



AUTHOR:
Steven D. Johnson¹ 

AFFILIATION:
¹Centre for Functional Biodiversity,
School of Life Sciences, University
of KwaZulu-Natal, Pietermaritzburg,
South Africa

CORRESPONDENCE TO:
Steven Johnson

EMAIL:
johnsonSD@ukzn.ac.za

DATES:
Received: 27 July 2020
Revised: 07 Aug. 2020
Accepted: 07 Aug. 2020
Published: 21 Sep. 2020

HOW TO CITE:
Johnson SD. Peer review versus
the *h*-index for evaluation of
individual researchers in the
biological sciences. *S Afr J
Sci.* 2020;116(9/10), Art.
#8700, 5 pages. [https://doi.
org/10.17159/sajs.2020/8700](https://doi.org/10.17159/sajs.2020/8700)

ARTICLE INCLUDES:
 Peer review
 [Supplementary material](#)

DATA AVAILABILITY:
 Open data set
 All data included
 On request from author(s)
 Not available
 Not applicable

EDITOR:
Jenni Case 

KEYWORDS:
bibliometrics, Hirsch's *h*-index,
National Research Foundation, NRF
rating, research administration

FUNDING:
South African National Research
Foundation (grant 46372)

Peer review versus the *h*-index for evaluation of individual researchers in the biological sciences

Past performance is a key consideration when rationalising the allocation of grants and other opportunities to individual researchers. The National Research Foundation of South Africa (NRF) has long used a highly structured system of 'rating' the past performance of individual researchers. This system relies heavily on peer review, and has seldom been benchmarked against bibliometric measures of research performance such as Hirsch's *h*-index. Here I use data for about 600 rated researchers in the biological sciences to evaluate the extent to which outcomes of peer review correspond to bibliometric measures of research performance. The analysis revealed that values of the *h*-index based on the Scopus database are typically 5–20 for researchers placed in the NRF's C rating category ('established'), 20–40 for those in the B rating category ('considerable international recognition') and >40 for those in the A rating category ('leading international scholars'). Despite concerns that citation patterns differ among disciplines, the mean *h*-index per rating category was remarkably consistent across five different disciplines in the biological sciences, namely animal sciences, plant sciences, ecology, microbiology and biochemistry/genetics. This observation suggests that the NRF rating system is equitable in the sense that the outcomes of peer review are generally consistent with bibliometric measures of research performance across different disciplines in the biological sciences. However, the study did reveal some notable discrepancies which could reflect either bias in the peer-review process or shortcomings in the bibliometric measures, or both.

Significance:

- NRF rating categorisations (estimates of standing in a research field based on peer review) are a reasonably good predictor of the *h*-index of individual researchers in the biological sciences.
- The relationships between rating categorisations and the *h*-index are remarkably consistent across five sub-disciplines in the biological sciences.
- Peer review and the *h*-index were compared in terms of their relative advantages and disadvantages and the combined use of both approaches is advocated for measurement of research performance.

Introduction

Peer review is one of the pillars of the global research enterprise.¹ At a political level, national governments often steer overall research activities in particular directions through funding allocations to programmes, but they also almost always devolve the final allocation of resources to a system of peer review. The logic is that peers are considered to be in the best position to evaluate the quality of past and proposed research, even though it is also acknowledged that reviewers can exhibit bias, which can be either explicit, such as overt competition among researchers, or implicit, such as underlying prejudice according to race, gender and the perceived status of the institution to which a researcher is affiliated.¹ The peer-review system is under enormous strain worldwide and grant administrators increasingly struggle to obtain quality reviews of funding proposals.² Peer review, whether of grants or submitted manuscripts, is usually performed without added remuneration and there are thus limits to the time that researchers are willing to allocate to this process, particularly if it involves an assessment of the entire track record of each individual researcher.

One possible way of reducing the burden on peer reviewers (and the institutions that manage the process) is to separate periodic evaluation of the track record of the applicant from the evaluation of proposed research, such that a single measure of the applicant's track record can be used for multiple decision-making processes. This approach has been adopted in South Africa for several decades and takes the form of peer evaluation of the research performance of researchers in the higher education and research sector. This evaluation is usually based on the opinions of about six peers (some suggested by the candidate and some suggested by a panel of discipline-specific experts) and is codified as a particular 'rating' which lasts for a period of 6 years. This information can then be used by administrative panels when making grant allocations or for other purposes such as informing a university selection committee about the past performance and general standing of an applicant in a research field.

The availability of massive computerised databases of publications that include citation information has led to the development of bibliometric measures of research performance. By far the most widely adopted of these is Hirsch's *h*-index, which is simply the number *h* of papers that have been cited at least *h* times.³ The *h*-index is intended to strike a balance between the total number of citations, which may be unduly influenced by a few very well cited papers, and the total number of publications, which may not reflect the actual impact of the research in terms of citations.⁴ The *h*-index has been shown to be closely associated with measures of academic standing in a field.^{5,6} Potential drawbacks of the *h*-index include its insensitivity to number of authors, author positions, and discipline-specific citation patterns.^{4,7–11} Like the metric of total citations, the *h*-index also shows a ratchet effect whereby it will continue to increase even after a researcher has become inactive.⁴

There have been very few attempts worldwide to determine the extent of agreement between peer review and bibliometric measures. Most of these involve studies of whether the *h*-index can predict the outcomes of applications for fellowships^{12,13} or future career trajectories¹⁴. Hirsch⁹ has characterised the *h*-index as 'an indicator

of the impact of a researcher on the development of his or her scientific field'. This is uncannily similar to the objectives of the National Research Foundation (NRF) rating system in South Africa. Studies of correlations between peer assessment of research standing, such as NRF ratings, and the *h*-index are particularly valuable, but remain rare.^{5,8,15,16} Lovegrove and Johnson¹⁷ analysed data for a small sample (163) of botanists and zoologists in South Africa and found that the outcome of peer review of research standing in the form of NRF rating categories was fairly well correlated with the ISI-based *h*-index. They did not attempt to analyse trends across other disciplines in biology, however, and the ranges and means they obtained for the *h*-index in each rating category need to be updated, given the large increase in publications worldwide over the past 12 years.

The aim of the present study was to establish, across different disciplines of biology, the relations between estimates of the standing of researchers in their field based on peer review and those based on bibliometric analysis of their *h*-index.

Methods

I used the public database of NRF ratings assigned to 4176 South African researchers, available at www.nrf.ac.za, which was last updated on 30 June 2020. I filtered this database down to 644 researchers who mentioned 'Biological Sciences' as one of their primary disciplines (Supplementary table 1). I was able to find the Scopus-based *h*-index values for 614 of these researchers (searches took place 18–19 July 2020). Values of the Scopus-based *h*-index are generally very similar to those based on the Web of Science.¹⁸ I did not use Google Scholar because Google Scholar profiles that are not frequently curated often include papers that are not authored by the researcher, thus inflating their actual *h*-index.

I was able to assign 569 researchers to a sub-discipline, usually based on their own statement of a sub-discipline in their NRF profile, or, more rarely, by examining the content of papers in their Scopus profile or consulting with other researchers in the sub-discipline. I used the following sub-discipline categories (number of researchers): 'Ecology' (137), 'Plant Sciences' (75), 'Animal Sciences' (127), 'Microbiology' (115) and 'Biochemistry, Molecular Biology and Genetics' (115). Palaeontologists were allocated to the Animal Sciences category only if their main focus is on animal structures and evolution rather than geology. These disciplines are based loosely on those used by Scopus and do not correspond exactly to the specialist committees used by NRF to oversee the results of peer review. For example, I have grouped all ecologists into one category, regardless of whether they would fall under the plant sciences or animal sciences committees.

I analysed values of the *h*-index of researchers to determine the degree to which these could be predicted by peer-review outcomes (the rating categories in which these researchers had been placed) as well as according to the specific discipline of the researchers. The three rating categories used by the NRF that were analysed (and the number of researchers in these) were A ('leading international scholars', $n=27$), B ('scholars with considerable international recognition', $n=129$) and C ('established scholars', $n=414$). I did not analyse rating sub-categories, e.g. C1, C2 and C3, as this information is not made public by the NRF. Even if the rating sub-category data could be obtained on the condition of confidentiality, as was the case in the analysis by Lovegrove and Johnson¹⁷, analysing them here without making the data available as an appendix would violate the principle of data transparency. I did not analyse researchers who applied for a rating and were unsuccessful as this information is not publicly available; nor did I analyse the Y and P rating categories as these specifically apply to early-career researchers who are evaluated largely according to future potential and not on their past track record in terms of impact within their fields.

Values of the *h*-index were skewed with a longer tail to the right and were thus analysed using a generalised linear model which incorporated a gamma distribution and log link function (implemented in SPSS 25, IBM Corp.). Rating category, sub-discipline and their interaction were treated as fixed predictors, and time (0–5 years) since the rating was awarded was treated as a covariate. Model significance was assessed using

likelihood ratios and post-hoc comparisons among means were based on the Dunn–Sidak method. In statistical terms, this model tests the null hypothesis that different rating category allocations are drawn from the same distribution of the *h*-index. An equally valid approach not used here would be to use a logistic model to assess whether the *h*-index is a good predictor of the probability of different rating categorisations.⁸

Results

The distributions of raw (non-adjusted) *h*-indices according to rating category revealed clear clustering, but with some overlap of values between categories (Figure 1). The *h*-index values for researchers differed significantly according to rating category ($\chi^2 = 399.9, p < 0.0001$), sub-discipline ($\chi^2 = 14.2, p = 0.006$) and time since rating ($\chi^2 = 399.9, p < 0.0001$), but not by the interaction between rating category and discipline ($\chi^2 = 10.4, p = 0.234$).

The overall model-adjusted means (\pm s.e.) for the *h*-index were 16.4 ± 0.29 for researchers in the C category, 29.8 ± 0.96 for researchers in the B category and 57.4 ± 4.16 for researchers in the A category (all category means differed significantly, post-hoc tests $p < 0.0001$). The ranges in mean *h*-index across disciplines were 15.4–18.0 for the C category, 25.8–31.8 for the B category and 38.0–81.5 for the A category (Table 1).

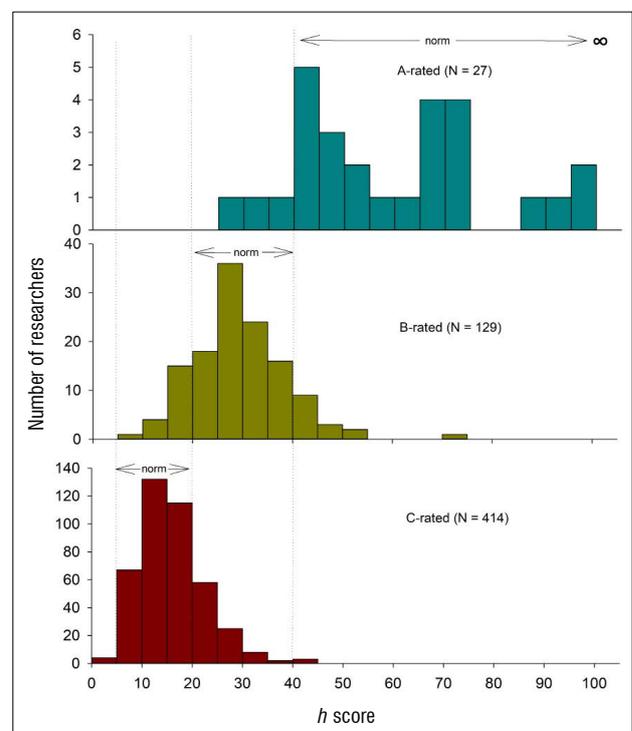


Figure 1: The distribution of values of Hirsch's *h*-index for researchers in the biological sciences placed in the three main rating categories used by the National Research Foundation of South Africa: A = 'leading international scholars', B = 'scholars with considerable international recognition' and C = 'established scholars'.

Discussion

These results indicate that there is substantial agreement between the evaluation of research performance based on peer review and values of the *h*-index based on bibliometrics across a wide range of sub-disciplines in biology (Table 1). The consistency across disciplines is particularly evident for the B and C rating categories (Table 1). The greatest variation in the mean *h*-index across disciplines, notably a twofold difference between animal sciences and molecular biology, is found in the A category (Table 1), but as this category has no upper bound and includes a relatively small number of researchers, the mean value for a sub-discipline can be shifted upwards considerably by one or two researchers with particularly high values of the *h*-index.

Finding that the *h*-index is largely congruent with NRF ratings does not automatically imply that measures of research performance based on peer review can simply be supplanted by bibliometrics. The reason for this caution is that both approaches have disadvantages, and a combination of both may offer the best safeguard against unfair evaluation of an individual. Given that the advantages and disadvantages of peer review, partially summarised in Table 2, are well known^{1,15}, I focus here on the advantages and risks of relying on the *h*-index as a measure of research performance.

The purpose of measuring research performance is usually to make decisions about allocations of public funds in the form of grants. It is therefore important that the measure reflects overall research competence and ability. In this sense, the *h*-index has some serious drawbacks, such as favouring authors who publish as part of large teams, including consortia.^{7,19} For example, an author who is a minor (middle) author of a large number of multi-authored papers may quickly develop a healthy *h*-index and yet may not have the requisite experience in managing research or writing papers.^{9,20} This issue would probably be flagged by peer reviewers. There is an increasing trend for papers to have large numbers of authors. The charitable view is that this reflects a genuine increase in collaboration among researchers, but a more cynical view is that it reflects a form of collaborative gaming of the system by authors who wish to collectively increase their numbers of publications and citations.²¹ A curious footnote to this issue is that in South Africa, government incentive funding for each publication is divided among authors, thus providing a perverse incentive (counter to that applied by the *h*-index) for researchers to minimise the number of co-authors. Another form of gaming by authors is to focus on writing review articles, rather than to conduct original laboratory or field-based research, simply because review articles are well known to garner more citations.²¹ The need to more fairly reward authors for original research has recently been recognised in the ‘San Francisco Declaration on Research Assessment’ (<https://sfidora.org/>) which includes an advisory for authors to cite original research rather than reviews wherever it is possible to do so.

The *h*-index ratchets upwards throughout a person’s career and even continues to increase after they have become research-inactive. This is potentially a serious problem given that the purpose of performance evaluation is usually to allocate resources to individuals who are

currently active. Reviewers for the NRF rating system in South Africa, for example, are expected to focus on the quality and impact of the research performed in the 8 years preceding the evaluation, and the overall *h*-index can be misleading in this regard. Only by focusing on the year-by-year trends in citations and publication quantity and quality over that period can a panel gain insight into the recent impact of a person in their field. Of course, the situation is completely different if the purpose of using the *h*-index is to award a prize for career impact to someone who has retired or who is approaching retirement, as opposed to using the information for allocation of resources. It also seems obvious that the *h*-index has limited utility for evaluating early-career researchers, although there have been some studies that have shown that early-career values of the *h*-index (medians around 2–3) do correlate to some extent with the peer-review outcomes of fellowship applications.¹³ Personally I would be sceptical of relying too heavily on the *h*-index to evaluate early-career researchers for purposes such as awarding postdoctoral grants as it may favour applicants in a manner that is directly proportional to the number of years since completing (or commencing) their PhD, even if those intervening years were not particularly productive.

Given these drawbacks of the *h*-index, should it be used in performance evaluation at all? The answer, I believe, is an emphatic ‘yes’. If the evaluation of a researcher by a peer-review process leads to an outcome that is incongruent with their *h*-index, then a panel needs to consider the basis for the incongruence. For example, a person with an *h*-index of 10 whose peers write reports that place them in the B category (‘considerable international recognition’) may have benefitted from reviewers who had been primed to provide favourable reports, and the case should be reconsidered. It would be very difficult to detect this misfeasance without the use of a benchmark provided by the *h*-index. However, it can also be the case that peer review is sometimes the more reliable measure. For example, a person whose peers write reports suggestive of a B rating may have an *h*-index of 60. Further investigation by a panel may reveal that such a person is no longer fully research-active, and the final decision may be more in accordance with the peer review than the *h*-index. It is even possible that the peer reviewers themselves make use of the *h*-index when asked to evaluate a candidate for rating, especially one with whom they are not very familiar.²¹ Thus it could be the case that measures of research standing in a field based on peer review, such as NRF ratings, are, in fact, already being informed by the *h*-index. This lack

Table 1: Model-adjusted (marginal) mean *h*-index values for sub-disciplines within each NRF rating category

Rating	Discipline	Researchers	Mean	s.e.	95% Confidence interval	
					Lower	Upper
A	Animal Sciences	4	38.0	6.82	26.69	53.98
	Biochemistry, Molecular Biology, Genetics	4	81.6	14.65	57.32	115.95
	Ecology	6	64.5	9.47	48.40	86.03
	Microbiology	9	59.7	7.15	47.18	75.46
	Plant Sciences	4	52.5	9.48	36.83	74.77
B	Animal Sciences	31	25.8	1.66	22.69	29.24
	Biochemistry, Molecular Biology, Genetics	24	29.3	2.16	25.41	33.89
	Ecology	34	30.6	1.89	27.15	34.59
	Microbiology	21	31.8	2.50	27.22	37.07
	Plant Sciences	19	31.7	2.62	27.00	37.31
C	Animal Sciences	92	15.4	0.58	14.26	16.52
	Biochemistry, Molecular Biology, Genetics	87	16.4	0.63	15.19	17.67
	Ecology	98	18.0	0.66	16.79	19.38
	Microbiology	85	16.9	0.66	15.62	18.21
	Plant Sciences	52	15.7	0.78	14.20	17.27

Table 2: A comparison of peer-review and the *h*-index in terms of advantages and disadvantages

Peer review		<i>h</i> -index	
Advantages	Disadvantages	Advantages	Disadvantages
<ul style="list-style-type: none"> • Can evaluate recent research performance • Can account for complexities such as author position and the contribution of an individual to a multi-authored paper 	<ul style="list-style-type: none"> • Reviewer bias • Reviewer fatigue • Administratively complex • Requires textual interpretation by panel members 	<ul style="list-style-type: none"> • Not sensitive to citations of single papers • Considers impact of papers, not journal impact factors • Inexpensive to administer 	<ul style="list-style-type: none"> • Does not decrease when research activity slows (ratchet effect) • Does not take number of co-authors into account • Does not take author positions into account • Not suitable for evaluating early-career researchers • Can be 'gamed' by self-citation of papers just below <i>h</i>-index threshold • Relatively insensitive to massively cited 'big hit' papers (can also be interpreted as an advantage) • May vary among disciplines

of clear separation between independent and dependent variables means that any statistical test of associations between the *h*-index and ratings of research standing, such as those in the present study, should be viewed as no more than crude heuristic approximations. This also means that any characterisation of peer review as 'subjective' and the *h*-index as 'objective' would be overly simplistic.²² It could even be argued, quite reasonably, that the *h*-index itself is already a form of peer review as most citations are essentially a form of validation by peers – although it is of course possible for papers to be cited as examples of faulty research.

An unexpected result of this analysis was the huge inflation (roughly doubling) of *h*-index values within each rating category in the biological sciences since the previous analyses by Lovegrove and Johnson¹⁷ and Fedderke⁹. For example, the lower *h*-index threshold for a B rating in the 2008 study was about 10 and is now about 20, and for an A rating this lower threshold has shifted from about 20 to 40 (Figure 1). It is not clear why this is the case, but one possibility is that the yearly increase in publications is outstripping the number of new researchers in the global system, thus leading to increasing values of the *h*-index. Another (related) possibility is that the number of authors per paper is increasing, leading to more papers for the same researcher effort. The most cynical and depressing explanation is that there is limited turnover of new researchers, particularly in the A and B categories, and so the NRF system is simply repeatedly re-evaluating much the same cohort of people whose *h*-index is increasing with time. Whatever the reason, it is clear that the *h*-index norms for each rating category are moving targets and, for administrative purposes, may need to be adjusted upwards every few years.

As fitted by eye, the *h*-index norms in the data for researchers in the biological sciences appear to be about 5–20 for the C category, 20–40 for the B category and >40 for the A category (Figure 1). These norms encompass 75.8% of researchers in the C category, 69.7% of researchers in the B category and 88.8% of researchers in the A category. In this analysis, there were 142 researchers (c. 25% of the total) whose ratings lie outside these norms and some of these represent marked outliers (Figure 1). These outliers could reflect failures of the peer-review system, such as review reports that are negatively biased or primed to be favourable when they should not be, or they could represent special cases fairly deliberated on by a panel, such as cases of researchers who have high a *h*-index, but whose research standing has waned or whose contributions to well-cited papers were relatively minor.

It should be noted here that the average gap between rating and recording of the *h*-index in this data set is c. 2.5 years, meaning that the actual *h*-index at the time of rating would have typically been slightly lower for each researcher. I controlled for the time gap between rating and scoring of each *h*-index in the statistical model by including it as a covariate. Although the marginal means for each category are adjusted according to this covariate, the mean *h*-index at the time that rating took place will still be slightly overestimated for some researchers. This problem

of overestimation of the *h*-index at the actual time of rating also applies to previous analyses^{8,17} and cannot account for the above-mentioned inflation of the *h*-index per rating category over the past decade and more.

It is uncertain whether the NRF will continue with the rating of researchers, even though it has not publicly indicated any intention to phase out the system. It has already been pointed out that the rating system seems to be quite disconnected with allocations of large state resources to the South African Research Chairs initiative⁸ and rating itself is no longer connected with any significant 'incentive' funding allocations to researchers.²³ On the basis of cost and effort relative to application, it seems unlikely that the rating system will persist, but the general need for an evaluation of the standing of a researcher in a discipline will remain relevant even if the process is eventually bundled into the evaluation of grant applications.

It would be straightforward to extend this study to other disciplines besides the biological sciences, as was done previously by Fedderke⁹ who used data derived from Google Scholar and found strong discipline-specific associations between the *h*-index and probabilities of ratings. Such an exercise would further clarify whether *h*-index norms are more or less universal across the natural and social sciences, or whether each broad discipline needs a different set of norms with which to benchmark the assignment of researchers to different rating categories. There is no doubt that the *h*-index can serve as a valuable benchmark when peer review is biased or blatantly unfair. Some authors have even expressed the view that a rating system based on bibliometrics would be viewed as being more progressive and transparent.^{8,24} In cases where the spread of data such as those in Figure 1 are available, I would even venture to suggest that under conditions of severe resource or time constraints, organisations such as the NRF could consider using the *h*-index in a discipline-adjusted manner as the sole measure of the standing of a researcher in their field. However, peer review will usually provide more nuanced interpretations of career impact than will a single number based on bibliometrics, and should be used whenever it is feasible. For early-career (<8 years) researchers, whose *h*-index will almost always be modest, my advice would be to simply ignore this whole debate and focus on publishing captivating high-quality papers, because that remains the only way to make an impact on one's research field, no matter how it may ultimately be measured.

Acknowledgements

I thank Jane Carruthers for inviting this submission to SAJS as a contribution to international Peer Review Week. I am also grateful to Theresa Coetzer, Barry Lovegrove, Timo van der Niet, Daniel Johnson and Kathryn Johnson and two anonymous reviewers for comments on the manuscript, and Theresa Coetzer for assistance with the assigning of some of the researchers to sub-disciplines (the responsibility for any errors is mine). This research was supported by NRF grant 46372 as part of the DST-NRF South African Research Chairs Initiative.



Competing interests

I declare that there are no competing interests.

References

1. Smith R. Peer review: A flawed process at the heart of science and journals. *J R Soc Med.* 2006;99(4):178–182. <https://doi.org/10.1177/014107680609900414>
2. Nguyen VM, Haddaway NR, Gutowsky LFG, Wilson ADM, Gallagher AJ, Donaldson MR, et al. How long is too long in contemporary peer review? Perspectives from authors publishing in conservation biology journals. *PLoS ONE.* 2015;10(8), e0139783. <https://doi.org/10.1371/journal.pone.0132557>
3. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA.* 2005;102(46):16569–16572. <https://doi.org/10.1073/pnas.0507655102>
4. Egghe L. The Hirsch index and related impact measures. *Annu Rev Inform Sci Technol.* 2010;44:65–114. <https://doi.org/10.1002/aris.2010.1440440109>
5. Benway BM, Kalidas P, Cabello JM, Bhayani SB. Does citation analysis reveal association between *h*-index and academic rank in urology? *Urology.* 2009;74(1):30–33. <https://doi.org/10.1016/j.urology.2008.10.045>
6. Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics.* 2006;67(3):491–502. <https://doi.org/10.1556/Scient.67.2006.3.10>
7. Barnes C. The *h*-index debate: An introduction for librarians. *J Acad Libr.* 2017;43(6):487–494. <https://doi.org/10.1016/j.acalib.2017.08.013>
8. Fedderke JW. The objectivity of National Research Foundation peer review in South Africa assessed against bibliometric indexes. *Scientometrics.* 2013;97(2):177–206. <https://doi.org/10.1007/s11192-013-0981-0>
9. Hirsch JE, Buéla-Casal G. The meaning of the *h*-index. *Int J Clin Health Psychol.* 2014;14(2):161–164. [https://doi.org/10.1016/S1697-2600\(14\)70050-X](https://doi.org/10.1016/S1697-2600(14)70050-X)
10. Iglesias JE, Pecharroman C. Scaling the *h*-index for different scientific ISI fields. *Scientometrics.* 2007;73(3):303–320. <https://doi.org/10.1007/s11192-007-1805-x>
11. Kelly CD, Jennions MD. The *h* index and career assessment by numbers. *Trends Ecol Evol.* 2006;21(4):167–170. <https://doi.org/10.1016/j.tree.2006.01.005>
12. Bornmann L, Daniel HD. Does the *h*-index for ranking of scientists really work? *Scientometrics.* 2005;65(3):391–392. <https://doi.org/10.1007/s11192-005-0281-4>
13. Bornmann L, Wallon G, Ledin A. Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. *PLoS ONE.* 2008;3(10), e3480. <https://doi.org/10.1371/journal.pone.0003480>
14. Hirsch JE. Does the *h* index have predictive power? *Proc Natl Acad Sci USA.* 2007;104(49):19193–19198. <https://doi.org/10.1073/pnas.0707962104>
15. Derrick GE, Haynes A, Chapman S, Hall WD. The association between four citation metrics and peer rankings of research influence of Australian researchers in six fields of public health. *PLoS ONE.* 2011;6(4), e18521. <https://doi.org/10.1371/journal.pone.0018521>
16. Norris M, Oppenheim C. Peer review and the *h*-index: Two studies. *J Informetr.* 2010;4(3):221–232. <https://doi.org/10.1016/j.joi.2009.11.001>
17. Lovegrove BG, Johnson SD. Assessment of research performance in biology: How well do peer review and bibliometry correlate? *Bioscience.* 2008;58(2):160–164. <https://doi.org/10.1641/B580210>
18. Meho LI, Rogers Y. Citation counting, citation ranking, and *h*-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *J Am Soc Inform Sci Technol.* 2008;59(11):1711–1726. <https://doi.org/10.1002/asi.20874>
19. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. *Science.* 2007;316(5827):1036–1039. <https://doi.org/10.1126/science.1136099>
20. Stallings J, Vance E, Yang JS, Vannier MW, Liang JM, Pang LJ, et al. Determining scientific impact using a collaboration index. *Proc Natl Acad Sci USA.* 2013;110(24):9680–9685. <https://doi.org/10.1073/pnas.1220184110>
21. Chapman CA, Bicca-Marques JC, Calvignac-Spencer S, Fan PF, Fashing PJ, Gogarten J, et al. Games academics play and their consequences: How authorship, *h*-index and journal impact factors are shaping the future of academia. *Proc Roy Soc B Biol Sci.* 2019;286(1916), Art. #20192047. <https://doi.org/10.1098/rspb.2019.2047>
22. Costas R, Franssen T. Reflections around 'the cautionary use' of the *h*-index: Response to Teixeira da Silva and Dobranszki. *Scientometrics.* 2018;115(2):1125–1130. <https://doi.org/10.1007/s11192-018-2683-0>
23. 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. <https://doi.org/10.17159/sajs.2019/5445>
24. 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. <https://doi.org/10.17159/sajs.2018/20170344>