the vacuum?”).

The disappearance of the Soviet Union and the profound crisis of the years 1989-1993 almost led to the disappearance of the revolutionary left of the communist matrix, to the dispersion of their members and voters. The new revolutionary left that emerged from its ashes has highly diversified characteristics, although it is mostly composed of long-standing militants of the communist movements, Trotskyists, Maoists and of the socialist left. Some legacy of the 20th century communism continues within it but only a small minority cultivates this legacy as the basis of a political project. The vast majority tried to amalgamate very different political-cultural references: Marxism, Keynesian, anarchism, social democracy, libertarian left, radical democracy, environmentalism and populism, trying to create a “modern” political organization, which is pluralistic, inclusive and hostile to neoliberalism. The political project is vague and shaky, evoking an idea of transition toward a distant socialist society, towards an anti-capitalism system defined now as communism, now as democratic socialism; a society that affirms the primacy of man over profit. In fact, the political identity of the radical left is undefined, ideology has little to do with political daily choices. It tries to stay focused on the representation of the interests of the working classes, the defense of the welfare state, and the promotion of the values of the libertarian left.

Engaged in the tradition of the communist left and revolutionary socialism, the radical left in contemporary Europe had moved by the 1990s toward a new ideological identity centered on antiliberalism and has thus claimed to be the authentic heir of both historical communist organizations, both of the socialist tradition and of the libertarian left. The radical left appears today as the product of three distinct elements: the decline of the historic tradition of communism and of the socialist left; the adoption of the founding themes that characterized the social democrats and the ecologist left in the 1970s and early 1980s; and finally it is the product of a new anti-neoliberal reflection.

The radical left must contend with three challenges and many contradictions: coherence between an antiliberalist position and unity of the center-left; between anti-neoliberalist and anticapitalist; and between loyalty to the tradition and the requirements posed by the economic and social transformations. Consequently, there are difficulties in relations with the other left-wing parties (social democrats

and environmental movements), the ideological oscillations, the organizational weakness and the continuous oscillations between fragmentation and regrouping.

According to Chiocchetti, the great recession of 2007-2008 offered new and great opportunities to the radical left in Europe but, at the same time, it highlighted its limitations. The non-homogeneous electoral successes of the radical left are a barometer of widespread rebellion in many areas of European society against the dominant neo-liberalism, but certainly they are not the only barometer. The successes of the new radical European left are the fruit of anti-austerity mobilisation however, it has not gained success everywhere or been seen as an alternative to existing coalitions of center-right governments or as a partner for the governments of the center left.

Radical left, neo-communist left, revolutionary left? Or just left? The question that arises after reading this thorough research concerns precisely the political-ideological boundaries of the study. And perhaps the answer is that however you translate this “radical” nature, it is still difficult to speak of a “family party”, of a radical left family in western Europe after 1989: though the author actually believes this is possible.

I believe that what Chiocchetti’s careful and very detailed reconstruction does is to highlight the great differentiation among partisan subjects who would like to be grouped into a single family. The non-homogeneity between parties who share the same anti-liberal orientation is very strong and, above all, does not tend to decrease over time, as demonstrated by the evolution of the parties of this shaky radical left in the period following the end of Chiocchetti’s research (2015) until today.

Carlo Baccetti, *University of Florence*

* * *

STEFANIA PANEBIANCO (ED.), *Sulle onde del Mediterraneo. Cambiamenti globali e risposte alla crisi migratoria* (Milan, Italy: Egea, 2016). 230 pp., €24.30 (paperback), ISBN: 9788823845275

In 2016, over 180,000 migrants crossed the Mediterranean in the attempt to reach Italy. After the March 2016 deal between the European Union and Turkey, the Central Mediterranean migratory route, heading from the Western Coast of Libya towards Sicily, has become the largest avenue of irregular migrations to Europe. The death toll – amounting to around 4,500 estimated casualties in 2016 only – has turned the Mediterranean into the theatre of a complex humanitarian emergency.

In spite of the policy-relevance of migrations across the Mediterranean and its salience in the public discourse, academic research on the subject has lagged behind. To be sure, existing scholarship has shed light on different aspects of the phenomenon, such as the securitization of migrations, EU and non-governmental organisations’ law enforcement and Search and Rescue (SAR) operations, and pub-

lic opinions' (mis-)perceptions of migrations to Europe. There is, however, hardly any research seeking to bring together all these different dimensions in order to provide a truly comprehensive overview of the so-called migration crisis. The absence of such an overarching analysis is regrettable, as all the issues mentioned above are tightly intertwined and can hardly be investigated in isolation. Decision-makers' policies are informed by public opinions' perceptions, but also shaped by social norms, and international law and institutions. Consequently, an in-depth explanation of the migration crisis requires a thorough examination of the material, institutional, and ideational factors affecting foreign and domestic policy decision-making processes.

Sulle onde del Mediterraneo. Cambiamenti globali e risposte alle crisi migratorie (On the Waves of the Mediterranean. Global Changes and responses to migratory crises) – edited by Stefania Panebianco – is the first attempt to provide such a comprehensive analysis. Based on an impressive amount of empirical research conducted at the University of Catania within the framework of the research project 'FIR 14', the volume systematically examines the nature, drivers, and implications of the Italian response to the latest surge in maritime migrations by analysing the phenomenon in each of its most relevant aspects.

Fulvio Attina's introduction places the present crisis within the framework of the academic scholarship on migrations and EU migration policies, a subject examined more in-depth in Francesca Longo's and Rosa Rossi's chapters. Longo's chapter examines the evolution of EU asylum and migration policies, arguing that EU policies are no longer capable of addressing large-scale migratory flows. Most notably, the Dublin regulations – which oblige refugees to embark in a dangerous journey and apply for asylum in the country of first entry – should be reconsidered to both guarantee a better protection of refugees and ensure fairer burden sharing across EU member states. Rossi's chapter broadens the perspective to other international organisations, presenting elite survey data of Italian elite perceptions of international organisations' response to the crisis. The contribution by Luigi Caranti goes beyond a legalistic understanding of the obligations enshrined by European and international law by examining the moral underpinning of the duty to rescue and provide for refugees and economic migrants alike.

The chapters by Stefania Panebianco and Daniela Irrera then turn to the operational aspects of the migratory crisis offshore Libya, examining SAR and law enforcement operations. Panebianco's contribution focuses on state-led migrant rescuing, and most notably the Italian Navy operation Mare Nostrum, launched in October 2013. In spite of being capable of rescuing over 150,000 migrants, Mare Nostrum was discontinued after one year due to Italy's frustration over the lack of burden sharing and other European states' criticism that the operation was a pull factor on migration. As argued by Panebianco, while Mare Nostrum did not become a template for future EU operations, it at least succeeded in putting maritime migrations at the centre of the EU policy agenda. The EU maritime operations that followed Mare Nostrum, Triton and EUNAVFOR Med do not have SAR as their primary mandate. Consequently, a number of non-governmental organisations started to conduct their own migrant rescuing operations to try and fill the gap left

by the end of Mare Nostrum. Daniela Irrera's chapter focuses on the non-governmental provision of SAR, providing data that illustrate NGOs' contribution to mitigating migrants' loss of life.

The chapters by Simona Gozzo, Fulvio Attina' and Rossana Sampugnaro rely on a large amount of survey data to look at how existing Italian and European policies are perceived by elite and public opinion alike, thereby providing an ideal conclusion to the volume.

As epitomized by this short summary, the volume edited by Panebianco examines Italy's perceptions of and response to the migration crisis in a truly comprehensive and multidisciplinary fashion, combining the use of survey data, the in-depth examination of existing policies and the legal frameworks they are embedded in, as well as legal normative perspectives on the responsibility to rescue and welcome migrants. *Sulle onde del Mediterraneo* is not only an important reading for scholars of international relations, comparative politics, political theory, and international and European law alike. Thanks to its empirical richness, comprehensiveness and clarity, the collection also provides a useful compass for the wider community of informed readers seeking to navigate a public debate that often provides more heat than light on such a complex subject.

In spite of its merits, Panebianco's volume cannot – nor does it seek to – provide a conclusive examination of the subject, suffering from the inevitable limitations associated with the timeliness and complexity of the issue it investigates. Given the ongoing nature of the migration crisis, examining the phenomenon is like shooting at a moving target. This leaves ample room for further research looking at very recent developments such as the growing role of and mounting criticism against NGO SAR operations. Likewise, the in-depth, empirically rich examination of the Italian case provided by the volume can only occur at the price of renouncing a larger comparative analysis. Contrasting the policy responses to migrations across the Central and Eastern Mediterranean routes would allow future research to better investigate the role played by public opinion and European and international law, norms and institutions in shaping policy responses to migration crises in Europe and worldwide.

Eugenio Cusumano, *Leiden University*

* * *

GIANLUCA PASSARELLI (ED.), *The Presidentialization of Political Parties. Organizations, Institutions and Leaders* (London, United Kingdom: Palgrave Macmillan, 2017). 300 pp., \$39.99 (paperback), ISBN: 9781349577675

This volume is an important contribution to the field of comparative political institutions because it focuses on the growing role of party leaders who assume a relevant institutional power in many advanced democracies. The notion of “presidentialization” is a

dentialization” is at the core of the volume’s theoretical framework. However, in contrast to other pieces of empirical research emphasizing the impact of institutional changes on the development of party structures, the volume endeavors to explore the phenomenon of the increasing importance of party leadership independently from the evolution of the institutional setting. This is, as noted by the editor in the introduction, the “missing link” in the study of presidentialization. More precisely, Passarelli aims to explain the varying intensities of “party presidentialization” one can observe by comparing certain countries using a simplified framework built on two separate dimensions: *institutional presidentialization* and *party genetic presidentialization*.

First, this volume is a very interesting contribution to our understanding of the changing role of party (and institutional) leaders in many contemporary democracies because it emphasizes the complicated relationship between historical party transformations and macro-institutional (or constitutional) changes. To solve this puzzle, Passarelli proposes a systematic analysis of the multifold dynamics of the process of *party presidentialization*), which should be in conflict with the “natural” attitudes of any European parliamentary democracy (dominated by collegial executives controlled by legislative bodies). Supposedly, it is very likely that a pure presidential form of democracy has been established in those systems. The structure of the volume confirms that such dynamics are, in the real world, much more compound and mutable, corroborating the arguments presented at the end of the comprehensive review on the literature on presidentialization and party personalization provided by Passarelli in his introduction. This ambitious proposition can be somehow tested by a large country-by-country comparison. For this reason, the rest of the volume, shaped on the idea of an extensive research strategy, includes eleven chapters devoted to different cases of parliamentary, semi-presidential, and presidential systems, thus covering a significant variety of political systems and democratic experiences.

Further, other examples of “institutional presidentialization” are covered in the first part of the volume, where the systems in Chile, the United States, and Brazil and the peculiar “semi-presidential” case of the French Fifth Republic are included. The cases of parliamentary democracies (or “premier-parliamentary democracies” like those recently developed in Central-Eastern Europe) included in the second part of the volume are also rather different from each other. For example, Poland and Ukraine represent the family of “newcomers,” while a good sample of the Western political systems from the UK to Australia and from Germany to Japan and Italy covers an evident variability including the typical “Westminster” and “power-sharing” examples of democracy.

Such a research strategy proves very useful in unveiling the complicated set of factors determining a great deal of variance in party presidentialization. The study of a relatively neglected case such as Chile (Chapter 2) shows, for instance, how the impact of party organizations has been rather malleable since the end of the Pinochet regime. On the other hand, some parliamentary democracies show that despite their stable rates of democratic performance and practices, their overall rates of presidentialization (or “missing presidentialization”) have changed considerably

over time. The presence of specific institutional devices and the emergence of hierarchical party organizational cultures, for instance, have determined high levels of personalization in Germany (Chapter 10) and, to some extent, in the UK (Chapter 8). Conversely, the expectation of crucial “majoritarian turns” connected to the emergences of strong leadership and the consolidation of personalized styles of electoral campaigns did not come true in typical power-sharing democracies, such as Italy (Chapter 12). This is due to the persistence of several institutional and partisan characteristics. In other words, following the theoretical framework used by Passarelli, personalization cannot be a surrogate for the absence of institutional presidentialization.

We have no space here to cover the myriad of findings included in all the empirical chapters of this rich yet extremely harmonized collection of studies. We can simply say that the deep complexity of party presidentialization emerges in all the diachronic analyses included in the volume. This brings the reader to the conclusion that different factors must be considered to understand the comparative evolution of the phenomenon. Among them, historical path dependencies and the different developments of the constitutional settings—including the actual powers of the legislatures, the various steering capabilities of the executives, and, not least, the electoral regimes—are revealed as crucial variables. However, historical transformations of party organizations can make the difference, especially when they are originated by the peculiar visions of strong, long-standing leaders.

The editor’s final chapter provides the volume with a precious element of comparative assessment. According to Passarelli, many of the empirical findings provided in the country chapters support the idea that the typical approach based on the role of institutional setting on the transformation of party leadership should be somehow completed considering the findings of a more comprehensive comparison of the evolution of party genetic presidentialization. This is the main message the reader receives from this volume, which seems to pave the way to a new generation of studies based on the idea of a mutual interactive influence between the macro-institutional framework and the historical evolution of the most significant and long-standing political organizations at the core of democratic competition.

As always happens to any new path-breaking piece of research, the highest, most provocative point becomes the weakest (at least in terms of empirical robustness), and the most debatable argument surfaces at the end. In this case, the map portraying the dynamics of party presidentialization in the 11 political systems covered by the volume (p. 257) looks impressionistic and, as admitted by the same author, rather vague, especially if one looks to the relative distance between the measures of the party genetic presidentialization dimension. However, the implications discussed by Passarelli are fascinating. The phenomena of presidentialization and party leader personalization must be discussed in their continuous interactions under an adequate comparative research framework. By providing important evidence for such a basic but not irrelevant proposition, this volume thus proves to be an important text for scholars concerned with the future of party politics and the perspectives of political leadership within the democratic sphere.

Luca Verzichelli, *University of Siena*