

https://dspace.org/

A review of South Africa's National Research Foundation's ratings methodology from a social science perspective

Callaghan, C

2018
Academy of Science of South Africa
https://hdl.handle.net/10539/28348
Callaghan, C.W. (2019) African Journal of Science, Technology, Innovation and Development

Downloaded from a DSpace repository. https://dspace.org/

A review of South Africa's National Research Foundation's ratings methodology from a social science perspective

AUTHOR:
Chris Callaghan[1]

AFFILIATION:
[1]Economic and Business Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Chris Callaghan

EMAIL:
chris.callaghan@wits.ac.za

DATES:
Received: 06 Oct. 2017
Accepted: 29 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
scientific methodology; NRF; rating methodology; South Africa; subjectivity bias

HOW TO CITE:
Callaghan C. A review of South Africa's National Research Foundation's ratings methodology from a social science perspective S Afr J Sci. 2018;114(3/4), Art. #2017-0344, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170344

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

© 2018. The Author(s). Published under a Creative Commons Attribution Licence.

One of South Africa's National Research Foundation's (NRF) activities is to award ratings to academics who apply according to predefined categories. Explicitly or not, these ratings are part of submissions academics make for promotions and for employment in South African universities. As such, methodological assessment of the validity of this system is important. This paper seeks to conceptually evaluate certain characteristics of this system against certain general principles of reliability and validity. On the basis of the results of this evaluation, it is argued that assumptions that the NRF rating system is always valid or reliable as a differentiator of individual academics cannot be made unconditionally. Using Management Science as an example of a social science field that draws from multidisciplinary theoretical and methodological frameworks, this paper identifies certain validity issues associated with the current NRF rating system, and makes recommendations for improvements.

Significance:

- Certain validity issues are highlighted and arguments are made to improve the methodology used by the NRF to rate researchers.
- Issues related to multidisciplinarity and mode two knowledge production are considered.
- Technological advances that have made it possible for scientific measurement of research productivity and impact are discussed.
- Problems with subjective methodologies are identified, together with their ethical consequences.

Introduction

If one sought to identify dominant tensions in the literature relating to the progress of science, a candidate would be the tension between Popper's[1] falsifiability thesis and Kuhn's[2] thesis that science progresses as much as a result of changes in human shared values as on the back of scientific advances in their own right. According to Kuhn[2(p.2)], 'science does not develop through the accumulation of individual discoveries and inventions' but through changes in the values and beliefs of scientists, termed 'paradigms' which typically resist evidence-based change until evidence has accumulated sufficiently to tip this balance of beliefs.

According to Still and Dryden[3(p.273)], Kuhn's theory 'seemed to put a distance between nature and scientific practice, and to undermine Popper's principles of demarcation'. What is of critical importance about Kuhn's[2] contribution is perhaps the way human subjectivity is placed centre stage in what was considered 'objective' natural science, thus invoking academic scrutiny around the role of subjectivity in holding back the progression of scientific progress, notwithstanding social scientific critique of objectivity itself and other questions around the legitimacy of the goal of scientific progress itself.[4]

It has long been known that systems theory underlies the workings of human systems, particularly in fields such as Management Science[5], and that there are fundamental differences between the natural and social sciences, not only in methodological approaches but also in terms of focus[6], which have important implications for the tensions between monodisciplinary versus non-monodisciplinary research. This tension is summarised by van den Besselaar and Heimeriks[7] as follows:

> *Interdisciplinarity is an important and complex issue. It is important as modern society increasingly demands application-oriented knowledge, and the usability of scientific knowledge generally requires the combination and integration of knowledge from various scientific disciplines. Traditionally, the disciplines have been very dominant in the organisation of the science system, in the reward system, and in the career system. Nevertheless, funding agencies are increasingly stressing the social relevance of research results, and consequently a new mode of application-oriented research is emerging, on top of traditional academic research.* (p. 1)

These changes have therefore essentially given rise to two modes of knowledge production, and to a differentiation of research according to the extent to which it is disciplinary versus interdisciplinary.[7] This longstanding differentiation is highlighted by Gibbons et al.[8], who argue that these trends 'amount, not singly but in their interaction and combination, to a transformation in the mode of knowledge production', which in turn 'is profound and calls into question the adequacy of familiar knowledge producing institutions' (p.1). Given the differentiation between modes of research described here, and the growing need for applied research seeking to solve societally important problems, which is defined more *by the problem* than disciplinary origin, and therefore necessarily interdisciplinary[7,8], it is argued here that researcher rating systems that are applied in such a way as to discriminate against interdisciplinary research *in the social sciences* can cause harm, as they might disincentivise societally important research in favour of monodisciplinary research, and may give rise to conditions which incentivise

'gaming', or in which research is conducted for the express purpose of meeting the goals of a system, or prioritising these goals at the expense of societal contributions. It is argued in this paper that the societal costs of such a system might be particularly salient in the South African context, and similar contexts, in which localised knowledge is particularly important, yet where localised contexts can be poorly represented in international high-impact journals.

Alvesson and Gabriel[9] decry the standardisation of research and publications 'into formulaic patterns that constrain the imagination and creativity of scholars and restrict the social relevance of their work' (p.245), and which therefore result in the proliferation of non-innovative research publications. This criticism is echoed in criticisms of the culture of 'publish or perish'[10], which seems to contribute to wasteful publication and unethical practices[11]. In light of certain potentially serious limitations associated with a system that creates a culture of accumulating points and impact factor scores, and at the same time rejects ratings applications on account of a lack of monodisciplinary focus, notwithstanding societal contribution, this paper seeks to strike a cautionary note, and to offer certain insights on the basis of the literature, which might be usefully incorporated into such a competitive system to reduce the harm it may cause.

Drawing from the relevant literature, this paper seeks also to make the argument that a system that rates academics through subjective rather than strictly objective evaluation might lack sufficient validity to be used to create perceptions as a differentiator of the quality of academics, based on their research. Similarly, given evidence that strong *cross-disciplinary differences* exist in terms of the relationship between objective criteria and the subjective NRF rating system's ratings,[12] research into ratings in the Management field is considered important, and perhaps timely.

Fedderke[12] found, for example, that, on average, 'C-rated scholars in the Biological Sciences have the same h-index as A-rated scholars in the Social Sciences' (p.3), and that ratings in the Business Sciences were the most difficult to attain for individuals with high h-indices, exceeded in difficulty only by those in the Medical and Biological Sciences.[12] Arguably, such attempts to prescribe a rating to an individual can suffer from a host of biases well considered in the scientific literature. This paper therefore seeks to identify certain potential biases associated with the application of the South African National Research Foundation's (NRF) rating system, and to link these potential biases to a discussion of the consequences of such a system, as well as to how these consequences accrue differently to different stakeholders, particularly societal stakeholders, who might be the most powerless in the face of a system that might not incentivise societal problem solving. These societal costs are expected to also result from decision criteria which subjectively deviate from relatively more objective measures of research performance.

Justification of the research

The arguments made in this paper are considered important for the following reasons. Firstly, the violation of central tenets of the academic process of gatekeeping itself might be considered in turn a violation of academic ethics, in that principles of anonymity and confidentiality of identity are not upheld in NRF rating assessments. This is perhaps especially problematic given the intensity of identity politics[13], and the racially oppressive history of the country associated with institutional racial discrimination on the part of the apartheid regime[14-16]. Given this historical context, to have the racial and gender identity of an individual known to assessors is perhaps unethical, given the historical context of the country, and given the career implications of rating. This is especially concerning if the objective evaluations of one's published work have *already been undertaken by expert peers* in the topic areas of journals, and therefore have already been vetted *under conditions of anonymity.*

Secondly, a similar violation of the principles of anonymity might relate to issues of academic freedom. The requirement for a 'coherent stream' of research has arguably been widely interpreted to suggest an applicant's research should fall into a 'silo', or into a largely monodisciplinary stream of research that does *not deviate in its focus*. Because an individual's entire portfolio of research is 'declared', any deviation from silo focus can be penalised. This is at odds with principles of academic freedom, for a number of reasons. Arguably, in doing so, the NRF rating system effectively shapes the growth of research to remain in silo areas, which might stunt important multidisciplinary or transdisciplinary innovations, as already stated above. This harks perhaps to Lysenkoism,[17] in that shaping research to grow in silos, or 'straight monodisciplinary' lines, might deny important changes in research trajectories, or might mitigate against important scientific advances in applied social sciences, particularly in socially important areas, particularly given that the 'second mode' of social science knowledge creation[8] is associated with applied interdisciplinary research that is necessarily defined by its problems (including those that are societally important). This might not be as big a problem for the natural sciences, as multidisciplinary work is arguably a characteristic of certain social research as a result of the multiple influences that can come to play in causing social conditions. Applied research in the social sciences, and in Management Science, can in many cases require transdisciplinary approaches, and for grant funding purposes, a multidisciplinary focus is often necessary. If Management Science researcher rating applications are rejected on account of a lack of a monodisciplinary focus, this issue should be the topic of further research and discussion.

Similarly, how scientific is it of a rating system to potentially penalise changes in a researcher's trajectory, away from a singular monodisciplinary focus, or even toward another? Arguably, denying a researcher a rating because of changes in trajectory (and hence a lack of a 'coherent' focus) could potentially count as harmful practice, as it can incentivise lack of innovation and constrain natural changes in the trajectory of an individual's research interests. Such systems might operationalise the exact problems identified by Kuhn[2].

Thirdly, another violation of the principles of academic freedom might be associated with the prescriptive nature of research 'authorities' in general. By not allowing subsidy for many good journals, yet officially including 'bad' journals in official lists ('white lists'), the stage is set for perverse incentives. It is common knowledge that journals that were identified as 'predatory' by Beall's List, were in the same year still fully accredited for subsidy by the South African Department of Higher Education and Training (DHET). However, Beall's List was discontinued at the start of 2017,[18] which has left academic staff, particularly those new to the system, at the mercy of official lists. The predatory journal phenomenon would be a non-issue if authorities implemented 'white lists' (lists of accredited journals) with the diligence required.

Indeed, who can forget the case of a journal that was fully accredited by DHET (and IBSS indexed) being de-accredited retrospectively, 2 years after South African academics had (perversely) accounted for a large share of its contents. One has to ask: has the NRF through its rating system not further reduced social science academic activity to that of a 'game'? Gamification of the system is hugely problematic if it results in the proliferation of ever-growing volumes of non-innovative research that is simply targeted at formulaic journal publication.[9] Have we created a monster? The test of this would perhaps be the extent to which research publication genuinely contributes to the benefit of societal stakeholders. If much of the research produced is not read by many, then what of the high levels of investment in the production of barely read research? If such a system incentivised innovative research or societally important research findings, it is possible that the system might be less wasteful. The NRF rating system, at least to the extent it relates to the rating of Management academics, might do well to take cognisance of these issues.

One may ask: who gets hurt in such games? Is it those established in publication, or is it the emerging cohorts of young academics who rely on the mentorship of those more established? Is the NRF rating system one which facilitates inclusion and development, or is its effect the opposite, acting as a mechanism of exclusion, or penalising innovative or societally oriented interdisciplinary research? Similarly, is this rating system acting as a catalyst to create a culture of competition which differentiates *publicly* between 'winners' and 'losers' in an academic game? If submission to such a system resulted in societal good, or was aligned with societally important needs, then tolerating the downsides

of such a system would be justified. If not, then further research and discussions into this topic are needed.

Perhaps it takes courage to speak truth to power, or to take a stand on issues that affect an academic's career progress within a powerful system in which many are invested. Nevertheless, such research is important if it leads to more transparent debate and scrutiny of a system that either directly or indirectly affects everyone in this country, either as academics or as societal stakeholders.

As indicated previously, given evidence that ratings outcomes are not consistent across different academic fields,[12] the objective of this paper is therefore to question certain of the assumptions that underpin the South African NRF researcher rating system, as it relates to the rating of Management researchers, in order to highlight instances in which principles of ethical and equitable assessment might not relate to practice. In doing so, certain suggestions for improved ethical use of such a system are made.

Context and background

The NRF is a South African state research funding agency that applies a peer-based evaluation system in rating researchers. The NRF's predecessor – the Foundation for Research Development – was established in the 1980s (see Pouris[19] for a useful history of the NRF and its origins). The mandate of the NRF is to 'promote and support research' through 'funding, human resource development and the provision of the necessary facilitates' in order to facilitate 'the creation of knowledge, innovation and development in all fields of science and technology, including indigenous knowledge' and thereby contribute to 'the improvement of the quality of life of all the peoples of the Republic'[20]. Its strategy is based on 'four core tenets': transformation, excellence, service culture and sustainability. Its mission statement includes the following corporate values: 'passion for excellence; world-class service; ethics and integrity; respect; people-centered; accountability'.

In terms of ratings, an individual is assessed, by peers, on their recent research outputs and impact as 'perceived by international peer reviewers'[20]. As the NRF rating methodology is based on qualitative, or subjective assessments, there should be no problem in developing an objective index of impact, according to Fedderke[12], based on either citations or on a formula that takes into account the impact factors of publications. Instead, what seems to happen is that an individual's research is subjectively assessed by a small group of evaluators, during which, for example, four reviewers can recommend rating, but two might object, resulting in the rejection of a rating. There seems to be a clear problem in that much variance exists in ratings – an issue expressed by Fedderke as follows[12]:

> *Given that such decisions can carry substantial bearing on the career prospects of scholars, it is not surprising that the decision making of funding bodies are viewed as being incompletely objective, and subject to the allegations of bias and inconsistency of various hues. Unfortunately, in general such claims are difficult to assess objectively against real data…This very substantial differential in research funding will invariably be critical in the career prospects of researchers. As such, that the peer review process is impartial, rewards true merit, and serves the development of scholarly activity is therefore of critical significance to researchers in South Africa…It is therefore not surprising that the review mechanisms of the NRF are often subject to controversy amongst the scholarly community. One of the sources of controversy arises from claims that the peer review mechanism is subject to bias at worst, or inconsistency across researchers and disciplines at best.* (p.1–2)

According to Fedderke, 'since the NRF undertakes its peer review under conditions of anonymity and in a closed review process, the claims of the NRF that its peer review process issues in reliable outcomes, is inherently not verifiable due to the lack of transparency of the process' whereby 'its peer review is not subject to the same public scrutiny that the publication process in peer reviewed scholarly journals faces once a scholarly contribution is accepted for publication'[12(p.2)]. Importantly, the individual being reviewed is not afforded the methodological protection of anonymity.

The individual applying for rating is not anonymised, unlike in the typical review process, so they are exposed to bias that was not experienced in the blind journal article review methodology. It is argued here that such issues cannot simply be discarded, as at best there will be inconsistencies between different fields and individual ratings. At worst, the inherent possible bias in these forums could be used maliciously to 'settle scores'. Such abuses are not necessarily occurring, but in the absence of comparative evidence of inconsistencies as a result of the lack of transparency in the process (the methodology), the potential for harm arguably exists. As scientists (natural or social), it is important to ensure the fairness of a rating system that allocates career (dis) advantages, particularly in a context that seems to prioritise research in promotion rather than teaching, notwithstanding the societal imperatives associated with massified higher education and the dramatic inequalities in access to opportunities in our society. NRF ratings are awarded according to the following categories[20]:

- A – Leading international researchers
- B – Internationally acclaimed researchers
- C – Established researchers
- P – Prestigious awards
- Y – Promising young researchers

In certain contexts, promotion to full professor is contingent on an NRF rating, which can skew promotion to favour those under 40 who are able to obtain a Y-rating under less stringent requirements than those seeking a C-rating. Thus, rating can have substantive career consequences and consideration of these consequences may warrant discussion, *particularly as one may not re-apply for rating for a period of time*. Thus, the methodology can intensify the potential harm from bias by barring an individual from applying again for a number of years. When methodologies are developed as part of research design, they typically need to pass muster with ethics committees, in order to ensure little harm comes to those subject to the research process. Are the same methodological design and ethical standards applied to applicants of the NRF ratings system? If they are, would it not be timely to initiate a debate on how the current methodology can be improved?

In order to understand these issues discussed here in terms of underlying regularities or deeper theoretical relationships, theory is now considered with a view to providing a more thorough perspective of the phenomena under discussion. Dominant in the discussions above is the notion that objective measures of scholarly impact do exist, and that subjective assessment can be harmful as a consequence of exposure to bias. Seminal literature might offer insights into these debates.

Theory and overarching principles

For Popper[1], subjective knowledge differs from objective knowledge. The former depends on our senses and the latter is associated at best only with tentative hypotheses. Subjective interpretations rely on probability theory as a result of incompleteness of knowledge, but at best remain beliefs, which can only be corroborated by degree. Objective interpretations, however, can be tested.[21] Hence we can only know what we can falsify, according to Popper's logics. From Popper's work, the alternatives to falsifiability are clear: failure to consider objective criteria can be associated with a cost. The cost of subjectivity in ratings applications falls not on the anonymous assessor but on the applicant, who must wait a number of years before applying again. This discussion brings to mind the problem of making a Type I or Type II error in significance testing, in which at least an objective measure can be used as a tool, but on the basis of objective probability evidence.

According to Popper, Einstein's theory that light would be influenced by gravitational forces is falsifiable, in contrast with Adler's psychological theory which seemed to fit 'all instances' of phenomena, which are 'compatible with the most divergent human behaviour, so that it was practically impossible to describe any human behaviour that might not be claimed to be a verification of these theories'[1(p.6)]. Burrell and Morgan[22] seminally differentiate between four paradigms relating to ontological and epistemological assumptions of social scientists along two axes, namely one related to objectivity versus subjectivity, and the other related to a 'sociology of radical change' versus a 'sociology of regulation'. This schema places work such as that by Marx in the objective/radical change quadrant, and what is typically regarded as natural science in the objective/status quo quadrant (one quadrant). The other quadrants include the radical humanist (subjective/radical change) and interpretive (subjective/status quo).[22] The point here is to acknowledge the seminal importance of subjectivity in social science research, and to relate it to the differentiation between subjective and objective assessments of phenomena – an example not least of which is the assessment of individual researchers in national rating schemes.

Arguably, Popper's[1] logic would therefore fall into but one of Burrell and Morgan's[22] incommensurate 'paradigms'. The other three paradigms, on the other hand, might accommodate Kuhn's[2] logic, where they are antithetical to Popper's.[1] In a nutshell, the purpose of revisiting seminal conceptions of the business of science is to simply show the contested terrain upon which we as academics ply our trade. To reduce this complexity and differences of views to a numerical rating scale justifies ongoing debate as to the validity of such exercises.

However, there is one aspect that, when discussed, academics typically are unanimous on. This aspect relates to the principle of objective fairness in matters relating to the employment relationship, and in the need for fairness in assessments of academic achievements which provide inputs to employment decisions of employers.

As explained, if there is widespread use of NRF ratings for employment and promotion purposes in South African institutions of higher education, even if not explicitly acknowledged, then it is perhaps our scientific imperative to ensure these are based on objective measures. Subjective measures might have been appropriate prior to the development of Internet-based measures of productivity and impact, but subjective measures are perhaps outdated in a context in which harm can come to applicants because of subjectivity bias.

The first issue relating to the validity of the system concerns the NRF rating process itself. A useful way to understand this is in its deviation from 'objective knowledge which is already known'. What is already known of a researcher applying for a rating? Using journal article publications as an example, there is usually clear evidence of the extent to which such an individual has been assessed by the academic community in general, in the form of published and accepted work that has been vetted by the academic gatekeeping system itself, in the form of reviewers and editors.

Journal reviewers are typically knowledgeable about the topic area of a journal article, and are accountable to journal editors. Undoubtedly, this system itself is far from perfect, but is arguably based on a systematic process. In the subjective assessments of NRF ratings applications, reviewers will only be knowledgeable of all the areas of research of the applicant in 'first mode' knowledge production.[8] If applied research has been undertaken, disciplinary divergence will necessarily occur, and a reviewer will only be able to attest expertly to a portion of the applicant's portfolio. The monodisciplinary approach might be useful for such a system to encourage.

In most instances of journal article quality control, however, double- or even triple-blind reviews are used, which if applied properly largely remove gender, racial or other forms of bias, even if they cannot remove bias associated with academic assumptions[22] and paradigmatic beliefs[2] identified by Burrell and Morgan and Kuhn, respectively. Thus the volume of an individual's work and its quality *has already been assessed* in a relative objective manner. The duplication of this process in an age of improving technological opportunities for objective measurement is costly, not only in terms of bias, but also in terms of time and resources.

In terms of the principles that guide the ratings process, the primary problem then, firstly, is the way the NRF rating system then takes this evidence, which is already in a quantifiable and objective form, and violates principles of anonymity, which this pre-existing body of evidence of publication history did not violate in its accumulation. Anecdotally, one hears constantly in corridors of researchers with prodigious volumes of research, including with articles published in some of the best journals in the world, who receive substantively lower ratings than others with fewer publications and publications in lower-ranked journals. One cannot but wonder as to the extent such inequity is the result of the violation of the anonymity principle which was not violated every time such researchers submitted their work for blind review. What makes the NRF rating process especially pernicious is perhaps that it occurs in a context in which discrimination was historically not only tolerated officially but was actually enshrined in apartheid laws and state institutions.[13] Change to democracy cannot be considered to have removed prejudice itself, even for academics. By removing the protection of anonymity, individuals are exposed not only to potential bias based on their race and gender, but also to bias that can be related to any aspect of their lives or lived identities – they are *known*.

Individuals have different personality endowments, different sexualities and different life choices; in a world of social media this information is but a click away from anyone with knowledge of a name and a person's basic information. This is not to say this bias is necessarily the result of the endemic inequities of the NRF ratings process, but on the level of principles, the possibility of bias in these ratings cannot be excluded. Moving from a body of work that has already been anonymously assessed by (a relatively larger number of) knowledgeable peers to a subjective and non-anonymous assessment by six or so reviewers is perhaps like moving (scientifically) from the properties of ratio data, to the adjudication of apples and oranges, or categorical data. This process is almost guaranteed to reproduce inequity in assessment of academic research because academic research is no longer the sole criteria for judgement.

And, importantly, what of the new emerging cohort of academics, who have arisen despite, or in the face of, the structural constraints of the country's past. *Can these scholars genuinely be guaranteed that the system will be as objective and fair* as the blinded reviews on which they have built up their portfolio of work? Who is most disproportionately affected by this system? Is it those entrenched in the system, who hold power over these new entrants? Is this system not institutionalising power, and the ability for those already established to exercise it at the expense of those seeking to enter the system? In the sections that follow, differential vulnerabilities to such a system are also considered in terms of power and its potential for harm.

Secondly, the NRF rating system works through reviewers chosen by the person being rated. This approach violates principles of objectivity in the most fundamental way. As previously argued, while different academic fields differ in their ontological and epistemological assumptions related to the tensions between subjectivity and objectivity, as well as between radical change and the maintenance of the status quo[22], there can be little disagreement that evaluations, and particularly *public* evaluations, of individuals, should be based on objective criteria. If we view the NRF rating system as a scientific methodology, and treat the harm it can cause in the same way we would for subjects in experiments, then ethical principles related to the use of a subjective methodology also need to be considered.

The ratings of an individual are clearly different in their validity from the work being assessed that had been through blinded peer-review. Arguably, those listing their lifelong friends as reviewers of their work can achieve extraordinarily high rankings. This is not to take anything away from those who do become highly ranked. These high rankings have largely been found to correlate with objective measures,[12] and it is typically those who have been unfairly rated on lower scales where the inconsistencies arise. After decades of work in an area of work, it

is likely that personal relationships will form. But to subject a ratings applicant to subjective ratings can prejudice those who are unfortunately not as adept at building personal relationships.

Thirdly, as has been mentioned previously, the NRF rating system as it stands might be vulnerable to abuse of power. Academic ranking is hierarchical, which is reflected in large status differentials between the elite and those not considered part of the elite. The academic context is one in which exclusivity is celebrated. The status differentials between the Ivy League and institutions of lesser stature are perhaps part of accepted everyday academic practice, as these differentials echo down the line, with almost every university and its academics 'ranked' tacitly against other institutions. Who can deny the differences in prestige associated with different institutions, and potential bias associated with assumptions of superiority or inferiority? When one foregoes the protection of anonymity in assessment of academic outputs, one also puts oneself at the mercy of the baggage that goes with one's institutional associations, and bias that can emerge from these associations. But central to the issue of power is the way in which power is a currency of sorts; and just like economic inequality, power has been used to exclude or marginalise those less powerful. In the words of Foucault[23]:

> *I would like to suggest another way to go further toward a new economy of power relations, a way which is more empirical, more directly related to our present situation, and which implies more relations between theory and practice. It consists of taking the forms of resistance against different forms of power as a starting point. To use another metaphor, it consists of using this resistance as a chemical catalyst so as to bring to light power relations, locate their position, and find out their point of application and the methods used.* (p.780)

The NRF rating system would be open to power abuses if it were inconsistent in its rating across different disciplines, or inconsistent across individuals. Arguably, the identification of inconsistencies should be reason enough to consider changing to a more objective system in the assessment of social science research for rating purposes.

The problem of power is particularly troubling if such systems can systematise power in such a way as to incentivise different types of research over others. The monodisciplinary approaches suited to the natural sciences differ from the multidisciplinary or even transdisciplinary approaches that are necessary when problem-solving or practitioner-focused research is undertaken,[8] representing mode one versus mode two research production, respectively.

Indeed, a monodisciplinary focus in applied research can be problematic, for instance in certain medical contexts.[24] In applied (second mode) social science research which focuses on solving societally important problems, it might be more difficult to demonstrate the singular, or rarified monodisciplinary focus that the NRF system seems to reward. In Management Science research, this might be a case of the misapplication of rating principles derived from a model better suited to the natural sciences rather than certain social sciences.

Indeed, in terms of objective measures, evidence suggests that inter-disciplinary research is not valued any less than monodisciplinary research according to qualitative (journal peer review based) or quantitative (bibliometric) measures.[25] This suggests that in academic publishing, rigour is not the exclusive domain of monodisciplinary research. Arguably, to address Kuhnian paradigmatic challenges,[2] interdisciplinary, multidisciplinary or transdisciplinary approaches might be particularly useful in social scientific contexts in which management practice is the focus of research. Even methodologically, enhanced discriminant and convergent validity of research[26] is increased through the use of different methods and frames of reference. Transdisciplinary approaches are particularly important for socially important research in the area of sustainability, which typically require new ways of knowledge production and decision-making,[27] and such approaches are perhaps particularly important in the context of a developing country like South Africa.

The way forward

As Fedderke[12] stresses, there are

> *a wide range of measures that measure both the number of scholarly contributions of scholars, and an even wider range of alternative measures that measure impact through citations and various derivatives of citation-based measures now available for consideration.* (p.2)

Although it is not possible to quantify the harm that has been caused by bias in the NRF rating system over time, Figure 1 illustrates the argument that the use of objective criteria for rating can reduce error variance, or bias in the rating system, notwithstanding the wider range of objective measures now available. In this figure, a bull's eye target heuristic is used to make the point that the more scientific the process, the less harm can come to individuals on account of subjective bias in the ratings evaluation process.

In an era when big data analytics can provide relatively comprehensive information on an ever-increasing range of phenomena, and when electronic measures can offer relatively comprehensive measures (and permutations of measures) of research productivity, these changes might promise not only a less harmful system (with less human cost), but also a system that can provide the cost and time benefits of an automated system. Hybrid (objective systems with a subjective component) are also considered to potentially be problematic. It is acknowledged that the design of an objective system would require qualitative, or subjective, engagement in order to weight its components, but if wide stakeholder agreement were obtained, once set, no subjective engagement would be required from ongoing assessments.

The arguments made in the above sections effectively reduce to one core issue – namely the need for *a more scientific methodology* in NRF rating assessments. Drawing from Fedderke's objective measures, alternative ratings methods are now considered, which might reduce subjectivity bias in ratings assessments. A ratings system should minimise bias as much as possible, but the shape and form of a ratings system will necessarily value certain types of research over others. The basis for a re-design of a system should therefore be widespread consultation across different societal stakeholder groups. Societal contribution should therefore be an anchor of such a system (amongst others). Evidence of prodigious output should not be penalised, as this is often the channel through which learning and development occurs. Worse, such an approach can incentivise perverse effects. It is not the role of such a system to hold back productivity, but to incentivise the societal contribution and impact of such productivity, as well as to ensure scientific progress, irrespective of whether it is in basic research or not. A more objective system might include any or all of the following measures, in any manner of permutation that captures what are clearly defined as measurement criteria:

(1) A measure of *raw output* is necessary, so that hard work is incentivised, as a path to researcher development, and innovative efforts are not penalised, as might occur under the present system.

(2) A *citation count* is also necessary, as it directly measures the impact of research, and could be divided by the number of years a researcher has been active, to mitigate the effect of accumulation of citations over time, which might not necessarily reflect improvements over time.

(3) The combination of output volume and its impact is captured in an *h-index*, which seeks to capture 'a robust single-number metric of an academic's impact', in that it 'corrects both for single high-impact publications, as well as for authors that publish a large number of uncited papers' whereby it validly favours those that publish 'a continuous stream of papers with lasting and above-average impact'[12].

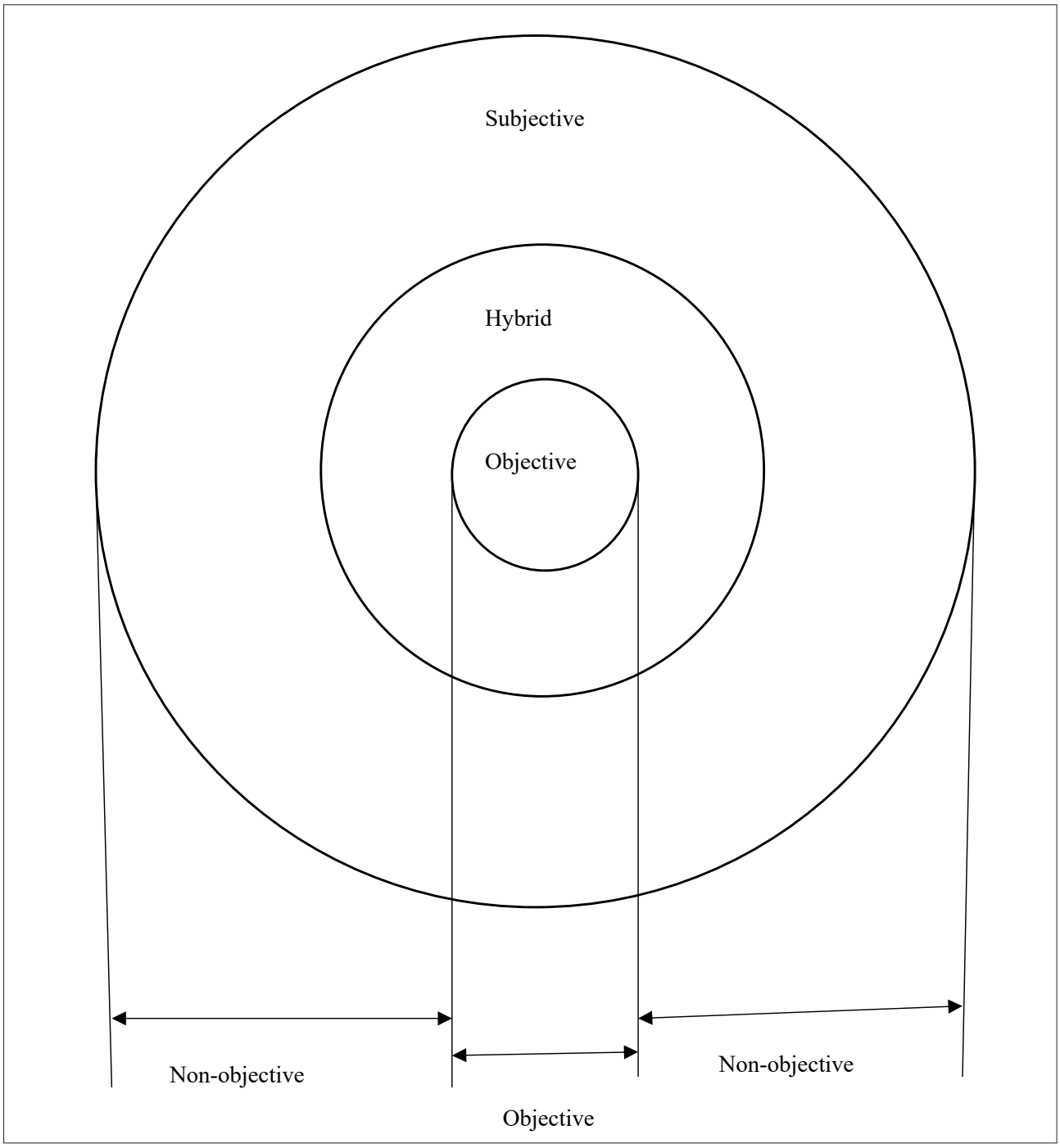


Figure 1: Error variance or 'potential for harm' of subjective versus objective NRF rating assessment.

The use of an h-index would arguably be more valid than the use of a system open to subjectivity bias. Someone with an h-index of seven would have at least seven papers with seven citations each, for example. Following Fedderke's suggestions, the h-index can be adjusted to better reflect current research, whereby it is adjusted for time since publication, whereby it is weighted according to the time since publication, or patterns in co-authorship, particularly in terms of the differences in co-authorship behaviours across fields, whereby citations for a paper might be divided by the number of co-authors (see Fedderke[12] for a useful summary of potential indices). Arguably, the use of these three methods can reduce the current subjectivity inherent in the ratings process. It was not the objective of this paper to explain the permutations of the h-index and similar measures, given that modifications to these objective measures have addressed previous critiques,[12] but to make the point that the dangers of subjective measures can be addressed by shifting the evaluations of researchers to electronic measures that are relatively more robust to the influence of subjective bias and inconsistencies.

(4) The current system might also be missing the point in evaluating researchers on an absolute basis, and should perhaps take recourse to a *relative approach*. New sites such as ResearchGate use a measure of a researcher's rating relative to all other included researchers, and this can perhaps provide an additional complementary objective measure against which rating can be conducted.

In all, a host of objective measures are now available for ratings systems to use, which arguably provide a holistic and relatively valid measurement process, particularly when taken together. The lack of innovativeness associated with the dominance of subjectivity identified

by Kuhn[2] is problematic, and it is perhaps time for the South African NRF system to move out of the past and into the present by changing to a technologically enabled system. Given the need to move away from past discrimination and inequality, technological systems would be unable to discriminate on spurious grounds unrelated to the impact of an individual's research, and would improve the validity and reliability of the evaluation process. Indeed, the elimination of potential discrimination in the South African context might warrant the use of an electronic system.

It is difficult to avoid continued reference to Fedderke[12], but this is considered necessary in that South African academics seem loathe to critique the NRF system as there are few who have challenged its conventions in the literature. So the final reference draws from Fedderke's[12] summary of the current limitations of the subjective NRF ratings process:

> *…it is based on subjective judgement; it favours a narrow disciplinary focus; it covers only a limited time span; it favours researchers in institutions with greater capacity; it does not even consider objective output and/or impact measures; the process is untransparent and unverifiable.* (p.19)

It is hoped that the practices of research rating that were developed in a time when the technology did not exist to record each and every indexed publication of an academic will change, and that the rating system will come into its own, as a valid and reliable system of evaluation. Until it does, it is argued that the costs of a subjective and inconsistent system will stay with us.

Conclusions and recommendations for further research

The objective of this paper was to highlight certain of the biases and inconsistencies associated with the South African researcher rating system, using as an example the field of Management Science, which is a social science. Given evidence of inconsistencies in ratings across disciplines,[12] it was argued that technological advances now offer a host of objective measures of both research productivity and its impact. It was argued that the current system is associated with certain validity and reliability issues, which are reflected in inconsistent ratings across individuals, with substantive career consequences for those who are rated unfairly, or who experience bias in ratings. It was also argued that the rating system as it stands violates certain principles of ethical research, namely that (1) it violates the principle of anonymity, in a context which has a history of possibly being among the most discriminatory in the world, (2) it violates the principle of objectivity, whereby subjective bias and inconsistencies have been shown to be present, and (3) it violates the principle of equality of power, as the academic context is one where exclusivity and Ivy League aspirations are associated with high power distance, and those with power have a mechanism to exercise it and to exclude. This exclusion is also on a public platform. Thus the subjectivity inherent in (2) arguably makes the power dynamics of (3) ethically untenable. An alternative system of evaluation was suggested, drawing on previous work that had suggested the same. It is concluded that, given that technological advances have made objective evaluations possible, the use of biased and inconsistent measures are not consistent with ethical practice. Indeed, if the NRF's mission statement[20] includes the values 'passion for excellence; world-class service; ethics and integrity; respect; people-centered; accountability', then these values would be expected to extend to the uptake of technological innovations in support of quality improvements in the research rating system itself.

References

1. Popper KR. Science: Conjectures and refutations: The growth of scientific knowledge. London: Routledge; 1963.
2. Kuhn TS. The structure of scientific revolutions. 2nd ed. Chicago, IL: University of Chicago Press; 1970.
3. Still A, Dryden W. The social psychology of "pseudoscience": A brief history. J Theor Soc Behav. 2004;34(3):265–290. https://doi.org/10.1111/j.0021-8308.2004.00248.x
4. Sokal A, Bricmont J. Fashionable nonsense. New York: Picador; 1998.
5. Boulding KE. General systems theory – the skeleton of science. Manage Sci. 1956;2(3):197–208. https://doi.org/10.1287/mnsc.2.3.197
6. Latour B. When things strike back: A possible contribution of 'science studies' to the social sciences. Brit J Sociol. 2000;51(1):107–123. https://doi.org/10.1080/000713100358453
7. Van den Besselaar P, Heimeriks G. Disciplinary, multidisciplinary, interdisciplinary – Concepts and indicators. Paper presented at: The 8th Conference on Scientometrics and Informetrics – ISSI2001; 2001 July 16–29; Sydney, Australia.
8. Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, Trow M. The new production of knowledge. The dynamics of science and research in contemporary societies. London: Sage; 2002.
9. Alvesson M, Gabriel Y. Beyond formulaic research: In praise of greater diversity in organizational research and publications. Acad Manag Learn Educ. 2013;12(2):245–263. https://doi.org/10.5465/amle.2012.0327
10. Richard JE, Plimmer G, Fam KS, Campbell C. Publishing successes of marketing academics: Antecedents and outcomes. Eur J Marketing. 2015;49(1/2):123–145. https://doi.org/10.1108/EJM-06-2013-0311
11. Rawat S, Meena S. Publish or perish: Where are we heading? J Res Med Sci. 2014;19(2):87–89.
12. Fedderke J. The objectivity of National Research Foundation peer review based ratings in South Africa. ERSA Working Paper 300 [document on the Internet] c2012 [cited 2017 May 15]. Available from: https://econrsa.org/system/files/publications/working_papers/wp300.pdf
13. Friedman S. South Africa: Electoral dominance, identity politics and democracy. In: Doorenspleet R, Nijzink L, editors. Party systems and democracy in Africa. London: Palgrave Macmillan; 2014. p. 47–68. https://doi.org/10.1057/9781137011718_3
14. Carrim N. Anti-racism and the 'new' South African educational order. Camb J Educ. 1998;28(3):301–320. https://doi.org/10.1080/0305764980280304
15. Habib A, Taylor R. South Africa: Anti-apartheid NGOs in transition. Voluntas. 1999;10(1):73–82. https://doi.org/10.1023/A:1021495821397
16. Crush J. The dark side of democracy: Migration, xenophobia and human rights in South Africa. Int Migr. 2001;38(6):103–133. https://doi.org/10.1111/1468-2435.00145
17. Hossfeld U, Olsson L. From the modern synthesis to Lysenkoism, and back? Science. 2002;297(5578):55–56. https://doi.org/10.1126/science.1068355
18. Strielkowski W. Predatory journals: Beall's List is missed. Nature. 2017;544(7651):416. https://doi.org/10.1038/544416b
19. Pouris A. The National Research Foundation's rating system: Why scientists let their ratings lapse. S Afr J Sci. 2007;103:439–441.
20. National Research Foundation [homepage on the Internet]. c2017 [cited 2017 May 15]. Available from: http://www.nrf.ac.za/
21. Popper KR. The propensity interpretation of probability. Brit J Philos Sci. 1959;10(37):25–42. https://doi.org/10.1093/bjps/X.37.25
22. Burrell G, Morgan G. Sociological paradigms and organisational analysis. London: Heinemann; 1979.
23. Foucault M. The subject and power. Crit Inquiry. 1982;8(4):777–795. https://doi.org/10.1086/448181
24. Verhey FR, Jolles J, Rudolf WHM, Rozendaal N, Plugge LA, De Vet RC, et al. A comparison between a monodisciplinary and a multidisciplinary approach. Neuroscience. 1993;5:78–85.
25. Rinia EJ, Van Leeuwen TN, Van Vuren HG, Van Raan AFJ. Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. Res Policy. 2001;30(3):357–361. https://doi.org/10.1016/S0048-7333(00)00082-2
26. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959;56:81–105.
27. Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, et al. Transdisciplinary research in sustainability science: Practice, principles, and challenges. Sustain Sci. 2012;7(1):25–43. https://doi.org/10.1007/s11625-011-0149-x