

**AUTHOR:**David A. Coldwell¹**AFFILIATION:**¹School of Economic and Business Sciences, University of the Witwatersrand, Johannesburg, South Africa**CORRESPONDENCE TO:**

David Coldwell

EMAIL:

david.coldwell@wits.ac.za

HOW TO CITE:Coldwell DA. A fair individualised university researcher rating system? A rejoinder to the current NRF debate. *S Afr J Sci.* 2019;115(3/4), Art. #5445, 3 pages. <https://doi.org/10.17159/sajs.2019/5445>**ARTICLE INCLUDES:**

- Peer review
- Supplementary material

KEYWORDS:

NRF rating; fairness; disability

PUBLISHED:

27 March 2019

A fair individualised university researcher rating system? A rejoinder to the current NRF debate

Given the impact of individualised research rating classifications on academic careers, further effort needs to be dedicated to finding systems and methods aimed at ensuring that rating evaluations are as fair as possible. However, the idea that this aim is possible through the application of a *more scientific approach* in the classification of individual academic researchers' 'pecking orders'¹ is questionable.

Boshoff's² 'rebuttal' points to some of the inaccuracies of Callaghan's¹ description of South Africa's National Research Foundation's (NRF) researcher evaluation process, but, in my view, he places too much faith in the *fairness* of the NRF's current researcher evaluation system and at times seems to rather uncritically ingratiate himself to the existing system despite its obvious faults. An issue that Boshoff² conflates, which is central to his argument against Callaghan's¹, is that of discrimination between different levels of academic performance. Boshoff² finds it unacceptable that Callaghan¹ regards discrimination between academics as elitist and unfair. However, the real issue is surely not whether it is unfair to discriminate between academic performance as such, but whether the NRF's discrimination in evaluating the excellence of researchers' outputs and the grade allotted to individual researchers is fair *sui generis*. I argue that it is not fair because it does not take account of the life circumstances of specific researchers which undermine their equality of opportunity to scarce research resources controlled by the NRF in its capacity as the sole official social institution in South Africa for this purpose.

Callaghan¹ suggests that the NRF rating system is not sufficiently scientific and that it makes, at least occasional, errors in rating categories allotted to individual scientists. And, in some cases, it is suggested these errors are gross and could be avoided by being more scientific. I argue that the realistic achievement of a formal scientific rating system is unattainable. However, with greater attention to qualitative aspects of the evaluative process, in particular, rather than greater objectivity, many of these gross errors of individual researcher prowess classification could be avoided. Of course in this approach, the assumption is that an individual form of researcher evaluation such as that used by the NRF in South Africa is the best form of assessment and the one most conducive to promoting more quality research across all disciplines.

From this point of view, South Africa currently operates an individualised form of research evaluation that is shunned by highly research productive countries in Europe. For example, in the UK, the Netherlands and Germany, a group-centred, departmental view to the quality and quantity of research output is adopted.³ This fact and the reasons for it, cast suspicion on the likely efficacy and durability of the South African system managed by the NRF whose viability has further recently become threatened through a drastic reduction in research funding for individual rated scientists (falling from a research subsidy of ZAR200 000 over 5 years to a once-off payment of ZAR30 000 over the same period for C-rated researchers). However, unlike Callaghan's¹ recent argument that the NRF is not scientific enough in its individualised rating assessments and should aim to be more so, I suggest that a truly objective and fair individualised rating system is not achievable and that rather than attempting to be more scientific in its evaluations, the NRF should rather be fairer in its assessments of a researcher by evaluating *more data* relating to individually specific, qualitative circumstances. Individual circumstances concerning ethnicity, gender and disability need to take central stage in individualised researcher grading if the grading is to be considered fair in a country such as South Africa with a long history of exclusion. The fact that it is not possible to consider in a scientifically rigorous way, the effects of ethnicity, gender and disability on research performance does not mean that consideration of such factors should be discounted altogether from NRF evaluations. In fact, and speaking generally, while race and gender have rightly attained central focus, and prescriptive remedial steps have been strongly implemented to try to address past discrimination, disability has become the 'Cinderella' of this thrust for redress and often is left out of equity deliberations and policies altogether. Although the proportion of disabled people in the South African general population (and therefore those who have been excluded by discriminatory practices) is comparatively smaller than those affected by past policies of race and gender exclusion, it does not mean that they should not also be dealt with fairly. In other words, disabled South African researchers must be given a fair opportunity to take up their rightful place by fairly assessing their contribution in relation to their particular circumstances and thus allowing a fair recognition of their talents to the overall research programme.

Of course, it should be stated at the outset that in any form of human assessment or judgement, errors are inevitable. It might appear to follow from this assertion then that the best way to avoid human subjective error is to use hard objective data, i.e. to become more objective and scientific as Callaghan¹ suggests, and thus eliminate as far as possible errors that arise from human subjective judgement. However, I argue that a truly scientific rating system worthy of that name is beyond reach and that objective assessments using hard, objectively verifiable data alone, would tend to *lessen* the fairness of such assessments, particularly in the context of South Africa where past injustices have left a legacy of inequality and injustice. It is my view that it is not that the NRF system is not scientific enough in its evaluations, as Callaghan¹ suggests, but that a more scientific approach to such evaluations is beyond reach, and that the adoption of such an approach, even if practically possible, would tend to *diminish* the validity of such evaluations (although it might increase their reliability), because crucial *qualitative* criteria for making fair assessments would summarily be ruled out of order.

Science and non-science demarcation and the NRF rating system

Callaghan¹ maintains that the current NRF system is not scientific. Perhaps the most generally accepted view of the demarcation between science and non-science is presented by the seminal work of Karl Popper in the philosophy of science. Popper's⁴ notion of demarcation between science and non-science is methodological, i.e. it is not

the subject matter itself that makes being scientific possible, but the methodological approach used. Thus, for Popper, sociology can be as scientific as physics. Being scientific is adopting a method which is, for Popper⁴, ‘falsificationism’; that is, scientific hypotheses must be both testable and open to falsification. Clearly, in this strict sense, evaluations in which expert reviews are used as a core resource for evaluation, such as those by the NRF, cannot be formally tested nor open to falsification. To that extent, NRF individualised researcher evaluations are not truly scientific. Such evaluations may be made more objective through the manner Callaghan¹ suggests, but by doing so, will lose a crucial criterion of fairness.

The review process

It is often said by NRF administrators that the process of grading depends very largely on the detail and content of the peer-review reports. Boshoff² gives a good outline of the NRF review process which he regards as both elaborate and sophisticated in its apparent thoroughness. However, I do not take this sophistication to mean, as Boshoff² apparently does, that the system is not prone to errors of evaluation because of the number of checks and balances in the evaluation process. Nor does the possibility of appeal against the outcome of the evaluation process seem to be one that bears much fruit for those who have taken this course of action to redress what they consider to be substantive errors in the outcome of such evaluations. For example, at the University of the Witwatersrand, very few appeals for reconsideration of grading were successful in 2017. Such an apparent mismatch between the number of appeals and successful outcomes thereof, casts doubt on the efficacy of the appeals process. The number of unsuccessful outcomes is surprising given that this University has expert knowledge and considerable experience of what is required for specific researcher grading categories, and that appeals are vetted for their reasonableness and cogency by the University before they are lodged with the NRF for consideration. On the face of it, this situation suggests that either the University is operating with a ‘pie in the sky’ notion of the NRF classification system, and one which seeks to maximise the chances of obtaining the highest number of research grades in the highest possible categories (which seems highly unlikely), or there are fundamental misunderstandings between the two bodies. Whatever the reason, there can be little doubt that the general tendency for appeals to be rejected evokes negative reactions in the researchers concerned and is generally highly counterproductive to the fundamental research enterprise.

As Callaghan¹ points out, the current review process is not blind. Not only do evaluation candidates indicate their reviewer choices, they are able also to request exclusion of reviewers who they feel might be most harshly opposed to them. This makes the evaluation process open to subjective assessment at best, and vulnerable to unbridled bias at worst.

I do not subscribe to the view that large sophisticated institutions simply because of their size and sophistication, or indeed the apparent ‘good’ for which they aim, cannot be wholly misguided.² History is littered with the failure of institutions precisely of that type, which may have been initiated with the intent to do good, yet later were found to have created at least as many problems as they solved. (The British National Health system might, arguably, serve as but one modern example.)

However, Rawl’s⁵ theory of justice which regards justice as *fairness* in social institutions, presents two basic priority rules which are worth considering in this context. The first rule states that principles of freedom – such as freedom of speech, association and religion – can only be restricted when doing so results in other freedoms. The second rule, which has much greater relevance to the argument here, is that justice or fairness is always more important than the efficacy or utility of outcomes, and that, in particular, equal opportunity is more important to justice and fairness than utility. As a social institution providing benefits and responsibilities of scarce resources, which in this case is research funding money to academic researchers, the NRF must ensure that fair opportunities exist for all researchers to fully enjoy these benefits through a fairly managed system of access to extant opportunities.

Lack of specificity and definition of graded categories

I have operated several times as a reviewer for the NRF and I have been amazed by the lack of specific guidelines on how to proceed with my review. For example, the grading system is aimed at categorising a particular individual into one of several possible grades:

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

However, even as an established researcher with some knowledge of the NRF grading system, it becomes difficult to make an assessment of individual candidates’ applications in terms of in which grades they might reasonably fit, because there are no clear guidelines in extant evaluator documents indicating how to do so. The reviewer is left to describe and assess an applicant’s research without having any formal parameters to make specific categorisations. In effect, the reviewer is tasked with the job of ‘measuring’ an applicant’s research prowess without having a common yardstick with which to perform this difficult task.

Weasel words in NRF grading

Merriam-Webster Dictionary⁶ defines a weasel word as ‘a word used in order to evade or retreat from a direct or forthright statement or position’. The American Heritage Dictionary⁷ defines it as ‘an equivocal word used to deprive a statement of its force or to evade a direct commitment’. As far as I am aware, the use of equivocal words in the NRF grading procedure is not a deliberate policy to make their validity in comparisons of individual rating outcomes difficult to judge, but it is, nevertheless, a direct consequence. In this regard, for example, the meaning and interpretation of the word ‘acclaim’ is crucial in that it affords the springboard to promotion from the NRF’s lowest grade grouping.

The classification into separate categories of research excellence by the NRF is relatively clear for Grades A and C. An A-rated researcher is defined as ‘a leading international researcher’, which is fairly clear in the sense of what it purports to cover, although *leading* in international research may refer to different spheres of influence and therefore impact. For example, a leading international scholar of South African English literature will have a smaller critical audience and rivalry for the accolade of ‘leading’, and significantly less international impact and recognition for their research, than a research scientist working on a cure for HIV or malaria. This would suggest that a leading researcher in the field of medical science will be required to reach a quite different level of research excellence and to find a space as a leader in a much larger competing group of potentially leading international researchers than their colleagues in the humanities. In this regard, for example, Fedderke⁸ found that biological C-rated researchers had on average the same h-index as A-rated researchers in the social sciences.

The NRF C-rating for established researchers is, similarly, a reasonably clearly defined and easily operationalised category, although again the requirements to become an ‘established researcher’ are quite different in stringency demands between disciplines in terms of what is regarded as minimum acceptable performance.

However, the concept of ‘international acclaim’ is largely a subjective one and one which cannot be effectively tested in the Popperian⁴ sense, because the concept itself cannot be effectively operationally defined, and the meaning and objective measurement of, for example, the ‘relationship’ between what is meant by ‘acclaim’ and what may be considered internationally recognised scientific contributions to knowledge may not be one and the same thing. The Cambridge Dictionary⁹ defines acclaim as ‘public approval and praise’, but in the context of scientific research can one reasonably expect all valuable and internationally recognised research to be ‘acclaimed’. Lakatos¹⁰ idea of a scientific research programme is that it is not scientific for all time



but slips over time from being 'progressive' to becoming degenerative¹¹. Lakatos¹⁰ talks of a scientific 'hard core', of theory and methodology which is treated as irrefutable by its promoters with a protective belt of auxiliary hypotheses that have still to be fully tested. The protective belt acts as a defensive mechanism, thereby allowing a period of 'normal science' before a paradigm shift occurs through a revolutionary change in scientific perspective.¹² These ideas are particularly problematic in this regard because research that is publicly approved and praised today, may not be so tomorrow. But more importantly for the work of a researcher who is undergoing evaluation, there are, in the social sciences, multiple methodologies and theories co-existing and while one group of the research audience may regard a researcher's work as attaining the 'holy grail' of theoretical and methodological perfection, another group in the same audience may regard it in much the same way as a natural scientist today would regard astrology or alchemy. It is true to say, however, that the NRF has recently tried to put the 'concept' of 'international acclaim' on a more intelligible footing, by offering further criteria for its identification and including an addendum to the existing definition that states that despite this change, it may 'not be considered exhaustive'. This change is tantamount to an admission that the word 'acclaim' is next to impossible to define clearly in any scientifically operational way that would allow objective measurement. In short, therefore, any attempt at making an evaluation more scientific in a rigorously testable way is likely to fail; and to generate as many Type 1 as Type 2 errors.

Of course, this is not to say that objective criteria of 'international acclaim', such as the number of invited plenary presentations or the number of international citations in quality journals, should not be considered – clearly they should, but holistic qualitative data regarding a particular scientist's specific circumstances also need to be carefully considered. For example, is it fair that a scientist who suffers from a sensory disability (e.g. blindness or deafness) should be judged in the same manner as other scientists in relation to key criteria for a NRF rating? Criteria such as the size of their scholarly network and reputation (obtained significantly through interaction and communication with colleagues at local and overseas conferences) and the number of invited plenary sessions hold the same weight for applicants whether or not they are disabled.

Conclusion

While agreeing with Callaghan¹ that there are flaws in the NRF's current grading of individual researchers, some of which may be endemic to a system that tries to distinguish between the excellence of research output between individual researchers, I disagree that the solution is for the institution to become more objective (or 'scientific') in its evaluations. My concern is that if it were to do so, it would not apportion fairness in its assessments and ignore the particular circumstances of, for example, researchers with disabilities, thus lessening their equality of opportunity and access to scarce research funding resources.

It needs to be said that the current debate has been ongoing in the literature for some time. Campbell¹³ for example refers to it as the binary option of university research evaluation; the binary aspects being:

- *peer review* consisting of evaluations based on subjective expert opinion which broadly corresponds to the current NRF system, and
- *indicators* consisting of judgements of research or researcher excellence based on objective quantitative data.

It is noted that while peer-review evaluations allow a higher degree of complexity in assessments, they have a strong dependence on the composition of the panels which can create personal, methodological and theoretical bias. Indicators can be more objective but tend to be superficial. Whether a fair balance can be found in NRF-type individual researcher evaluations, to fairly assess equality of opportunities of particular researchers presented with specific life realities, remains doubtful and is a significant factor motivating the use of the more generally applied *group-centred* researcher evaluation systems in most other countries today.

References

1. Callaghan C. A review of South Africa's National Research Foundation's rating methodology from a social science perspective. *S Afr J Sci.* 2018;114(3/4), Art. #2017-0344, 7 pages. <https://doi.org/10.17159/sajs.2018/20170344>
2. Boshoff C. Chris Callaghan's criticism of the National Research Foundation's rating methodology: A rebuttal. *S Afr J Sci.* 2018;114(7/8), Art. #a0278, 4 pages. <https://doi.org/10.17159/sajs.2018/a0278>
3. Barker K. The UK Research Assessment Exercise: The evolution of a national research evaluation system. *Res Eval.* 2007;16(1):3–12. <https://doi.org/10.3152/095820207X190674>
4. Popper KR. *Science: Conjectures and refutations: The growth of scientific knowledge.* London: Routledge; 1963.
5. Rawls J. *A theory of justice.* Cambridge, MA: Harvard University Press; 1971.
6. Merriam-Webster English Dictionary [online]. Weasel [cited 2018 Jul 27]. Available from: <https://www.merriam-webster.com/dictionary/weasel%20word>
7. *American Heritage Dictionary.* 5th ed. Boston, MA: Houghton Mifflin Harcourt Publishing Company; 2016.
8. 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
9. The Cambridge English Dictionary [online]. Acclaim [cited 2018 Jul 27]. Available from: <https://dictionary.cambridge.org/dictionary/english/acclaim>
10. Lakatos I. *The methodology of scientific research programmes: Philosophical papers volume 1.* Worrall J, Currie G, editors. Cambridge: Cambridge University Press; 1978.
11. Blaug M. *The methodology of economics.* Cambridge: Cambridge University Press; 1980.
12. Kuhn TS. *The structure of scientific revolutions.* 2nd ed. Chicago, IL: University of Chicago Press; 1970.
13. Campbell DFJ. *Conceptual framework for the evaluation of university research in Europe: Working paper.* Washington DC: Center for International Science and Technology Policy (CISTP), George Washington University; 2002.