

**THE ASSESSMENT OF RESEARCH QUALITY
LA VALUTAZIONE DELLA QUALITÀ DELLA RICERCA**

**GIOVANNI MASINO
UNIVERSITÀ DI FERRARA**

Abstract

A reflection on the research quality assessment process in the Italian academia is provided. It is argued that both the general logic and several specific technical details of the current assessment process is inconsistent with the evidence that decades of organizational research have shown on the most common organizational solutions adopted by companies and research centers that are able to systematically generate significant levels of knowledge production and innovation. It is also argued that a system like the current one will create, in the medium-long term, serious negative consequences on the Italian research system, both in cultural and behavioral terms. A few suggestions about how to change the assessment system are provided.

Keywords

Assessment, Research quality, Innovation, Organizational choices, Knowledge.

The assessment of research quality / La valutazione della qualità della ricerca, Masino Giovanni. Bologna: TAO Digital Library, 2019.

Licenza: CC BY-NC-ND 4.0
© Copyright 2019 degli autori

ISBN: 978-88-98626-21-2
DOI: <http://doi.org/10.6092/unibo/amsacta/6286>



The TAO Digital Library is part of the activities of the Research Programs based on the Theory of Organizational Action proposed by Bruno Maggi, a theory of the regulation of social action that conceives organization as a process of actions and decisions. Its research approach proposes: a view on organizational change in enterprises and in work processes; an action on relationships between work and well-being; the analysis and the transformation of the social-action processes, centered on the subject; a focus on learning processes.

The contributions published by the TAO Digital Library are legally deposited and receive an ISBN code. Therefore, they are to be considered in all respects as monographs. The monographs are available online through AMS Acta, which is the institutional open archive of the University of Bologna. Their stable web addresses are indexed by the major online search engines.

TAO Digital Library welcomes disciplinary and multi- or inter-disciplinary contributions related to the theoretical framework and the activities of the TAO Research Programs:

- Innovative contributions presenting theoretical or empirical analysis, selected after a double peer review process;
- Contributions of particular relevance in the field which are already published but not easily available to the scientific community.

The submitted contributions may share or not the theoretical perspective proposed by the Theory of Organizational Action, however they should refer to this theory in the discussion.

EDITORIAL STAFF

Editor: Bruno Maggi

Co-editors: Francesco M. Barbini, Giovanni Masino, Massimo Neri, Giovanni Rulli, Angelo Salento

International Scientific Committee:

Jean-Marie Barbier	CNAM, Paris	Science of the Education
Vittorio Capecchi	Università di Bologna	Methodology of the Social Sciences
Yves Clot	CNAM Paris	Psychology of Work
Renato Di Ruzza	Université de Provence	Economics
Daniel Faiña	Université de Provence	Language Science
Vincenzo Ferrari	Università degli Studi di Milano	Sociology of Law
Armand Hatchuel	Ecole des Mines Paris	Management
Luigi Montuschi	Università di Bologna	Labour Law
Roberto Scazzieri	Università di Bologna	Economics
Laerte Sznclwar	Universidade de São Paulo	Ergonomics, Occupational Medicine
Gilbert de Terssac	CNRS Toulouse	Sociology of Work

ISSN: 2282-1023

www.taoprograms.org – dl@taoprograms.org
<http://amsacta.cib.unibo.it/>

Pubblicato nel mese di dicembre 2019
da TAO Digital Library – Bologna

The assessment of research quality

Giovanni Masino, Università di Ferrara

Introduction

The assessment of research quality is one of the most debated subjects in the Italian academy. The discussion concerns the appropriateness of both an assessment process tightly connected to the financing of Universities, the methods that are utilized and its consequences in terms of research quality, productivity and researchers' behaviors.

In simple terms, one could distinguish between opinions that support the general usefulness and / or the appropriateness of the current methods and others that, on the contrary, are very critical of the specific methods and the general logic that is currently utilized. This is a relevant issue because the consequences of the assessment process concern, directly or indirectly, all the areas of interest of the academic community – not just research, but also teaching and other activities through which universities influence the social, economic and cultural development of the Country. The debate happens informally and formally at various levels. While the various positions are often rooted in deep reflections and beliefs, and discussions are quite lively, I believe that not always such discussions are based on sufficient, specific knowledge. It seems that, paradoxically, the scientific community not always utilizes a “scientific” approach to this issue. Quite often data are anecdotic, and ideas do not derive from scientific disciplines that could be useful to at least guide the discussions. Such paradox leads me to propose this article. The goal is not to provide a complete picture on the issue, but to offer a point of view based on a specific research literature, the field of organizational theory and studies, in which pertinent knowledge about the issue can be found but it is not utilized in the debate. Obviously, this is not the only research literature that could provide useful insights. There are already other contributions from different fields. For

example, some authors propose critical analysis on methodologies and indicators used in the Italian assessment process (Baccini, De Nicolao 2016; Abramo, D'Angelo, 2015; 2016), with a specific focus on technical and statistical problems. It is also important to recall the reflection by Viesti (2018). The author describes an exhaustive and very concerning picture of the Italian academic policies in the last few decades. Viesti illustrates the issue of the research quality assessment, among others. These are important contributions. However, I believe that the organizational literature may provide ideas and evidence that are complementary to statistical and political sciences contributions.

It is necessary to specify that I will use the term “assessment of research quality” in a broad way. I will not only refer to the assessment that is used for the distribution of government funds (the so-called “VQR”), but also to other assessment processes that are established for other goals, such as the scientific qualifications of researchers and professors, career advancements, attribution of local funds and others. These are assessment instances that, while slightly different in some technical and procedural details, are all informed by the same cultural background, one that sanctions according to criteria that are similar or identical in their fundamental approach and logic. Thus, I believe that many of the general reflections that follow can be relevant for all these instances.

Why should we assess the research quality?

It is essential to begin from the following question: why should we assess the research quality? Trying to answer such question implies reflecting on the possible objectives of the assessment process. If the objectives of the assessment are not clear, it is impossible to propose any argument about its effectiveness. Any assessment may have multiple objectives: general, specific, direct, indirect, etc. In general terms, we agree with Vergnani (2005), when he defines the assessment as an action that attributes value. However, in the specific case of research quality assessment we should go beyond such definition. The mere attribution of value does not seem a satisfactory goal. One should also consider the *generation of value* as a desirable goal, since the assessment may contribute to

improve the assessed processes in relation to some relevant dimensions. According to this perspective, it is possible to distinguish two general classes of purposes aimed at improvement.

The first purpose is to identify / measure the level of achieved outcomes in the past. This may be useful for several sub-goals, such as increasing the self-awareness of individual and / or collective strengths and weaknesses, identifying processes and resources that have allowed to achieve the desired results, comparing one's achievements with competitors or other comparable entities, and so forth.

The second purpose is to orient the behaviors and improve the future outcomes. It concerns the ability to transform the system, its processes and resources, in order to create conditions allowing to pursue the desired outcomes.

Thus, the first purpose refers to the past, as it aims to provide a description of what already happened. The second purpose refers to the future, as it aims to identify improvement opportunities.

It is quite obvious that the two purposes are sequentially connected, as the second one is hardly achievable unless the first one is also achieved. It is not possible to imagine conditions for improvement unless one is aware of the outcomes that have been already achieved.

It is important to consider that if the assessment is implemented in a way that generates any kind of consequences on subjects and resources implied, then there is going to be an unavoidable effect in terms of orientation of future choices of such subjects. Even just a simple communication about the current state of affairs, or the achieved outcomes, may have significant psychological influences on individuals. The mere fact of knowing whether one is considered a "bad performer" or a "good performer", regardless of practical consequences, may change the behavior of the subject. The same is true when prizes or sanctions are applied in relation to past performance, even without any reference to the future, just because expectations are created about the utilization of the assessment results. Even more so, obviously, when the assessment results are explicitly used

to orient future behaviors, for example when the former are used as criteria to allocate resources, career opportunities, etc.

Thus, if on the one hand there is a logical connection between the two purposes – the realization about the past, and the creation of new conditions for the future – there is also a “practical” connection: it is very hard, maybe impossible, to engage in an assessment process without having direct or indirect effects on future behavior. This is a very significant aspect, as it emphasizes that fact that when mistakes are made in designing or implementing an assessment system, there are negative consequences in the medium-long term which may become much more significant and serious than one may be able to predict. I even argue – and I will explain later why the following consideration is particularly pertinent to the case of research quality – that the negative effects of a badly designed assessment may generate a worse final outcome than an absence of assessment.

If we turn to the assessment process, one should ask the following: what are the outcomes that it should achieve? If we talk about research quality, then it should be obvious that the desired outcome of an assessment should be the improvement of the scientific production. The problem is how to articulate such a general concept into more specific terms. I believe it is possible to propose some essential elements that all parties involved would easily agree upon – in fact, I argue that these basic elements should be very obvious, even tautological. They should be so basic and simple, but specific just enough to allow a useful reflection on the current research quality assessment process. I would like to insist on the need to stop at very minimal specifications, as I believe that more detailed articulations are not necessary and, indeed, they may even be damaging for the discussion.

The first specification is that “quality” does not coincide, necessarily, with “quantity”. Even a single research discovery may change the world, may initiate new scientific paradigms and characterize an era. There are innumerable examples. On the side of the spectrum, a large quantity of marginal, redundant, scarcely significant scientific results may have, even if taken together, a negligible

“impact” on the scientific progress. I am not saying, of course, that the quantity of scientific production should not be considered in the assessment process at all. It should be. One cannot deny that, quite simply, it is better to achieve a larger number of significant scientific results than a smaller number of analogous results. I am just arguing that “quality” should not be confused with “quantity”, as the two dimensions refer to different aspects. One cannot achieve good research quality just by aggregating a large number of mediocre studies. This is the first, entirely obvious specification. Notwithstanding its simplicity and obviousness, we will see that this specification is quite important if we want to reason about the current Italian research quality assessment process.

The second specification concerns the nature of “quality” in the field of scientific production. What does it mean to have “high quality” research? Again, I am proposing a specification so basic that seems hard to refute: every scientific result to which we would attribute a non-zero value should improve our knowledge about the world. The production of useful knowledge is the final “proof” to which no scientific result could be exempted from. Obviously, knowledge production may happen in many ways, but not in *any* way. Let us see some examples.

First, research may generate new knowledge on a phenomenon on which no knowledge was available in the past. In this case, we have a maximum degree of novelty.

Also, new knowledge can also be generated about a phenomenon about which incomplete knowledge was already available. In this case, the novelty is lower, but still significant.

Third, new knowledge can be confirmative or corrective about a phenomenon in relation to which the available knowledge was not adequately validated or even imprecise or incorrect. Thus, this new knowledge is useful to decrease uncertainties and / or to correct past imprecisions and mistakes. In this case the novelty is lower, but usefulness – in other words, the progress that is generated in the capacity to understand the world – is positive.

A final case should be added, one where the scientific production aims at providing a new interpretation, or a new way of explaining a certain phenomenon based on a different epistemology from the one upon which the previously available knowledge was based. This is a frequent occurrence in social sciences, where different epistemological choices lead to different interpretations about the same phenomenon. The debate that is triggered in this case concerns different explanatory modalities. Thus, the progress of knowledge concerns the improvement of critical reasoning and analysis which the debate itself enables. The novelty of knowledge may be more or less significant, but the value generated by a competition of ideas is undeniable, at the very least because it helps scholars to develop new understandings of reality.

As we have seen, there are different ways of identifying the production of knowledge. The degree of novelty and usefulness may vary, sometimes quite significantly. In some cases, there may be self-proclaimed “scientific products” which, in reality, provide very little or even no valuable knowledge about the world, for different reasons. Maybe because the study merely replicates results that are already available and adequately validated, or maybe because it is just a “repackaging” of already available analysis, or simply because it contains methodological or conceptual mistakes which compromise its validity. Regardless of these “extreme” cases – which are not necessarily rare – the key point is the following: if one agrees that research quality is fundamentally connected to the production of new / valuable knowledge about the world, then one must logically conclude that any assessment of research quality should mostly focus on just that, and avoid (or at least be extremely cautious about), assessments based on “proxies”, that is, parameters and indicators that are presumed to be correlated with the production of knowledge. Such correlation must be proven with highest attention and rigor, otherwise the assessment will concern “something” unrelated to the new, valuable knowledge – that is, not related to actual research quality.

To summarize: if one accepts that the purpose of research quality assessment is not merely to describe the outcomes of past research processes –

hence, not just an “attribution of value” – but also to orient future behaviors – hence, to “generate value”, and if one also accepts the research quality strictly concerns the production of valid knowledge about the world, then the design and implementation of research quality assessments could greatly benefit from studying of those organizational contexts in which the production of knowledge and innovation has reached systematic, long term, positive results. There is, indeed, a vast literature that for decades has studied, with different methodologies, exactly how the most innovative companies and research centers are able to generate new knowledge and innovation in a successful, systematic way. It should be possible to learn useful lessons about the organizational conditions that facilitate the production of knowledge and innovation. There are general organizational principles that are well known and sufficiently reliable. However, it should be noted that these principles cannot be transformed into “models” or “formulas” that guarantee a perfect replicability of processes and results. These are not “best practices”. These are organizational conditions that seem to facilitate the production of knowledge and innovation. These conditions are translated by each company in specific choices, practices, techniques and processes, which are entirely or at least partially distinctive of the specific situation. Every organizational context is unique, and cannot be perfectly reproduced in space and time. Nonetheless, those principles are still quite useful to reflect about “what to do” in order to create favorable conditions to the production of knowledge in different contexts, such as the academic research. Here, I will propose the following exercise: a) I will summarize some of these well-established principles, as studied and validated by decades of organizational literature; b) I will illustrate the (mis)alignment between such principles and the organizational conditions that the current Italian research assessment system is implicitly and / or explicitly promoting; c) I will introduce some general proposals in order to better align those principles and the organizational conditions of the academic research system.

Organizational principles that facilitate the production of knowledge and innovation

The literature on the subject is quite extensive, so it is not possible here to explore in detail every single relevant aspect. A few premises are necessary. As it is always the case in social sciences, there are competing theories, different empirical studies carried out through different methodologies. The the conclusions may vary to some degree. Also, there are atypical organizational cases which may differ from what is observed in the majority of cases, usually because of very specific contextual, cultural or historical reasons. Finally, one should also consider that this literature is continuously evolving, and new ideas and results emerge all the time. Notwithstanding all these premises, I believe it is possible to identify some general principles that are sufficiently validated and uncontroversial so that they can be considered as useful, reliable guidance about how to create organizational conditions that facilitate the production of knowledge and innovation. The following list is not complete, and some of these principles are tightly connected with each other. Again, the goal is not to propose an in-depth review of these ideas, but to emphasize the relationship between this area of the organizational literature and the organizational conditions that are directly or indirectly promoted by the Italian research quality assessment system.

General approach to the creation conditions favorable to innovation

Decades of studies clearly show that the most innovative companies and research centers conceive the production of knowledge and innovation as a collective process – in other words, the outcome of shared activities by a community which operates within specific enabling conditions (Sawyer, 2007). This does not mean, of course, that the production of knowledge cannot results, sometimes, from a strictly individual process. It does, occasionally. However, the idea of a solitary genius who, in perfect solitude, comes up with a revolutionary idea (a very common picture in the public imagery about science) is the exception, not the rule. What differentiates organizational contexts that systematically are able to generate innovation is their collective, organizational

conditions that allow individuals and groups to be at their best in terms of creativity, problem solving and motivation. The most effective approach is not to hire individual geniuses, but to create the conditions that enable competent, motivated individuals to combine into organized groups which may become a “collective genius” (Hill *et al.*, 2014).

There is a very clear, radical difference with the approach that is implicitly encouraged by the current Italian quality research assessment system. There is nothing, in that system, that allows to assess those contextual conditions. There is no assessment of the collective elements. And even when an apparently collective assessment seems to be attempted, the focus is always on the outcomes, not the conditions, while the “collective” dimension is only conceived as a mere aggregation of individual performances. This is a truly fundamental difference which has cascade effects on many other aspects, as we will see shortly. One may argue that this is a truly general conception of the assessment process in antithesis with what the available organizational literature clearly suggests as the most productive approach to knowledge production and innovation.

The motivation of individuals

Literature on creativity shows that intrinsically motivated individuals have higher chances to succeed – this is true in general, but it is particularly true for jobs and activities characterized by a high degree of creativity and complexity (Amabile, 2013). Intrinsic motivation refers to the idea that the activity itself represent the main motivational “trigger” in the workplace. For example, intrinsically motivated individuals are those with a true passion for their job; those enjoying significant discretion or even autonomy in what they do; those enjoying opportunities for personal growth and learning; those with the possibility to understand, agree with, or even contribute to, the goal of their own activity and the purpose of their organizational context (Ryan, Deci, 2000). In other words, intrinsic motivation is about those situations in which individuals attribute to their jobs deep, relevant meanings, a higher sense of purpose, passion and interest.

On the other side of the spectrum of the motivational continuum we find individuals that are extrinsically motivated, where the motivational triggers are “external” to the activity and the job itself, such as various forms of compensation, incentives, prizes, benefits and sanctions. While these are also relevant motivational devices, a vast literature shows their limitations and even their negative consequences – not just in terms of lower long-term satisfaction and well-being, but also in terms of decreased performance. One of the problems is a well-documented “cannibalization” effect, where the presence of extrinsic incentives decreases intrinsic motivation (Amabile, 1998). This dangerous effect is particularly evident in creative, intellectual work situations. Intrinsic motivation, on the contrary, does not show significant negative effects. Thus, organizational context that are more able to generate knowledge and innovation are those that, while providing adequate economic conditions for their employees, are mostly concerned about their intrinsic motivation. Those companies recruit people that are genuinely passionate or at least interested in their activities, allow them to enjoy significant discretion or even autonomy, insist on a participated process of goal definition and provide relevant opportunities for personal growth and learning (Hill *et al.*, 2014).

The current quality research assessment process in Italy seems to move towards the opposite direction. The assessment is used to extrinsically motivate the researchers. The assessment outcomes are used to define career advancements, allocations of funds, even the reputation of individual is increasingly associated to the level of parameters which have nothing or very little to do with both the content of their research efforts (parameters such as the number and type of publication, citation indexes, etc) and their actual contribution to the scientific progress. Obviously, I am not claiming that researchers are not intrinsically motivated. On the contrary, I believe that this is a profession in which intrinsic motivation is still particularly high and widespread. And that is where the problem lies: the assessment system goes toward a direction which is the opposite of what is desirable, and all the literature evidence clearly suggests that such approach generates, over time, disastrous

motivational consequences (intrinsic motivation is replaced by extrinsic motivation) and negative, unavoidable repercussions in terms of what really should matter, the generation of new knowledge and scientific progress.

Tolerance for errors

William McKnight, famous manager of 3M, one of the most innovative companies of the last century, famously declared that one of his biggest concern was to make sure that, every year, a high enough number of ideas and projects failed. In his book “Work Rules”, Lazlo Bock, Human Resource Manager of Google, describes the reasons and the ways through which his company tries to “compensate failure” (Bock, 2015). These are not crazy ideas. A high tolerance for errors, for failed attempts, for projects that do not achieve what they were hoping to, is a very common characteristic of the most innovative companies (Farson, Keyes, 2002). Obviously, such tolerance should not be granted to errors and failures due to incompetence or insufficient effort, but for initiatives that imply uncertain outcomes because of their complexity, projects that try to innovate in a radical way, to go beyond the current frontier of knowledge, to achieve particularly valuable outcomes. All these initiatives have an inherently lower probability to succeed. If a company does not have enough projects of this kind, it is not “daring” enough, it is not trying to “really” innovate. If the rate of failure is too low, that means that the innovation pursued is incremental and not radical, and nothing really “new” is attempted. In these contexts, these “failures” are not considered a waste, but the necessary investment for innovation. Innovation cannot be planned or programmed. What one can and must do is to create the conditions that facilitate it and increase the probability to succeed. A healthy tolerance for “errors” is among the most important conditions for innovation.

The current Italian research quality assessment system, once again, seems to do exactly the opposite. Only “successes” are considered (even worse, these “successes” are ill-defined, as we will see below). There is no consideration for the researchers’ ability to propose new ideas and projects, or for the investments of researchers towards particularly innovative pathways. Even more, the

emphasis that the assessment logic puts on both “quantity” (rather than actual “quality”), especially for career advancements, and on a competition essentially focused on bibliometric parameters (or similar ones), pushes scholars to embrace “prudent” research pathways characterized by a higher probability to succeed and with a shorter term expected life-cycle. A clear disincentive is created towards riskier, more innovative, more long-term oriented projects. The current system does not incentivize courage but prudence, not radical innovation but incremental innovation or, worse, no innovation at all.

Competences and disciplinary focus

History shows that both disciplinary, inter- and multi-disciplinary research processes play an important role in the progress of knowledge and innovation. However, the most significant innovations are often the outcome of inter-disciplinary collaborations. Cross-fertilization of different competences creates the most fertile conditions for the generation of truly new ideas. It is not by coincidence that a widespread concern in the most innovative companies is to make sure that a high variety of available competences is present, as well as a high degree of interactions – collaborations, information exchanges, shared projects etc. – between people with different disciplinary backgrounds (Hill *et al.*, 2014). It is worthwhile repeating that disciplinary research is also important, as it may lead to very valuable results of even radical ideas and innovations. And, in any case, even incremental innovations that disciplinary research is typically able to generate – the refinement, improvement and validation of available knowledge – are certainly essential. But it is a fact that the majority of the most innovative ideas emerge in situations in which the interaction of different ideas, competences and abilities represent the ordinary way of working, not the exception.

Once again, the current Italia research quality assessment system seems to lead to the opposite direction. The emphasis is clearly on disciplinary research. The so called “research products” (the publications) are assessed within a disciplinary context, the lists of scientific journals in which scholars “must”

publish have a clear disciplinary connotation. Not only interdisciplinary research is not incentivized, it is indeed punished. This argument is tightly connected with the one illustrated in the previous paragraph, as interdisciplinary research is more complex and “risky” – meaning that its outcomes are more uncertain and the probability of success is lower. The current assessment system discourages taking risks, and the risk of inter-disciplinary research is particularly penalized.

Collaboration vs Individualism

This element is a corollary of the previous ones, but a few specifications may be useful. As the most innovative companies facilitate the interaction between different backgrounds and competences, an obvious consequence is that the development of new ideas and knowledge can only happen, in such circumstances, through collaborative efforts. The organizational literature suggests exactly that well managed groups – which is not an easy condition to achieve, as empirical research on group pathologies shows – are generally much more effective than individuals in complex activities (Forsyth, 2010). This is true not only when group members are carriers of diverse competences, but it is in such cases that the advantage of collaborative processes may become very substantial. A collaborative approach to research and knowledge production, if carried out properly, may bring quite large advantages in terms of effectiveness, creativity and innovation, even when the available variety of disciplinary background is not so wide. Obviously, a properly designed research quality assessment system should promote collaborative research, rather than hinder it.

Once again, we must observe that the current Italian research quality assessment system does not take into account such fundamental principle. The research “products” are evaluated as individual products. There is no incentive to collaborative research. On the contrary, in some instances (for example, in the evaluation of individuals for career progressions), single-author publications are given more value than the ones with several authors, according to a truly bizarre and self-defeating “logic”. The rationale behind this approach is that it is worthwhile evaluating the “quantity of work” behind every publication. So, it is

assumed that when several authors co-sign a publication they had to work “less” than authors publishing alone. Regardless of the fallacy of such assumption, the very idea that a research quality assessment system should even worry at all about the “quantity of work” rather than the quality of the ideas and the contribution to the progress of knowledge is just absurd.

Leadership and culture

In the most innovative companies, leadership is typically methodological, not substantive. In other words, the role of leaders is not to predetermine the content of the research, which is hard to plan substantively – or even impossible to plan when the goal is to innovate in a radical way – but to create the most adequate conditions that facilitate innovative processes. By “methodological leadership” I mean that leaders play the role of methodological experts, as they implement and enforce the best possible working methods of processes aimed at producing new knowledge. On the contrary, a “substantive” leader predetermines goals and solutions, he intervenes on the merit and the substance of the research process. The latter approach does not work simply because the complexity of innovation and research processes very rarely allows to predict exactly the goals, the solutions, or even just the pathways to possible solutions, and it is even harder to do for single individuals with narrow, specific backgrounds. In Google, for example, leadership in R&D projects is seen as “emergent”, meaning that in every project the substantive leadership is distributed and contingent to specific problems that arise along the way, while the methodological leadership is ensured by an organizational culture that is developed specifically to create the right conditions for innovations, and by individuals that make sure that such culture is translated into the right research methodology (Bock, 2014). The most capable leaders, in the field of innovation, do not play the role of “main actors”, instead they “set the stage” – the methodological conditions – in which the various individuals of the team are in the perfect situation to play “leading roles” along the way, and thrive (Hill *et al.*, 2014).

The current Italian quality research assessment system is generating a culture in which a substantive leadership seems to emerge, rather than a methodological one. First of all, the assessment system seems to encourage conformity rather than openness. For example, lists of predetermined journals are used to assess scientific publications, and the “weight” of each journal is determined by the citation indexes, in other words, by the fact that the journal is well established within the mainstream by mostly promoting dominant ideas, research approaches and epistemologies. Thus, the assessment process influence scholars in substantive terms, in relation to the content of their research projects. Studies that propose alternative, new or just non-mainstream ideas will have a much harder time to be published in journals that are well ranked by the assessment system. The resulting behaviors of scholars – the choice of research themes, techniques, concepts, epistemologies etc. – will be unavoidably biased towards the disciplinary mainstream. A culture of conformity is then produced and reproduced over time. Exploration of new ideas does not pay off. The assessment system indeed works as a “virtual” substantive leader, not a methodological one. A leader that decides, *ex-ante*, what kind of research is worth pursuing, what is valuable and what is not. Which is exactly the opposite of what most innovative companies do.

Form and content

It is obvious, even trivial, that the production of new knowledge focuses mostly on content, not on form. In most innovative companies this is evident in terms of several conditions of working life, in which formal aspects are not so important, because what matters are those organizational elements that are mostly productive in terms of content: from a high level flexibility of working hours to a significant autonomy in terms of working places and movements, the ways interactions are carried out, even the way the physical layout of work spaces are designed etc.

If we consider the research quality assessment system in Italy, it seems that form matters far more than content. This is really evident if we consider the main

“parameters” that define the attributed “value” of research products. A journal article is considered more valuable than a book chapter, a journal is more valuable than a book, an international publication is, by default, more valuable than a national publication. The container prevails on the content – or, more precisely, the container becomes an approximation of the content, and in many cases the content is completely ignored. Obviously, I am not claiming that these “formal” aspects should not be considered at all. There may be some value – however modest – in establishing, for example that, *ceteris paribus*, an international publication should be given more value than a national one. The problem is that “*ceteris paribus*” (the actual content of the research product) is completely ignored. Form replaces content. Publications are less and less assessed in relation to their content. The problem is not just about publications. It may also concern, for example, the ability to propose research projects, the participation to research networks and other similar elements. The production of knowledge must be assessed by focusing mostly (if not solely) on content. Form “proxies” should be used, with great caution, only when strictly necessary.

Illusory results and other problems

The critical picture that I depicted above is just a summary. One could go much deeper. It would be desirable that key institutions (ANVUR, the Government agency in charge of defining and managing the research quality assessment process, the Universities and even the individual Departments) take into careful account the clear, well established suggestions coming from relevant literature about the most effective organizational arrangements that actually improve the quality of research. This is obviously based on the idea the goal of the assessment system is to actually improve the quality of research, not to promote illusory improvements. In this respect, it is worth paying attention to some results that, if not carefully considered, not only may be very deceiving about the actual results of the current system, but also reveal further, serious problems.

Baccini *et al.* (2019a) recently analyzed the 2016 SciVal Analytics report, written for the British Government. In the report is shown that the publications of the Italian Academy not only surpassed the US and other countries in terms of weighted citation impact, but has reached the second place worldwide, behind just the UK. Even more, the report shows that the Italian Academy seems to be well on its way to achieve the first overall position, based on the current trend. Baccini *et al.* (2019a) underscore that Anvur attributed, even with official statements, to the positive influence of the current assessment system this radical, sudden improvement of the quality of the Italian research. But, is this really the case? And what exactly is the nature of this astonishingly positive result?

Baccini *et al.* (2019a) provide an interesting answer. They prove that such result not only is utterly deceiving, but it also shows a serious distortion generated by the assessment system itself. According to the authors, “several studies show that the massive use of bibliometric indexes in assessing the quality of research promotes opportunistic behaviors by the researchers”. The same authors recently published a study in which they show that, since 2010 (the year when the Italian University reform was implemented, and the new assessment system was put in place), the Italian research has seen its “inwardness” rate rise rapidly and very significantly (Baccini *et al.*, 2019b). By “inwardness” the authors refer to the number of auto-referential citations used in a purely opportunistic and instrumental way in order to increase the citation indexes, thereby generating an artificial, unrealistic increase in the number of citations for the Italian research community. According to the authors, “there is only one plausible explanation: the need to achieve the bibliometric goals determined by Anvur created a strong incentive to self-citation and to create citation clubs. These behaviors are so widespread that the value of inwardness of the whole country is greatly increased, both globally and within most disciplines. In this situation, the increased citation impact of the Italian system as recorded in the country rankings is a mirage, explained only by a colossal, collective citation doping” (Baccini *et al.*, 2019a).

It is worth noting the many negative effects that the current system produces. On the one hand, it seems that the system encourages behaviors that one may define as incorrect or unfair, or at the very least not aligned with a proper professional *ethos*, which should imply that citations should be used in relation to the actual content of the publication, not to artificially improve the bibliometric indicators. Extrinsic motivation, as we already argued before, not only replaces the intrinsic one, but it seems to corrupt it. On the other hand, the meaning of “research quality” is completely altered, and it loses any connection with its true, higher meaning – the production of knowledge – as it is demeaned to a mere calculation of citations. Also, and maybe this is the most paradoxical point, one should notice that even if the bibliometric parameters were a good approximation of research quality – they are not, but we can pretend that they are for just a moment – one should still admit that the current system generates unacceptable distortions in relation to the same parameters that it uses: the system is fallacious not only because of its general logic, but it is also technically faulty, because it can be easily “gamed” so that the very result that the system itself promotes (increasing the number of citations) becomes insignificant.

This example is useful in many ways. First, it is an invitation to extreme caution, when considering the bibliometric results of the current system. Second, it represents a clear example of how a badly designed system may generate behavioral consequences that are completely inconsistent with its goals. Every assessment system will have an influence on behaviors. And the outcome of such influence is not easy to predict. A much better approach is needed, one guided by relevant, specific, well established knowledge about the connection between the human and organizational conditions that the assessment system promotes, and the desired outcomes, an approach that the designers of the current system have so far ignored.

Suggestions for an alternative approach

So far, I have described some of the serious problems that characterize the current research quality assessment system in Italy. This is obviously just a summary. A much more in depth analysis could be carried out, and other problems could certainly emerge. I strongly believe that a more careful consideration of what the existing organizational literature suggests in order to achieve better results in terms of innovation and creativity would greatly help in reimagining and redesigning the assessment system. In this paragraph I would like to briefly explore such possibility. Again, this is going to be just a quick description of some general, useful principles. Every principle could and should be operationalized in more detail.

The first point is not really a general principle but rather a very general premise, which nonetheless I think it is worthwhile emphasizing. While the academic context would certainly benefit from a good, properly design research quality assessment system, I would argue that no assessment is preferable to a badly conceived assessment. The distortions of a bad system are so many and so consequential that, in the unfortunate scenario where a significant improvement could not be made, it would be much better to rely of scholars' intrinsic motivation - in other words, to their passion, professionalism and sense of responsibility, which is still a widespread trait of most researchers, notwithstanding the negative press, the "bad apples", the misguided incentives and the gravely insufficient investments in research by the Italian governments in the past several decades. An absence of assessment is not the ideal solution. Proper assessment is important and useful, but if the current system leads to decreased performance on truly meaningful outcomes, to a poisoning of culture, to the corruption of intentions and motivations, then it is better to abolish it. In this specific context, no assessment is much better than a bad assessment.

The second point, just like the subsequent ones, follows logically from what we described in previous paragraphs. It is necessary that the assessment system pays much more attention to the contextual, organizational conditions within which research and scientific production is carried out. Those conditions

should be assessed. It's exactly those conditions that allow to achieve a "collective genius", as stated by Linda Hill and colleagues (2014). At the very minimum, it is necessary that the focus is shifted from individual results to collective results.

The third point is about the focus that should be put on the recruitment phase, to make sure that future researchers are genuinely passionate about science and research and intrinsically motivated. Attention should also be devoted to subsequent moments of training about all the skills and competences that good researchers must acquire. In other words, an assessment of individuals should be mostly, if not only, about inputs instead of outputs: about motivation, competences, abilities. Also, individuals could also be assessed in relation to their contribution to the creation and development of those contextual conditions that enable innovation and knowledge production. For example, the creation of new research collaborations, the active participation to research networks and to the debates and discussions that animate the academic community, the initiative about new projects.

Fourth point: the assessment should be oriented towards the medium-long term. The current obsession for the short term should be avoided altogether. Most ambitious and important projects may require many years to develop, sometimes decades. And if we look at the "form" of the so-called research "products", one may observe that the most innovative ideas sometimes require the length of a book to be fully developed, not the short summary of an article. In social sciences, for examples, there is no doubt that the most significant contributions are historically found in important, classic books - books that literally changed the way we think about reality, books that are the outcome of long-term research projects and programs. Very rarely the same thing can be said for single journal articles.

Fifth point. It is necessary to reduce to a very minimum, in the assessment, the relevance of the "quantity" of the scientific production, while increasing the relevance of actual quality, which should be measured not through unreliable, biased proxies, but through a specific, careful consideration of actual contents of

research. Quality and quantity are not commensurable. In order to assess the quality of a research paper, one has to actually read it.

Sixth point. It is necessary to consider not just the research results, but also the ability to propose, to take initiative, to be proactive, to put forward project ideas. The tolerance for error, that we discussed in previous paragraphs, can be easily translated into a positive assessment of individuals and groups that have the capacity to come up with competent, well informed new ideas and projects, regardless of their final success. Not all projects achieve the desired results, and in a knowledge production-oriented work environment, that is a good thing. Competent new proposals and projects should not be ignored, instead, they should be positively evaluated.

Seventh and final point. It is also necessary to consider positively the ability of the academic system to have a dialogue with the “outside” world, with society in general. On the one hand, it should be defended the right (and duty) of Academia to autonomously define its own goals and preserve its role of cultural guidance. On the other hand, dismissing the value of rich, frequent interactions with the world, of listening and collaborating with external actors, would be to ignore a very important element of a wider and wiser idea of “research quality”.

I am very convinced that each of the previous general points can be transformed into operational processes. For some of them it would be relatively easy, for others it would be challenging but still feasible. And others could be added as well. I do not believe that it is possible to achieve a “perfect” research quality assessment system, but the margin for improvement of the current system are so vast and evident that no technical complication, no difficulty, no imperfection should justify inertia and blind acceptance of the current system and all its distortions. I believe that the negative effects of the current system are propagating quickly, and they can be easily observed in some undesirable behaviors (but “rational”, in relation to the disastrous incentives that are implicit in the current system) by some researchers, especially the younger ones who

“grew up” within the current system and are obliged to respond consistently to such incentives in order to satisfy their legitimate career expectations.

I would like to conclude with a simple invitation: let us rethink the current Italian research assessment system by adopting a “scientific” approach. Let us learn from the various disciplines that may provide effective ideas based on well-established evidence about how to design a system capable of systematically generating innovation. Let us identify those ideas and apply them to the assessment process in order to create an Academia able to achieve research goals that are truly coherent with its institutional mission: the progress of knowledge.

References

ABRAMO G., D'ANGELO C.A.

2015 The VQR, Italy's second national research assessment: Methodological failures and ranking distortions, *Journal of the American Society for Information Science and Technology*, 66, 11: 2202-2214.

2016 Refrain from adopting the combination of citation and journal metrics to grade publications, as used in the Italian national research assessment exercise (VQR 2011-2014), *Scientometrics*, 109, 3: 2053-2065.

AMABILE T. M.

1998 How to kill creativity. *Harvard Business Review*, 76, 5: 76-87.

2013 Componential theory of creativity, in Kessler E.H. (Ed.), *Encyclopedia of Management Theory*: 134-139, Thousand Oaks (CA): Sage Publications.

BACCINI A., DE NICOLAO G.

2016 Do they agree? Bibliometric evaluation versus informed peer review in the Italian research assessment exercise, *Scientometrics*, 108, 3: 1651-1671.

BACCINI A., DE NICOLAO G., PETROVICH E.

2019a Citation gaming induced by bibliometric evaluation: A country-level comparative analysis, *PlosOne*, 14, 9.

2019b Citarsi addosso. Ascesa scientifica dell'Italia? No, solo doping per inseguire i criteri ANVUR, *Roars*, <https://www.roars.it/online/citarsi-addosso-ascesa-scientifica-dellitalia-no-solo-doping-per-inseguire-i-criteri-anvur>

BOCK L.

2015 *Work rules!: Insights from inside Google that will transform how you live and lead*, London: Hachett.

FARSON R., KEYES R.

2008 The failure-tolerant leader, *Harvard Business Review*, 80, 8: 64-71.

FORSYTH D.R.

2010 *Group dynamics*, Belmont (CA): Wadsworth.

HILL L., BRANDEAU G., LINEBACK K., TRUELOVE E.

2014 *Collective genius: The art and practice of leading innovation*, Boston: Harvard Business Review Press.

RYAN, R.M., DECI E.L.

2000 Intrinsic and extrinsic motivations: Classic definitions and new directions, *Contemporary Educational Psychology*, 25: 54-67.

SAWYER K.

2007 *Group genius: The creative power of collaboration*, New York: Basic Books.

VERGNANI A.

2005 Valutazione e sociologia: qualche nota introduttiva, *Studi di Sociologia*, 43: 211-236

VIESTI G.

2018 *La laurea negata. Le politiche contro l'istruzione universitaria*, Bari: Edizioni Laterza.