South African Journal *of* Science

volume 114
number 3/4

The Square Kilometre Array and South Africa's economic development

Water, sustainable development and SDG 6

Usability of online banking services in South Africa

Temporal ranges and ancestry: The case of *Australopithecus sediba*

ASSAf
ACADEMY OF SCIENCE OF SOUTH AFRICA

Volume 114
Number 3/4
Mar/Apr 2018

South African Journal of Science

eISSN: 1996-7489

Leader

News & Views

Book Review

Scientific Correspondence

Commentary

Published by
the Academy of Science of
South Africa (www.assaf.org.za)
with financial assistance from the
Department of Science & Technology.

Design and layout
SUN MeDIA Bloemfontein
T: 051 444 2552
E: admin@sunbloem.co.za

Correspondence and enquiries
sajs@assaf.org.za

Submissions
Submissions should be made at
www.sajs.co.za

Review Article

Research Article

Cover caption
South Africa's MeerKAT
radio telescope under construction
in the Karoo, August 2016 (photo courtesy
of Dr Fernando Camilo). MeerKAT is a precursor to
the Square Kilometre Array. In an article on page 78, Bhogal
explores the factors affecting the impact of the Square Kilometre
Array on South Africa's economic development.

Suffer, little children: Paying the price of 'free' higher education

That the 75.1% pass rate for Grade 12 learners in 2017 is a fiction (*real* 'fake news') is common knowledge. About half of the learners who enter Grade 1 never make it to Grade 12 – the loss between Grades 10 and 12 alone is around 41%, making the real pass mark closer to 37% than to 75%. Equally well known are South African schools' poor performances in reading and in mathematics and science, while significant numbers of young people learn in mud, wood, zinc or asbestos schools. Open pit latrines are not unknown.

These reminders are a prelude to concerning data set out in the South African National Treasury's *Estimates of National Expenditure 2018*.[1] But before turning to Treasury, it is worth noting that both Treasury and the Department of Higher Education and Training (DHET) have very recently been at pains to remind universities that 'free' higher education will be implemented for the approved category of students. Treasury noted that, although the funding details are still under consideration, the plan will go ahead, while DHET reminded education leaders that universities and TVET colleges must not deviate from their approved enrolment plans. Costs to the state of 'free' education in 2018 are estimated to be about ZAR15–17 billion, growing to an estimated ZAR40 billion within 3 years as the funding is extended into students' second and third years of study.

It is not clear just how the newly necessary allocation of funds over the next 3 years (presently set at ZAR57 billion) will affect other aspects of public funding for the Higher Education and Training sector. What is clear from the *Estimates of National Expenditure 2018*, however, is that part of the needed funding, amounting to almost ZAR11 billion, will come from reductions in the budget for Basic Education.

The largest portion of the 'savings' will come from various aspects of Infrastructure Funding; the originally budgeted sum of ZAR42.6 billion will be reduced by ZAR10.9 billion to ZAR31.7 billion – slightly more than a 25% cut. Treasury's view is that these reductions will collectively result in 'delays in completing outstanding projects'[1(p.265)]. These reductions will also mean that many rural learners will continue to go to school in mud classrooms.

The second budget reduction is in the area of Improving Matric Completion Rates and is focused specifically on the Second Chance Programme. Here, the budget of ZAR261 million is reduced by ZAR117 million (45%) which is expected to result in 'slower expansion of the programme [in] all priority subject areas [including mathematics and science] and districts'[1(p.266)]. Interestingly, by focusing on the Second Chance Programme, improving matric completion rates does not begin to address the problem of learners who do not ever reach Grade 12.

In the Curriculum Implementation and Monitoring sub-programme, which aims to address South Africa's poor international performances in reading and in mathematics and science, the budget of ZAR1.2 billion will be reduced by ZAR50.5 million (a modest 4%), while bursaries intended to improve the supply of qualified teachers in mathematics, science and technology will not be reduced, but will progressively result in a reduction of some 2500 bursaries (about 18%) in each of the next two budgetary periods.

There are three reasons why reductions in parts of the budget for Basic Education are unacceptable – in fact, counterproductive. The first is that, based on performance and conditions in the sector, there is clearly a need for additional, rather than fewer, funds – and for better schools and improved management and teaching. That funding is being reduced to help cover the costs of an ill-conceived decision regarding Higher Education makes no sense at all. It simply reduces the chances of young people in the school system ever entering a university as they will not reach or pass Grade 12. As Sean Mfundza Muller, an economist at the University of Johannesburg, explains:

> *When the Fees Must Fall movement emerged, it insisted its fundamental demands were based on concern for poor South Africans. The movement argued that this group was effectively excluded from higher education or disadvantaged in their studies because they could not afford the fees and other costs of studying. The idea that the movement for free higher education is based on a concern for poor youth is clearly absurd when you consider that only 5% of South Africans aged between 15 and 34 are students in universities, while 34% are unemployed.*[2]

The second reason why reducing the Basic Education budget to help pay for Higher Education makes no sense is that it effectively means that more tax payers will pay for Higher Education while Basic Education continues to be woefully inadequate and will serve to increase inequality rather than reduce it. According to Sean Archer from the University of Cape Town:

> *There are persuasive arguments that free higher education will be unambiguously regressive. This is because it involves a transfer of resources from lower to higher income individuals within a national population.*[3]

Or, as some have put it, 'free' Higher Education implies that taxi drivers (amongst others) must help to pay to educate lawyers.

Finally (and this has reference to Dr Muller's point quoted above), there are almost twice as many young people in the 18 to 25 age group who are not in employment, education or training as there are in universities and TVET colleges. Almost none of them receives a social grant, and many live on the streets. There are no state funds available to help this population.

Weakening the Basic Education system rather than strengthening it, is, itself, an intrinsically regressive step that puts more young people at risk.

References

1. South African National Treasury. Estimates of National Expenditure 2018. Pretoria: National Treasury; 2018. Available from: http://www.treasury.gov.za/documents/national%20budget/2018/ene/FullENE.pdf
2. Mfundza Muller S. Free higher education in South Africa: Cutting through the lies and statistics. Mail & Guardian. 2018 January 25;Education. https://mg.co.za/article/2018-01-25-free-higher-education-in-south-africa-cutting-through-the-lies-and-statistics
3. Archer S. The case against free higher education: Why it is neither just nor ethical. The Conversation. 2018 February 27;Education. Available from: https://theconversation.com/the-case-against-free-higher-education-why-it-is-neither-just-nor-ethical-92193

HOW TO CITE: Butler-Adam J. Suffer, little children: Paying the price of 'free' higher education. S Afr J Sci. 2018;114(3/4), Art. #a0266, 1 page. http://dx.doi.org/10.17159/sajs.2018/a0266

Open access in South Africa: A coherent strategy is needed

AUTHORS:
Malcolm S. Hartman[1]
Shangmou (Samuel) Wu[2]

AFFILIATIONS:
[1]Bachelor of Health Sciences (Honours) Program, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
[2]Department of Health Sciences, McMaster University, Ontario, Canada

CORRESPONDENCE TO:
Malcolm Hartman

EMAIL:
malcolm.hartman2@gmail.com

KEYWORDS:
health policy; knowledge translation; health research

HOW TO CITE:
Hartman MS, Wu S. Open access in South Africa: A coherent strategy is needed. S Afr J Sci. 2018;114(3/4), Art. #a0259, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0259

PUBLISHED:
27 Mar. 2018

If you asked an idealist to describe the scientific endeavour, the phrase 'seeking the objective truth' might come to mind. But how is this truth obtained? The sharing and translating of knowledge is perhaps the key ingredient, along with a dash of persistence and luck. As budding researchers, this notion excites us – the thought that our ideas can cut across the globe, serving a greater purpose of helping others along their path of truth seeking.

One of us (MH) left Canada after graduating to work with the South African Department of Planning, Monitoring and Evaluation. The work involved a literature review and replying to requests for evidence summaries that flowed into the department. However, there was immediately a barrier to overcome that we have never encountered within our bubble of Canadian academia. This barrier – both a literal and virtual wall – most often presented itself as a paywall; colleagues in the department simply could not access a significant amount of published scientific research without having to pay exorbitant fees for access.

Hailing from Canada means our home university affords us the privilege of institutional access to more publications and data than we would ever use. Encountering paywalls for the first time, brought up the question of how it was possible to function in a research unit without this ability to tap into the shared scientific body of knowledge. How can research units, let alone a branch of the South African government, be expected to deliver quality, evidence-informed deliverables in a timely and feasible manner if the very evidence it relies on is inaccessible? The more we questioned, the clearer it was to us that this was not an isolated challenge.

A recent survey of the capacity for evidence synthesis in Africa amongst 176 researchers, government officials and NGO workers indicated that a lack of access to research literature and data was one of the greatest hindrances to their work – second only to lack of funding.[1]

As of 2015, only 48% of the scholarly outputs from South Africa were openly accessible.[2] The challenges of openly sharing scientific publications are not new. The largest abstract and citation database of peer-reviewed literature, Scopus, listed only 6.8% of articles as openly accessible in 2009.[3] This research climate may contribute to the fact that in the South African government, the use of research outputs as a main source of evidence is tepid, at best. In a recent survey, only 6 of 54 senior officials in the government indicated that they rely on research, while the majority (38) responded that they rely on their networks.[4] While inability to access publications was not included in this survey, it would be reasonable to hypothesise the role it plays.[5] Although tacit knowledge from strong networks can be useful, research data and publications remain imperative for evidence-informed decision-making.[6] In situations where institutional subscriptions are not feasible, open access provides another paradigm to facilitate evidenced-informed decision-making.[7,8]

The case for open access is perhaps best understood in the context of health care. In the era of evidence-based medicine,[9] the public has an underlying assumption that physicians and frontline health-care workers are updated with high-quality health information so as to provide the best care possible. Prior to June 2017, this assumption would have been partially inaccurate in South Africa as health-care workers did not have access to the Cochrane Library, the gold-standard source for health-care reviews.[10]

Beginning in the early 1990s, the push for open access has gained traction in the mainstream research community, albeit at a slow pace.[3] This lag is why the call for immediate open access for all articles published in the European Union (EU) by 2020 was welcomed as a firm step forward.[11] As the EU moves forward with open access policies, it will be crucial for African countries, especially South Africa, to take a strong stance in support of open access, not only to protect their research interests, but also to maintain the goal of increasing the global sharing of scientific knowledge.[12,13]

With these efforts, South Africa may also benefit from tailored open access policies that better meet the needs of their knowledge users and producers. There are two pathways to open access that the EU will rely on: green (self-archiving) and gold (publishing in open access journals).[11,14] However, both green and gold open access pose some challenges for researchers in South Africa. Gold open access usually shifts the costs from subscribers to researchers through an article-processing fee, which may be too high for authors/institutions, whereas green open access typically requires the institutional infrastructure to host a repository of articles that can be accessed via online searches.[13,15]

The reinvigorated open access movement in the EU presents a critical point for South African research communities. If open access is not addressed on a national level, South Africa may lose out on contributing to and accessing this growing platform. One model to look towards is the Scientific Electronic Library Online (SciELO)-SA, South Africa's largest existing open-access database.[16] SciELO is funded by the South African Department of Science and Technology and could be built upon to form the backbone of a future national open access strategy.

Knowledge is not useful if it cannot be shared, accessed and built upon. Decision-makers in the government, researchers in a university or health-care workers all share the need for accessible scientific literature. As nations push forward with national open access policies, South Africa must act coherently to develop open access policies which allow the country to access this growing market while addressing the challenges that come with the current paradigm of open access publishing.

References

1. Stewart R, Nduku P, Langer L. Capcity in Africa: The results of a survey on support for and production of evidence maps and evidence syntheses, including systematic reviews. Johannesburg: Africa Centre for Evidence; 2017.
2. Academy of Science of South Africa. South African scholarly journal landscape demonstrates growth in open access [document on the Internet]. c2015 [cited 2017 Jul 31]. Available from: http://www.assaf.co.za/newsletter/?p=1179
3. Laakso M, Welling P, Bukvova H, Nyman L, Björk B-C, Hedlund T. The development of open access journal publishing from 1993 to 2009. PLoS ONE. 2011;6(6), e20961, 10 pages. https://doi.org/10.1371/journal.pone.0020961
4. Paine G, Sadan M. Use of evidence in policy making in South Africa: An exploratory study of attitudes of senior government officials. Afr Eval J. 2015;3(1), a145,10 pages. https://doi.org/10.4102/aej.v3i1.145
5. Bowen S, Erickson T, Martens PJ, Crockett S. More than "using research": The real challenges in promoting evidence-informed decision-making. Healthc Policy. 2009;4(3):87–102. https://doi.org/10.12927/hcpol.2009.20538
6. Lavis JN, Robertson D, Woodside JM, McLeod CB, Abelson J. How can research organizations more effectively transfer research knowledge to decision makers? Milbank Q. 2003;81(2):221–248. https://doi.org/10.1111/1468-0009.t01-1-00052
7. Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14(1), 1144, 9 pages. https://doi.org/10.1186/1471-2458-14-1144
8. Zachariah R, Kumar AM, Reid AJ, Van den Bergh R, Isaakidis P, Draguez B, et al. Open access for operational research publications from low- and middle-income countries: Who pays? Public Health Action. 2014;4(3):142–144. https://doi.org/10.5588/pha.14.0028
9. Sackett DL. Evidence-based medicine. Semin Perinatol. 1997;21(1):3–5. https://doi.org/10.1016/S0146-0005(97)80013-4
10. Cochrane. National one-click access to evidence-based Cochrane Reviews for South Africans: The Cochrane Collaboration [homepage on the Internet]. c2017 [cited 2017 Sep 04]. Available from: http://www.cochrane.org/news/national-one-click-access-evidence-based-cochrane-reviews-south-africans
11. Enserink M. In dramatic statement, European leaders call for 'immediate' open access to all scientific papers by 2020. Science News. 2016 May 27. https://doi.org/10.1126/science.aag0577
12. Czerniewicz L, Goodier S. Open access in South Africa: A case study and reflections. S Afr J Sci. 2014;110(9/10), Art. #2014-0111, 9 pages. https://doi.org/10.1590/sajs.2014/20140111
13. Tempest D. Open access in Africa: Changes and challenges: Elsevier Connect. 2013 June 04 [cited 2017 Sep 04]. Available from: https://www.elsevier.com/connect/open-access-in-africa-changes-and-challenges
14. Harnad S, Brody T, ValliÃ¨ res F, Carr L, Hitchcock S, Gingras Y, et al. The access/impact problem and the green and gold roads to open access. Serials Rev. 2004;30(4):310–314. https://doi.org/10.1080/00987913.2004.10764930
15. Swan A, Willmers M, King T. Costs and benefits of open access: A guide for managers in southern African higher education [document on the Internet]. c2014 [cited 2017 Sep 04]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.899.9086&rep=rep1&type=pdf
16. Academy of Science of South Africa. SciELO South Africa [homepage on the Internet]. No date [cited 2017 Sep 04]. Available from: http://www.scielo.org.za/?lng=en#about

Launch of the ASSAf Presidential Roundtable: University rankings

AUTHOR:
John Butler-Adam[1]

AFFILIATION:
[1]Academy of Science of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
John Butler-Adam

EMAIL:
j.butleradam@gmail.com

KEYWORDS:
ranking systems; indicators; Quacquarelli Symonds; Times Higher Education

HOW TO CITE:
Butler-Adam J. Launch of the ASSAf Presidential Roundtable: University rankings. S Afr J Sci. 2018;114(3/4), Art. #a0265, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0265

PUBLISHED:
27 Mar. 2018

On 7 February this year, the Academy of Science of South Africa (ASSAf) launched its Presidential Roundtable series on Science, Scholarship and Society at an event in Stellenbosch, with a discussion on the subject of 'University Rankings: Helpful or Harmful?'

The roundtable discussions are a quarterly roundtable of experts in specific fields, in each case addressing a critical issue percolating in society that requires the deliberation of the best minds on the topic.

In 2013, the Journal's Leader, titled 'Being the best? Yes – but best for what?'[1], expressed several concerns about rankings:

> *The ranking system assumes that there is just one kind of university, with common criteria for measuring comparative success, while in many countries there are institutions that differ in terms of their markets and purposes in the higher education system.*

That concern remains true 4 years later, and while the presenters at the first roundtable provided a wide range of (often differing) views about, and insights into, the ranking systems, the old theme was a common view, although it was expressed in a variety of ways. Professor Lis Lange, for instance, expressed one of her concerns as follows:

> *One of the unintended consequences of rankings is that the idea of being in the top 100 becomes the strategy of universities. The whole being of the university is reduced to being one in the top 100 and this has very serious implications.*[2]

Jonathan Jansen, President of ASSAf, who moderated the roundtable put it this way:

> *Ranking for the sake of claiming bragging rights or boosting national egos is a problem, for then the practice of rank-ordering universities serves simply as a hurtful reminder of the academic inequities embedded in the global system of knowledge production.*[3]

Only one system may escape these concerns – the U-Multirank system, which is both more sophisticated and more complicated than other major ranking systems.

This article is not, however, an overview of the four presentations (which is given elsewhere[4]) but a consideration of some of the implications of the different indicators, definitions and variables, and varying metrics, used by the major ranking systems. There are about 30 'global' ranking systems for universities, and 31 countries have their own (often multiple) internal ranking systems. Of the global rankings, there are really just four that are consistently taken seriously – Academic Ranking of World Universities (ARWU, formerly Shanghai Jiao Tong Rankings); Quacquarelli Symonds (QS); Times Higher Education (THE); and University Ranking by Academic Performance (URAP). QS and THE rely primarily (but not solely) on information submitted by institutions in response to the questions posed by the ranking system while ARWU relies on Internet sources and URAP specifically on information available from the Web of Science and InCites.

There are two major implications of the different indicators, definitions and variables, and varying metrics, used by the major ranking systems. The first is that the systems are not comparable with one another and so relating a ranking on, say, the QS and ARWU lists makes no sense. The second is that the systems change their methodologies in various ways from time to time, and the participating institutions change in number from year to year, so that longitudinal comparisons for any one university most often make little or no sense. To make the point about variables and weightings, consider the QS and THE systems shown in Table 1.

Table 1: A comparison of the different indicators used by two university ranking systems: Quacquarelli Symonds (QS) and Times Higher Education (THE)

Indicator	QS	THE
Reputation	Academic reputation (survey) (40%)	
	Employer reputation (survey) (10%)	
Teaching	Faculty:student ratio (20%)	Five sub-measures (survey; staff:student ratio; PhD:undergrad ratio; PhD awarded/staff; institutional income) (30%)
Research	Citations per paper (20%)	Three sub-measures (survey; research income; research productivity) (30%)
		Citations (30%)
International profile	International:local staff ratio (5%)	Three sub-measures (staff; students; collaboration) (7.5%)
	International:local students ratio (5%)	
Institutional income	N/A	Industry income (2.5%)

Different ways of measuring, varying definitions, different weightings and, in three instances, different indicators, mean that, other than in exceptional cases, there can be little or no comparability. And even in the 'top 10' case, the specific rankings vary despite the tight, high-level competition. Figure 1 shows the 2018 rankings for the top 10 institutions as determined by QS and THE – where only the 'bottom' three institutions have consistent ranks, while Princeton University does not appear in the QS list, nor University College London on the THE list.

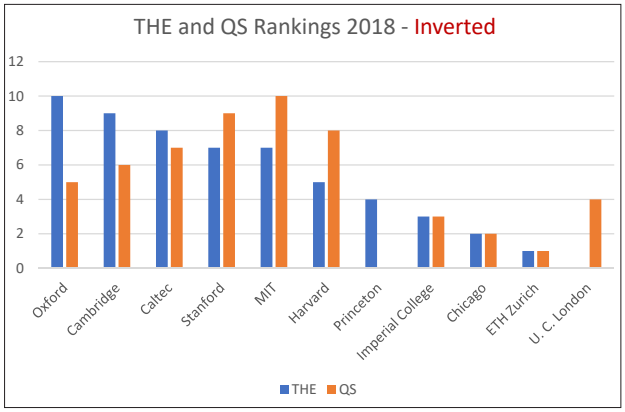


Figure 1: 2018 Rankings for the top 10 institutions as determined by Quacquarelli Symonds (QS) and Times Higher Education (THE) ranking systems. Princeton University does not appear in the QS list and University College London does not appear on the THE list.

As far as year-on-year comparisons of rankings outcomes go, these are made very difficult by regular, often yearly, methodological changes, including changes to citations and survey data window periods, bibliometric data and periods that are considered, and percentages assigned to local and international perceptions. In addition, the expansion of rankings lists increases the pool of ranked universities each year and this renders trend conclusions meaningless by varying the scale. It also tends to make ranking a zero-sum game. This is also complicated by the proliferation of ranking systems in recent years in all rankings spheres: global, regional, young, subject rankings and employability.

Although varying in their approaches to the question posed by the ASSAf Presidential Roundtable, the presenters agreed on one key matter: although rankings are often decried (even derided) in public, they are assiduously followed by universities and their leaderships, and so they are (for the meantime) an unavoidable reality, one which may serve to influence institutional decision-making – and spending. At the same time, they are also dubious measures to use in any attempt to undertake a systematic analysis of their results within and between the systems.

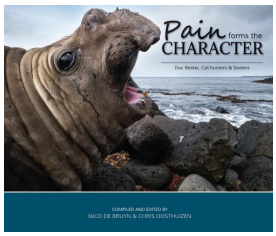
References

1. Butler-Adam J. Being the best? Yes – but best for what? S Afr J Sci. 2013;109(9/10), Art. #a0038, 1 page. http://dx.doi.org/10.1590/sajs.2013/a0038
2. Basson A. Rankings ignore local contexts of universities, say experts [webpage on the Internet]. c2018 [cited 2018 Mar 15]. Available from: http://www.sun.ac.za/english/Lists/news/DispForm.aspx?ID=5419
3. Jansen J. Rankings not whole story. Herald Live. 2018 February 15. Available from: http://www.heraldlive.co.za/opinion/2018/02/15/jonathan-jansen-rankings-not-whole-story/
4. Makoni M. The great global rankings debate. University World News. 2018 March 09. Available from: http://www.universityworldnews.com/article.php?story=20180306114540487&query=great+global+rankings+debate

A tribute to storytelling, camaraderie and the Prince Edward Islands

BOOK TITLE:
Pain forms the character: Doc Bester, cat hunters & sealers

BOOK COVER:



EDITORS:
Nico de Bruyn and Chris Oosthuizen

ISBN:
9780620749121

PUBLISHER:
Antarctic Legacy of South Africa, Stellenbosch; ZAR650

PUBLISHED:
2017

REVIEWER:
Bettine van Vuuren

AFFILIATION:
Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Johannesburg, South Africa

EMAIL:
bettinevv@uj.ac.za

HOW TO CITE:
Van Vuuren B. A tribute to storytelling, camaraderie and the Prince Edward Islands. S Afr J Sci. 2018;114(3/4), Art. #a0261, 1 page. http://dx.doi.org/10.17159/sajs.2018/a0261

PUBLISHED:
27 Mar. 2018

The Prince Edward Islands were claimed by South Africa in 1947/1948, and officially annexed in 1948 through the *Prince Edward Islands Act, 1948* (largely under pressure to have a presence in the Southern Ocean after World War II). The first Biological–Geological Expedition to Marion and Prince Edward Islands arrived at Transvaal Cove on 4 January 1965. Although it was initially intended as a once-off visit to gather scientific information, this first expedition actually became the cornerstone of a strong (scientific) house built by ongoing scientific expeditions to the islands. Through the efforts of a large number of researchers – both national and international, young and old, world-leader and student – an impressive body of scientific literature has emerged with the Prince Edward Islands as its focal point. Detailed and in-depth knowledge has been gathered of the biotic and abiotic environment on and around the islands; knowledge which has contributed to our understanding of change in the global context. However, our South African National Antarctic Programme is more than just strong terrestrial and oceanographic research; there is an equally important human component to it. There are human characters and stories behind the research, and this is what makes the South African National Antarctic Programme special.

Understanding and knowing our history helps us to understand ourselves. Storytelling is an important component of passing on information from one generation to the next. Although our world is moving towards digital and binary codes, it is essential that we do not lose our ability to tell stories, and to appreciate the joy that these bring. As the editors point out in their note, a 'handful of characters – by a complex mix of individuality, context, societal influence, opportunity, determination, enduring involvement and hard work – have had a proportionately large impact on several research programmes and indeed life at Marion Island as a whole'. One such person is Marthán Bester, and this book pays tribute to the enormous contributions that Marthán has made to the Cat Eradication Programme, and the establishment of the Marine Mammal Programme on Marion Island. But this book is more than just a tribute to Marthán; it is a compilation of stories about people and events on Marion Island over many years – some humorous, some philosophical, but all revealing intimate feelings and the strong camaraderie that one typically experiences whilst visiting one of the most beautiful places on earth: the Prince Edward Islands. It includes memories from other Marion giants such as Valdon Smith, John Cooper, Rudi van Aarde, Peter Ryan, Bruce Dyer, and Steven Chown, as well as numerous others who have worked on the Islands. It tells stories about the Islands themselves, about spending time in the field and on various ships, about falling and getting up (literally and figuratively), about friendship, but all of these are underlined by a strong sense of purpose – to be the best that you can possibly be at the task in hand.

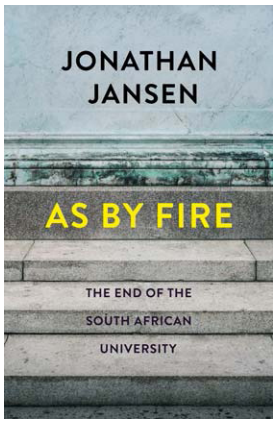
Marion Island has, over the years, seen its fair share of 'weird and wondrous'. From sealers to sheep, from a proposed landing strip (thankfully never realised) to an actual misguided light aircraft landing. There were also several very unfortunate events, including the loss of human lives, shipwrecks, the partial destruction of the base by a fire, and a few medical emergency evacuations. Arguably, one of the most unfortunate events was the introduction of the first cats in 1948, which paved the way for an eventual exponential increase in very skilled feline hunters across Marion Island. The negative impacts that these cats had on several bird species could potentially be matched only by the current mouse infestation. Rudi van Aarde played an important part in initially highlighting the negative impacts that cats had, and these initial efforts resulted in the conclusion of the Cat Eradication Programme under Marthán's watchful eye and skilled hands. Marthán also set up the long-term Marine Mammal Programme, which has been gathering information for a large number of years (perhaps in the running for the most valuable long-term monitoring programme in South Africa).

Reading this book raises a complex set of emotions for me, as I am sure it will for any person fortunate enough to have ever visited these sentinels in the Southern Ocean. The absolute awe when one first sees the Islands, frustration brought on by sheer exhaustion in the field, happiness from spending time with friends and colleagues, motivation and inspiration brought on by research ideas and collaborations, anticipation of seeing loved ones back home, but above all, absolute gratitude for the opportunity to have experienced these Islands. The collection of stories collated here touches on all of these. These stories serve the purpose to highlight and acknowledge the contributions of one such a giant (Marthán Bester), but they also serve to inspire others to experience for themselves what Marion Island is about. *Pain forms the Character* is a tribute to stories and storytelling, to camaraderie, and above all, to the Prince Edward Islands – our sentinels in the Southern Ocean.

A slice of higher education – the widening gyre

BOOK TITLE:
As by fire: The end of the South African university

BOOK COVER:



AUTHOR:
Jonathan Jansen

ISBN:
9780624080305 (softcover)

PUBLISHER:
Tafelberg, Cape Town; ZAR280

PUBLISHED:
2017

REVIEWER:
Robert J. Balfour

AFFILIATION:
Faculty of Education, North-West University, Potchefstroom, South Africa

EMAIL:
robert.balfour@nwu.ac.za

HOW TO CITE:
Balfour RJ. A slice of higher education – the widening gyre. S Afr J Sci. 2018;114(3/4), Art. #a0260, 1 page. http://dx.doi.org/10.17159/sajs.2018/a0260

PUBLISHED:
27 Mar. 2018

Reading any one of Jonathan Jansen's books is like taking a slice of South Africa at a given moment in time. At the tip of the slice is a point. *Knowledge in the Blood* (UCT Press; 2009), took as its 'point' the experience of students in a period of profound change: captured still by the imaginary of a racialised past; that book described the experience of racism in a post-apartheid, but by no means post-racist, state. In attempting to understand what racist beliefs offered a generation of students (white and black), Jansen illustrated how identity politics and prospects could be shifted pending the intervention of a leader. So too in *As by Fire,* the 'point' of leadership in the #FeesMustFall movement becomes the focus of intense reflections, often beginning as personal, and then widening to a social focus on the fissures of an increasingly divided South Africa in which the fractures tearing universities apart are revealed to be symptomatic of the post-apartheid state's approach to welfarism, the massification of education at the cost of limited and even declining expenditure on education in general, and higher education in particular.

The book demonstrates a researcher's skilful treatment of data, reportage, history and analysis and how the narratives of vice chancellors during the Fallist events understood their histories, the histories of their students and the context in which both came to forceful, and often times painful, confrontations between people tasked to lead and manage a crisis not of their making, against the students to whom their very professions, and sometimes even lives, are dedicated. The text casts vice chancellors in the same dialogic as that represented by a far more critical and self-serving press, as noted by Jansen himself. If the dialogic remains the same, the representations of leaders in this text seems to be a wilful correction to what Jansen obviously perceives as the vilification of vice chancellors as hapless subjects. By the time the book nears its conclusion, the reader comes to feel as though South Africa, as presented through the very different institutional contexts of 11 South African universities and their leaders, comes to be known, albeit from a very specific angle. The angle is not simply the individual in the text, but also the way in which the text is treated and the individual account is 'couched'. South Africa sceptics will almost inevitably attribute the Fallist crisis to a cocktail of failed African Nationalist Congress policies aided by a combination of greed and incompetence. But vignettes described by vice chancellors, all of whom are shown to be deeply committed to the broader transformation of education in South Africa, suggest a failure of another kind – a fundamental series of misrecognitions: of the potential of higher education to address wider social issues concerning centuries old inequalities, of the role of the university as a response to the welfare needs of the poor youth particularly, and of the capacity of the universities to adequately support let alone redress the inequities of a poor quality education system in South Africa. It is this last aspect that remains somewhat underexplored against the lack of adequate planning to enable the radical transformation of higher education commenced by Kader Asmal in 2004, to come to its full realisation in terms of the lived experience of transformation as a narrative of success, rather than a series of brutally experienced frustrations and disappointments (p.178).

The reality is evident all around us; government, through negligence, has helped create a two-tier education system: that which can be afforded by the enabled and that which cannot. Jansen's analysis of the missing middle affirms that the new apartheid cuts across race and gender lines of past exclusions, neither effacing nor erasing what has come before, but rather adding to inequality at the inter-sectionalities of race, class and gender. Although Jansen analyses the causes of the many stresses within the higher education sector, and although the book takes as its context the university as the setting for various spectacular confrontations, the role of the institution in South Africa (as opposed to in the UK, for example) receives too little attention in the book – a point to which I return later.

In *As by Fire*, the brunt of this public disappointment with the promise of education is carried by students and universities, and vice chancellors whose responses and reactions reveal the extent of the lack of foresight and sheer indifference by the state. In *As by Fire,* that moment or historical event begins with the rupture of the #RhodesMustFall movement and charts the widening gyre (a deliberate reflection on Yeat's 'The Second Coming' and Achebe's echo of *Things Fall Apart*) associated with #FeesMustFall and other Fallist movements across South Africa from 2015 to 2016. With the exception of Chapter 7, the chapters of this book rely on the testimony of university vice chancellors, as coordinated through the narrative of another vice chancellor – Jansen himself. The book aims at accessibility in terms of its anticipated audience and eschews any claim at rigorous research. However, underlying the narratives and the encompassing text, those who are aware of Jansen's research skills (and the many accounts he has written on higher education on far-reaching themes) will know that the insights provided arise from an acute and nuanced understanding of how an historical event comes to be generated; for example, in Chapter 8, the ambiguous connections between welfare and education as a perceived form of social welfare are described incisively.

These evident strengths noted, this book is one that offers sympathetic portraits of vice chancellors, not all of whom have been found – since its publication – to be as worthy of the praises sung by Jansen. The conclusions drawn are not unexpected: address the funding crisis and the schooling quality crisis; limit state interference; support the development of the black and female professoriate from within (and not through external mechanisms such as the NIHSS); and protect universities against the ravages of instability and insecurity. South Africa has known violent protest within institutions since the 1960s; thus it is not surprising that the book concludes with a sobering concern about the institutionalisation of violence. The antidote Jansen suggests is a renewal of a social contract – what is termed a solidarity between society and universities in opposition to elements or impulses towards the destruction of vulnerable institutions.

Do specimens attributed to *Lystrosaurus murrayi* and *L. declivis* (Triassic Therapsida) represent one species?

AUTHOR:
J. Francis Thackeray[1]

AFFILIATION:
[1]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Francis Thackeray

EMAIL:
francis.thackeray@wits.ac.za

KEYWORDS:
holotypes; dicynodont; conspecific; *Lystrosaurus*; therapsid

HOW TO CITE:
Thackeray JF. Do specimens attributed to *Lystrosaurus murrayi* and *L. declivis* (Triassic Therapsida) represent one species? S Afr J Sci. 2018;114(3/4), Art. #a0258, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0258

PUBLISHED:
27 Mar. 2018

Lystrosaurus murrayi and *L. declivis* are two dicynodont (Therapsida, or mammal-like reptile) species commonly represented in Triassic deposits which are referred to as the '*Lystrosaurus* Assemblage Zone' of the South African Karoo, postdating the Permo-Triassic boundary at 252 million years ago. The holotypes of these two species are curated in the Natural History Museum in London. Unfortunately, both are fragmentary (personal observation). Despite the fragmentary condition of the holotypes, Cluver[1], Cluver and King[2] and Brink[3] accepted the view that at least two species of *Lystrosaurus* (*L. murrayi* and *L. declivis*) could be distinguished.

Lystrosaurus murrayi was first described by Thomas Henry Huxley in 1859. A year later, Richard Owen described a similar specimen as *L. declivis*. Recognising similarities in the two holotypes, Broom[4] regarded Owen's specimen as 'probably' representing the same species as that which was described by Huxley in 1859. This view is consistent with those of Thackeray et al.[5] By contrast, Botha-Brink et al.[6] retain the view that two Triassic taxa of *Lystrosaurus* can be distinguished. In an impressive study, they examined as many as 97 crania attributed by them to *L. murrayi*, and 99 crania attributed to *L. declivis*.[6]

Two hypotheses

Thackeray's hypothesis addressed in this paper (H_T) is that specimens attributed to one or other of the two Triassic species are conspecific. H_{BB} is the alternative hypothesis (reflected for example by Botha-Brink et al.[6]) which is that two distinct Triassic species of *Lystrosaurus* can be recognised.

In order to test these two hypotheses, attention was given to the maximum basal skull length (BSL) dimensions in a total of 196 specimens of *Lystrosaurus* in Triassic deposits.[6] Two other species of *Lystrosaurus (L. maccaigi* and *L. curvatus*), both of which occur in Permian deposits, were not considered here but deserve attention in future studies.

A summary of measurements obtained for BSL for specimens examined by Botha-Brink et al.[6] and attributed by them to two Triassic species of *Lystrosaurus*, is given in Table 1. The results presented in Table 1 can be used to determine whether H_T is supported.

Table 1: Comparison of basal skull length measurements for *Lystrosaurus murrayi* and *L. declivis*

	Lystrosaurus murrayi	*Lystrosaurus declivis*
Number of crania	97	99
Mean basal skull length (mm)	106.7	118.1
Standard deviation	31.1	30.0
Coefficient of variation	29.1	25.2

From Table 1, the following observations are made:

1. The mean BSL dimensions in specimens attributed to *L. murrayi* (106.7 mm) is only slightly smaller than that for specimens attributed to *L. declivis* (118.1 mm).
2. The means are not significantly different (Students *t*-test, $p = 0.05$).
3. The standard deviations are almost identical (31.1 and 30.0 mm).
4. The coefficients of variation are almost identical (29.1 and 25.2).

Two important additional observations are:

5. Specimens attributed to the two species are both represented in the *Lystrosaurus* Assemblage Zone, with *L. declivis* apparently occurring only slightly later than *L. murrayi*, which apparently occurs first at the time of the Permo-Triassic boundary, 252 million years ago.[6] Notably, the specimens are coeval in most of the *Lystrosaurus* Assemblage Zone.
6. Specimens attributed to *L. murrayi* and *L. declivis* both display an 'early breeding pattern', which is associated with relatively low percentages of large individuals in the *Lystrosaurus* assemblages in Triassic deposits.[6]

It can be concluded that H_T is supported by all six of these observations.

Further analyses should be undertaken on measurements in addition to BSL to test H_T using a morphometric analysis of the kind developed by Thackeray et al.[7], updated by Thackeray[8] and discussed by Thackeray and Schrein[9] in the context of a probabilistic definition of a species, related to sigma taxonomy[10,11], as opposed to

the general tendency to use alpha taxonomy[12] whereby specimens are classified in terms of discrete taxa, assuming clear boundaries between them.

The six observations presented here, in support of H_T, are consistent with the view held by Broom[4] who stated that Owen's specimen attributed by him to *L. declivis* was 'probably' the same species represented by the holotype of *L. murrayi* described by Huxley in 1859. The six observations presented here are also consistent with those of Cosgriff et al.[13]

Thackeray et al.[5] suggested that much of the variability in specimens attributed to *L. murrayi* and *L. declivis* may be a result of sexual dimorphism (associated in part with 'bossing' on crania above the orbits), or ontogeny within one species. It was indicated that bosses were likely to occur primarily in male specimens, as reflected by the following statement: 'Where supraorbital bosses are present, these occur mainly in relatively large individuals. We consider the presence of supraorbital bosses in almost 50% of large individuals and the absence of such bosses in about 50% of similar-sized individuals attributed to the same species, as a potential indication of differences between adult males and females'[5].

If only one species of *Lystrosaurus* species is represented in Triassic deposits of the South African Karoo, as hypothesised through H_T, the nomen '*murrayi*' would have precedence over '*declivis*' according to rules of nomenclature, as Huxley's specimen was described in 1859, a year earlier than that reported formally by Owen. A possibility to be considered is that *L. murrayi* and *L. declivis* are components within a chronospecies, recognising also the importance of hybridisation in a diversity of modern taxa.[9]

Acknowledgements

The National Research Foundation, the DST-NRF Centre of Excellence and the Andrew W. Mellon Foundation are acknowledged for their support. Jennifer Botha-Brink kindly facilitated access to measurements of specimens analysed by her and her colleagues. Angela Milner provided access to holotypes of *Lystrosaurus* at the Natural History Museum in London. Sir Andrew Huxley (grandson of T.H. Huxley) gave encouraging comments when the holotypes were examined.

References

1. Cluver MA. The cranial morphology of the dicynodont genus *Lystrosaurus.* Ann S Afr Mus. 1971;56(5):155–274.
2. Cluver MA, King GM. A re-assessment of the relationships of Permian Dicynodontia (Reptilia, Therapsida) and a new classification of dicynodonts. Ann S Afr Mus*.* 1983;91(3):195–273.
3. Brink AA. Illustrated bibliographical catalogue of the Synapsida. Handbook 10, Part I. Pretoria: Government Printer; 1982.
4. Broom R. The mammal-like reptiles of South Africa and the origin of mammals. London: Witherby*;* 1932. p. 244.
5. Thackeray JF, Durand JF, Meyer L. Morphometric analysis of South African dicynodonts attributed to *Lystrosaurus murrayi* (Huxley, 1859) and *L. declivis* (Owen, 1860): Probabilities of conspecificity. Ann Transv Mus. 1998;36:413–420.
6. Botha-Brink J, Codron D, Huttenlocker AK, Angielczyk KD, Rutas M. Breeding young as a survival strategy during earth's greatest mass extinction. Sci Rep. 2016;6, Art. #24053. https://doi.org/10.1038/srep24053
7. Thackeray JF, Bellamy CL, Bellars D, Bronner G, Bronner L, Chimimba C, et al. Probabilities of conspecificity: Application of a morphometric technique to modern taxa and fossil specimens attributed to *Australopithecus* and *Homo*. S Afr J Sci. 1997;93:195–196.
8. Thackeray JF. Approximation of a biological species constant? S Afr J Sci. 2007;103:489.
9. Thackeray JF, Schrein CM. A probabilistic definition of a species, fuzzy boundaries and 'sigma taxonomy'. S Afr J Sci. 2017;113(5/6), Art. #0206, 2 pages. https://doi.org/10.17159/sajs.2017/a0206
10. Thackeray JF. Sigma taxonomy in relation to palaeoanthropology and the lack of clear boundaries between species. Proc Eur Soc Stud Hum Evol. 2015;4:220.
11. Thackeray JF. *Homo habilis* and *Australopithecus africanus*, in the context of a chronospecies and climatic change. Palaeoecol Afr. 2016;33:53–58.
12. Mayr E, Linsley EG, Usinger RL. Methods and principles of systematic zoology. New York: McGraw-Hill; 1953.
13. Cosgriff, JW, Hammer WR, Ryan WJ. The Pangaean reptile, *Lystrosaurus maccaigi,* in the Lower Triassic of Antarctica. J Paleontol. 1982;56(2):371–385.

A book regarding Kromdraai: Comments on Herries (2018)

AUTHORS:
Jose Braga[1,2]
J. Francis Thackeray[2]

AFFILIATIONS:
[1]University of Toulouse (Paul Sabatier), Toulouse, France
[2]Evolutionary Studies Institute, School of Geosciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Jose Braga

EMAIL:
jose.braga@univ-tlse3.fr

KEYWORDS:
Paranthropus; early *Homo;* Plio-Pleistocene; stratigraphy

HOW TO CITE:
Braga J, Thackeray JF. A book regarding Kromdraai: Comments on Herries (2018). S Afr J Sci. 2018;114(3/4), Art. #a0264, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0264

PUBLISHED:
27 Mar. 2018

The book[1] entitled *Kromdraai: A birthplace of* Paranthropus *in the Cradle of Humankind, a South African heritage site*, which we co-edited, was published by SUN PReSS (AFRICAN SUN MeDIA) after peer review. The book relates to the Kromdraai Research Project (KRP), which is an enterprising and exciting scientific venture aimed at promoting palaeontological science and transfer of knowledge in the Cradle of Humankind, declared as a World Heritage Site by UNESCO in 1999. The KRP is supported inter alia by the National Research Foundation (South Africa); the Department of Science and Technology (South Africa); the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Union; the Centre National de la Recherche Scientifique (CNRS, France); the Institut des Déserts et des Steppes (France); the French Ministry of Foreign Affairs; and the French Embassy in Pretoria.

New results are presented in the book by an international team (French, South African, Italian and German) to report on the latest developments of research and fieldwork at Kromdraai, a hominin cave complex which has yielded remarkable Plio-Pleistocene fossils attributed to *Paranthropus* and early *Homo*, constituting distant relatives of all humankind. An important outcome presented throughout the book is that the Kromdraai hominins and the associated fauna do not represent one single temporal period. As yet, the chronological interpretations of the Kromdraai hominin-bearing deposits have been essentially based on the unsubstantiated combination of three sets of assemblages successively collected by Robert Broom, Bob Brain and Elisabeth Vrba into Partridge's Member 3.[2] The new discoveries presented in our book are in strong opposition to these previous assumptions, as illustrated by the chronological interpretations made inter alia by Andy Herries and his colleagues.[3]

Herries has recently published a book review[4] of our new volume[1]. As detailed in our book, the vast majority of the fossil sample (including hominins) collected at Kromdraai before 2002 could not be tied precisely to any breccia members identified by Partridge[2]. Until 2002, most of the Kromdraai hominin samples were unprovenanced, including the type specimen of *Paranthropus robustus* (TM 1517), discovered and described by Robert Broom in 1938. However, remarkable hominin and faunal samples have been discovered by our team, and are provenanced in the context of three distinct depositional phases now securely tied in a very detailed stratigraphic succession, in which the distinction between KA and KB localities is no longer justified (Chapter 3). In the new book on Kromdraai, reference is made to Member 2, previously considered as essentially sterile. This is far from reality. With some 2200 newly discovered fossils presented in Chapter 5, Member 2 represents the (as yet) oldest faunal assemblage from Kromdraai where more than 22 hominins have recently been found.[1,5]

All the data presented in the new volume on Kromdraai have better stratigraphic control and provide a more coherent (temporally controlled) assemblage for analysis. Contrary to Herries' statements[4], the Member system used in the book is not incompatible with the reappraisal of our proposed detailed stratigraphy (Chapter 3). Indeed, some members that Partridge[2] differentiated can be associated throughout the whole site (between Kromdraai East, Kromdraai West and even Kromdraai A).

Unfortunately, Herries does not comment on the main scientific issue raised (see above) in our book. Instead, he comments on specific issues in his review of the recent book.[4] Some of them are discussed below.

Regarding the study of fauna from Member 2 (Chapter 5), Herries incorrectly states that it is 'rather devoid of actual data...no actual data are presented...no primary descriptions or metrics...A species list with no primary descriptions and evidence of what the fossils were compared to, is like stating the answer is 42, without any maths'. In response, attention can be drawn to the following. For birds, Table 5.1 presents metric data and descriptions. For primates, Fig. 5.2 summarises an extensive metric study associated with 'evidence of what the fossils were compared to', and metric data are also presented in Tables 5.2 – 5.4. For bovids, Table 5.5 presents NISP metrics (number of identified specimens) in a faunal list. The section is brief and preliminary on account of the fragmentary nature of the material, most of which can be identified only in terms of size groups (medium sized alcelaphines and large to medium sized Bovidae) such that (at this stage of KRP) it is simply not possible to identify most of the specimens to the level of species. Nevertheless, taken together, the ungulate samples from the Kromdraai Extension Site reflect semi-arid grassland, which is a relevant and important palaeoenvironmental assessment even though the bovid samples are so fragmentary. For carnivores, descriptions are given in the text, and Table 5.6 refers to an important 'composite faunal list' which serves to summarise data from a set of excavations undertaken by Brain, Vrba, Thackeray and Braga. Metrics for the carnivores are given in the text and in Tables 5.7 and 5.8.

Herries states that 'One of the *truly fundamental outcomes* of this book' (emphasis added) is 'that we will likely never know the exact provenience, and thus age, of the type specimen of *Paranthropus robustus* [TM 1517]'. This flies in the face of cited research[6], giving exciting results of chemical analyses identifying titanium, manganese, rubidium and zinc in relation to Si and Ca, in samples K1 and K2 which were associated directly with hominin specimen TM 1517, and which were compared with other (provenanced) samples at Kromdraai, with important results. Ongoing geochemical analyses will help to refine these extremely important results.

Herries criticises the use of a 'pay-to-publish publisher', in this case SUN PReSS (AFRICAN SUN MeDIA). The words 'format' or 'formatting' are used seven times in the review. For example, he says that the book is 'an edited volume of scientific articles packaged in an odd rectangular, coffee-table format'. What is wrong with a book being rectangular? Books have been rectangular for centuries. And what is the definition of being a coffee-table book? Herries states: 'It does make you wonder why the authors chose this format of publishing for research.' SUN PReSS was selected at the recommendation of the editor of the *South African Journal of Science*, and was

chosen because it provided an appealing way of publishing extensive peer-reviewed results in colour and *in detail* (emphasis added).

Herries states that 'many readers will not come away trusting the information presented'. Essentially he is saying that many readers will distrust the content of the book. This derogatory statement is particularly disturbing. Contrary to Herries' statements, Professor Bob Brain (in his Foreword of our volume) praises the 'great detail' in which research at Kromdraai is being conducted.

Herries mentions 'grammatical and spelling errors throughout' the book. One may teach lessons but 'never throw a stone and hide his hand'. Herries should thoroughly proofread his own papers and correct his incorrect spelling of Kromdraai (not 'Kromndraai', p. 22)[3].

To conclude, the Kromdraai book continues to be accessible through the reputable publisher, SUN PReSS, and a revised edition is planned in an electronic form to make the detailed studies accessible worldwide. Moreover, the KRP has already and will soon continue to supplement the results presented in the book.[7] In the near future, attention will be focused on sedimentology, taphonomy, chronology (both relative and absolute), faunal and hominin palaeobiology, taxonomy and phylogeny.

We hope that readers will decide for themselves whether the review by Herries[4] is an accurate reflection of the KRP which is truly exciting in terms of new discoveries of hominin and other fossils, within a controlled stratigraphic context.

Acknowledgements

We acknowledge the support of the National Research Foundation (South Africa); the Centre of Excellence for the Palaeosciences (South Africa); the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Union; the Centre National de la Recherche Scientifique (France); the Institut des Déserts et des Steppes; the French Ministry of Foreign Affairs; and the French Embassy in Pretoria.

References

1. Braga J, Thackeray JF, editors. Kromdraai: A birthplace of *Paranthropus* in the Cradle of Humankind, a South African heritage site. Stellenbosch: SUN MeDIA; 2016.
2. Partridge TC. Some preliminary observations on the stratigraphy and sedimentology of the Kromdraai B hominid site. Palaeoecol Afr Surround Isl. 1982;15:3e12.
3. Herries AIR. A multi-disciplinary seriation of early *Homo* and *Paranthropus* bearing palaeocaves in southern Africa. Quat Int. 2009;202:14–28. https://doi.org/10.1016/j.quaint.2008.05.017
4. Herries AIR. Kromdraai evolved, but poorly packaged. S Afr J Sci. 2018;114(1/2), Art. #a0241, 2 pages. https://doi.org/10.17159/sajs.2018/a0241
5. Braga J, Thackeray JF, Bruxelles L, Dumoncel J, Fourvel J-B. Stretching the time span of hominin evolution at Kromdraai (Gauteng, South Africa): Recent discoveries. CR Palevol. 2016;16:58–70. https://doi.org/10.1016/j.crpv.2016.03.003
6. Thackeray JF, McBride VA, Segonyane SP, Franklyn CB. Trace element analysis of breccias associated with the type specimen of *Australopithecus* (*Paranthropus*) *robustus* from Kromdraai. Ann Transv Mus. 2003;40:147–150.
7. Fouvel J-B. *Civettictis braini* nov. sp. (Mammalia: Carnivora), a new viverrid from the hominin-bearing site of Kromdraai (Gauteng, South Africa). CR Palevol. In press 2018. https://doi.org/10.1016/j.crpv.2017.11.005.

Four proposals for a more reliable scientific literature

AUTHOR:
Jason Bosch[1]

AFFILIATION:
[1]Gregor Mendel Institute, Vienna, Austria

CORRESPONDENCE TO:
Jason Bosch

EMAIL:
jason.bosch@gmi.oeaw.ac.at

KEYWORDS:
publishing; reproducibility; retractions; citation

HOW TO CITE:
Bosch J. Four proposals for a more reliable scientific literature. S Afr J Sci. 2018;114(3/4), Art. #a0257, 2 pages. http://dx.doi.org/10.17159/sajs.2018/a0257

PUBLISHED:
27 Mar. 2018

The publication of Richard Harris' new book *Rigor Mortis* once again brings the issue of poorly conducted science and irreproducible research into focus.[1] This book is the latest in a string of publications which point out flaws in the current scientific system. Earlier examples from the scientific literature include John Ioannidis' provocatively titled article 'Why most published research findings are false'[2] and the disclosure that researchers at Amgen could only confirm 6/53 landmark cancer studies[3]. These articles have suggested and inspired ways to improve the situation that address, among other things, incentives in science[4], guidelines for better experimental design[5] and AllTrials greater insistence on the registration and reporting of clinical trials. I think pre-registration would also benefit basic science.

I wish to advance a series of proposals regarding the way in which we publish scientific literature that, I believe, will benefit science by making research rapidly available, easier to search and more reliable.

Scientific articles should be short and focused

One should be able to grasp the gist of a paper from its title and abstract alone. However, in a long paper with multiple experiments, many pieces of relevant or interesting information can remain hidden. Often, entire experiments are heavily summarised; compressed into just a few paragraphs or sentences, which can hinder understanding. Over the past three decades, there has been a dramatic increase in the size of the average publishable unit, i.e. the number of pages, figures, references and authors per paper[6], which makes it more difficult for researchers to gain a true understanding of the paper.

A move to shorter papers will make it easier to identify and judge the contents of a paper and allow authors to clearly explain the experiment and its limitations without worrying about different messages competing for space. Articles would ideally cover a single experiment or a single, carefully defined question. For example, instead of writing a single paper in which one screens for a new molecule, characterises its behaviour and determines its structure, one should publish a separate paper for each step. Having multiple experimental approaches to answer a specific question will increase our confidence in the results but, if experiments appear to contradict one another, may lead to neither experiment being published or only the publication of experiments which favour the authors' hypothesis. Step-by-step publication may help avoid this problem.

I recognise that I am essentially promoting the least publishable unit – the smallest amount of data which can be successfully published – which has been criticised since the early 1980s.[7] Many of the arguments originally raised against it have already been addressed; diffuse responsibility and unnecessary authorships can both be counteracted by listing author contributions. Criticisms like the 'inflated' number of publications reflect problems with how scientists are evaluated. There are drawbacks to the least publishable unit which should be kept in mind, such as the same data series being used in multiple publications while appearing to be independent sets of data.[8] These issues can be addressed in other ways, such as publishing and referring back to a data set, and are not general enough to advise against the practice.

Shorter papers could significantly speed up scientific progress. With longer papers, it can be that the first of a series of experiments is completed months or even years before publication. It is quite likely that there are several years' worth of research on a particular topic, relevant to others' ongoing work, sitting unpublished in labs around the world. Rapid publication could prevent researchers from wasting time following dead ends that others have already tried or better inform their approach to a current problem. It also removes the risk that failure of a downstream experiment results in the collapse of the 'story', preventing the publication of earlier, valid results.

Short papers will prove more robust to retractions. A longer paper with more authors means there is a greater chance that one of the authors will do something which leads to a retraction. In such cases, this will have a negative impact on many authors and result in valid experiments being lost as collateral damage. Shorter publications would mitigate these issues.

Retractions and corrections should cascade through the literature

Publishers and reference manager software should take advantage of the digital landscape to cascade corrections and retractions. Almost all papers are published digitally and identified by unique tags such as the DOI (digital object identifier). These data can be used to automatically inform researchers of problems in the scientific literature. This is necessary because, despite the importance of such information, the current system does not make it obvious when papers have been retracted.

For example, take the paper by Hunter and Prüss-Ustün[9] which was published in October 2016 and retracted in May 2017. The web version of the paper links to the retraction notice and the new PDF makes it clear that the paper was retracted. However, the new PDF neither links to the retraction nor does the DOI entry offer such information. Although the Hunter and Prüss-Ustün paper remains cited after its retraction, this fact is neither indicated in previous reference lists nor is this information retrieved by Mendeley with the DOI. If someone had downloaded the paper in the months between publication and retraction, they could, quite easily, never become aware that it was retracted.

This issue could be avoided if reference managers could identify and flag retracted papers using information retrieved from the DOI. Furthermore, digital publishing allows us to follow citations through the literature. This means that papers which cite papers which have been, or later become, retracted could also be identified and flagged. This would inform readers that there are problematic references cited and warn them to look closely at

information coming from retracted or corrected sources. Together, this should help to prevent the spread of incorrect information and thereby increase our confidence in the scientific literature.

Journals should be required to publish replications of their articles

Although novelty is essential for science to advance, science builds on work that has come before; thus, replication is equally essential to ensure reliability. Therefore, we should not rely on a single publication and place an undue emphasis on novelty. This emphasis leads to absurd situations in which attention-grabbing work is published in a high-profile journal, while a failed replication of that same work is not considered because of a lack of novelty. This results in an asymmetry, in which novel but incorrect research can have a higher impact than less original but correct research. One unfortunate consequence of this asymmetry is the reluctance of scientists to do the important work of replicating previous studies.

I would like to propose that journals should have an obligation to publish scientifically sound replications of work that they have previously published. In addition, building on my previous point, replications should also be linked. Linking would allow readers to see whether someone has attempted to replicate a paper and the result. Scientists being aware that journals will publish replications should help address the problem in which negative results are seldom published, which is important because simulations have shown that publication of negative results is important to prevent incorrect results being accepted as fact.[10] My first three proposals combined would result in a much clearer view of the reliability of a specific piece of knowledge.

We should separate data-generating articles from storytelling articles

With an emphasis on short, focused papers and abandoning the idea of complete 'stories', how do we advance conceptually? The answer is by separating the scientific stories from their constituent parts. By complete stories, I mean a series of different but linked experiments which follow logically and build on one another to come to a combined conclusion all within a single paper. As the idea of publishing stories is currently widespread in science, I have no doubt that this proposal will be a controversial notion. Stories have limits, however, and it does not serve the interests of science for researchers to publish only when they believe that they can construct a story or to force results into a story before there is sufficient evidence to support one.

Short, focused articles – ideally linked to replications – will create blocks of data which can stand on their own. It is the function of review or 'story' papers to collect these blocks and combine them into a coherent narrative. The same data will be published as currently but the difference is that the data will come faster and enable the synthesis of up-to-date results from multiple labs instead of many narratives built on incomplete data.

Our narrative explanations of phenomena may change as new data become available, but the original data should remain valid regardless of the interpretation. This alone suggests that it might be wise to separate the data from the narrative as one is likely to remain valid much longer than the other.

The freedom to publish without the limitations of a story will open the way for sharing many more observations. There are journals being established which are supporting such approaches. *BMC Research Notes* publishes 'scientifically valid research outputs that cannot be considered as full research or methodology articles', and *Matters* is a journal where 'Stories can wait. Science can't.' In fact, in discussions with co-workers I have been told that this suggestion does not go far enough. Some believe that data papers could be entirely replaced by structured databases, leaving only review or outlook papers.

Conclusion

To address concerns about the accuracy and reproducibility of scientific publishing, I have presented a series of proposals which will improve the quality and reliability of scientific publications. Short publications will present data as building blocks which can be combined to form scientific narratives. By keeping papers focused, ensuring replications are published and dynamically linking replications and papers, we ensure that we treat results as pieces of data rather than individual stories. By cascading replications and corrections along a chain of citations, we can build a higher level of confidence in what is published. These proposals will require action from many different parties, but I believe that the benefits of this new system will outweigh the costs.

Acknowledgements

I thank my colleagues, particularly Angelika Czedik-Eysenberg, for discussions which led to this article. I also thank Angelika Czedik-Eysenberg, Nancy R. Gough and J. Matthew Watson for providing helpful feedback on this manuscript.

References

1. Harris R. Rigor mortis. New York: Basic Books; 2017.
2. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8), e124, 6 pages. http://dx.doi.org/10.1371/journal.pmed.0020124
3. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483:531–533. http://dx.doi.org/10.1038/483531a
4. Nosek BA, Spies JR, Motyl M. Scientific utopia II: Restructuring incentives and practices to promote truth over publishability. Perspect Psychol Sci. 2012;7:615–631. http://dx.doi.org/10.1177/1745691612459058
5. Masca N, Hensor E, Cornelius V, Buffa F, Marriott H, Eales J, et al. RIPOSTE: A framework for improving the design and analysis of laboratory-based research. eLife. 2015;4, e05519, 27 pages. http://dx.doi.org/10.7554/eLife.05519
6. Cordero RJB, De León-Rodriguez CM, Alvarado-Torres JK, Rodriguez AR, Casadevall A. Life science's average publishable unit (APU) has increased over the past two decades. PLoS ONE. 2016;11, e0156983, 14 pages. http://dx.doi.org/10.1371/journal.pone.0156983
7. Broad WJ. The publishing game: Getting more for less. Science. 1981;211:1137–1139. http://dx.doi.org/10.1126/science.7008199
8. Dupps WJ, Randleman JB. The perils of the least publishable unit. J Cataract Refract Surg. 2012;38:1517–1518. http://dx.doi.org/10.1016/j.jcrs.2012.07.020
9. Hunter PR, Prüss-Ustün A. Have we substantially underestimated the impact of improved sanitation coverage on child health? A generalized additive model panel analysis of global data on child mortality and malnutrition. PLoS ONE. 2016;11, e0164571, 17 pages. http://dx.doi.org/10.1371/journal.pone.0164571 [RETRACTED]
10. Nissen SB, Magidson T, Gross K, Bergstrom CT. Publication bias and the canonization of false facts. eLife. 2016;5, e21451, 19 pages. http://dx.doi.org/10.7554/eLife.21451

Diversity of participant representation within the 66th Lindau Nobel Laureate Meeting

AUTHORS:
Sphumelele C. Ndlovu[1,2]
Adriana Marais[3]
Promise T. Shabangu[4]
Luyanda L. Noto[5]
Johannes N. Greiner[6]
Nicholas E. Engelbrecht[7]
Asma Z. Dhunny[8]

AFFILIATIONS:
[1]Department of Air Quality Information, South African Weather Service, Pretoria, South Africa
[2]Space Geodesy Programme, Hartebeesthoek Radio Astronomy Observatory, Krugersdorp, South Africa
[3]Quantum Research Group, School of Chemistry and Physics, University of KwaZulu-Natal, Durban, South Africa
[4]International Relations and Cooperation, National Research Foundation, Pretoria, South Africa
[5]Physics Department, University of South Africa, Johannesburg, South Africa
[6]Third Institute of Physics, University of Stuttgart, Stuttgart, Germany
[7]Centre for Space Research, North-West University, Potchefstroom, South Africa
[8]Department of Physics, University of Mauritius, Moka, Mauritius

CORRESPONDENCE TO:
Sphumelele Ndlovu

EMAIL:
Sphumelele.Ndlovu@weathersa.co.za

KEYWORDS:
physics; research; Lindau Nobel Laureates; young scientists

HOW TO CITE:
Ndlovu SC, Marais A, Shabangu PT, Noto LL, Greiner JN, Engelbrecht NE, et al. Diversity of participant representation within the 66th Lindau Nobel Laureate Meeting. S Afr J Sci. 2018;114(3/4), Art. #a0263, 4 pages. http://dx.doi.org/10.17159/sajs.2018/a0263

The 66th Lindau Nobel Laureate Meeting (LiNo16) was dedicated to the field of physics. A total of 29 Nobel Laureates, 1 A.M. Turing Award recipient and 400 young scientists attended LiNo16 in Lindau, Germany. Young scientists from 80 countries attended the meeting. Out of the 400 young scientists at the meeting, only 30% were women scientists, and only one of the Nobel Laureate attendees was a woman.

Natural sciences have often been dominated by a single stereotype: white men from predominantly Western or developed countries. Although a lot is being done to increase human capacity in the sciences in the less developed and developing world, it has become apparent that there is a persistent lack of diversity in the sciences.[1,2] Extensive studies have also revealed that the rate of participation of women and minority groups in science, technology, engineering and mathematics (STEM) is significantly lower than the representation of women and minorities in society at large. As a strategy to improve women and minority representation, more countries are encouraging the involvement of these groups in science from a young age – a strategy which is the norm in many developed countries and which has been shown to have a positive correlation with high representativity of women and minority groups.

It is evident that global scientific collaborations encourage unity and inclusion regardless of power, race, beliefs and gender.[3,4] It has become imperative for society to be well capacitated in order to deal with global issues. The exclusion of certain groups based on gender and race, among other factors, means the game is played with less than half of the team. Diversity should be the basis for increased collaboration and not grounds for marginalisation. Nowotny et al.[5] describe diversity as a prominent theme in science and technology to determine technical processes, economic systems and social structures.

Since 1951, 350 Nobel Laureates have committed to the exchange among scientists with the aim of fostering education, inspiration and connection,[6] leveraging on diversity for increased scientific output as a result of collaboration and sharing of best practice experiences. This initiative, which is realised through annual meetings, has fostered and increased exchange amongst young scientists within their respective fields by exposure to a diversity of thinkers and new ideas. The meetings have encouraged global engagement on the unprecedented scale of the global problems we face today. Furthermore, the meetings enable the exploration and encouragement of the scientific diversity that lies in the world and the potential for addressing global issues.

We analysed diversity in terms of representation of gender and countries of origin at the 66th Lindau Nobel Laureate Meeting. Additionally, we ponder here on the lessons that came with the meeting. It is expected that our analysis will help to sensitise, motivate and improve the number of women and underrepresented regions with respect to participation in STEM meetings and forums. A diversity of attendees ensures a balanced benefit from the lessons that are acquired during such meetings. Through diversity, we believe that current scientific global challenges can be re-evaluated and innovations towards solutions developed more objectively, independent of gender, beliefs and race bias.

The Nobel Prize: A background

The Lindau Nobel Laureate Meetings were formed 66 years ago to contribute to reconciliation among the countries that were involved in World War II.[7] The initial aim of the meetings was to foster a peaceful and prosperous future for all, particularly in Europe. Current global challenges are not limited to interstate wars but include issues around transformation, women's empowerment, diversity and new issues around the unprecedented development taking place in science, technology and innovation. The Lindau Nobel Laureate Meetings have always been about science and bringing together Nobel Laureates and young scientists. During these meetings, young scientists are provided with a chance to interact, learn and be inspired to create a better future.

Taken from the archives of the Lindau Nobel Laureate Meetings,[8] 30% of Nobel recipients who attended the Lindau Nobel Laureate Meetings in the past 65 years have been from the discipline of physics (Table 1). This figure represents the largest number of attendees, followed by the discipline of chemistry with 26.6%.[8] Only three Nobel Laureates had connections with Africa, either by birth or relocation. Africans are therefore considered to be a minority when it comes to representation in these prestigious meetings. Given that Africa is expected to account for more than half of the world's population growth between 2015 and 2050,[9] the urgency and importance of incorporating young scientists from minority regions such as Africa is an imperative. One of the five women to have won a Nobel Prize in natural science, Marie Curie, has been honoured twice with a Nobel Prize. Her first Nobel Prize was in Physics in 1903. She shared the prize with Antoine Henri Becquerel and Pierre Curie. It was in recognition of the extraordinary services she rendered by her joint research with Pierre Curie on the radiation phenomenon discovered by Professor Henri Becquerel.[10]

The data presented in Table 1 indicate the total number of Nobel Laureates who participated in past Nobel Laureate meetings. These data include the majority of Nobel Laureates who have participated in the meetings since 1951. The few women and minorities among the Nobel Laureates who were part of the meetings heavily depended on the few women and minorities with Nobel Prizes in science. This raises a concern with regard to the number of female and minority participants in future science meetings. Would an increased number of female and minority participants in future science meetings boost or encourage these groups to be involved in breakthrough science

and hence obtain Nobel Prizes? This question remains to be answered through an experiment waiting to be conducted by the Lindau Nobel Laureate committee.

Table 1: The total number of Nobel Laureates per discipline who attended the Nobel Laureate Meetings in the past 65 years

Discipline	Number of Nobel Laureates	Percentage (%)
Chemistry	123	26.60
Interdisciplinary	1	0.22
Peace	9	1.95
Physiology/Medicine	132	28.57
Economic Sciences	56	12.12
Literature	2	0.43
Physics	139	30.00

Diversity of LiNo16

We analysed data on diversity within the meetings using the total number of participants in the 66th Lindau Nobel Laureate Meeting. The data used were obtained from the meeting handbooks.[7] The LiNo16 handbook contained all the relevant information on the participants. The handbook groups participants according to their country of origin and their affiliated countries. A world map is provided in Figure 1 which indicates the total number of participants (by nationality) at the meeting. The total number of participants was dominated by young men (70%). Africa, as a continent, was only represented by a total number of 25 participants, when compared to 41 and 89 participants from the USA and Germany, respectively. This indicates that women and African participants were under-represented during the 66th Lindau Nobel Laureate Meeting.

The lack of diversity was also observed from the total number of Nobel Laureates who attended the 66th Lindau Nobel Laureate Meeting. In total, 10 countries were represented and the USA had the majority of attendees. The USA is ranked top for all prizes except Literature, where France, Germany and the United Kingdom perform better.[11] Israel was the only country with a female Nobel Laureate who was present at the meeting. However, as she is the only woman alive to have won a Nobel

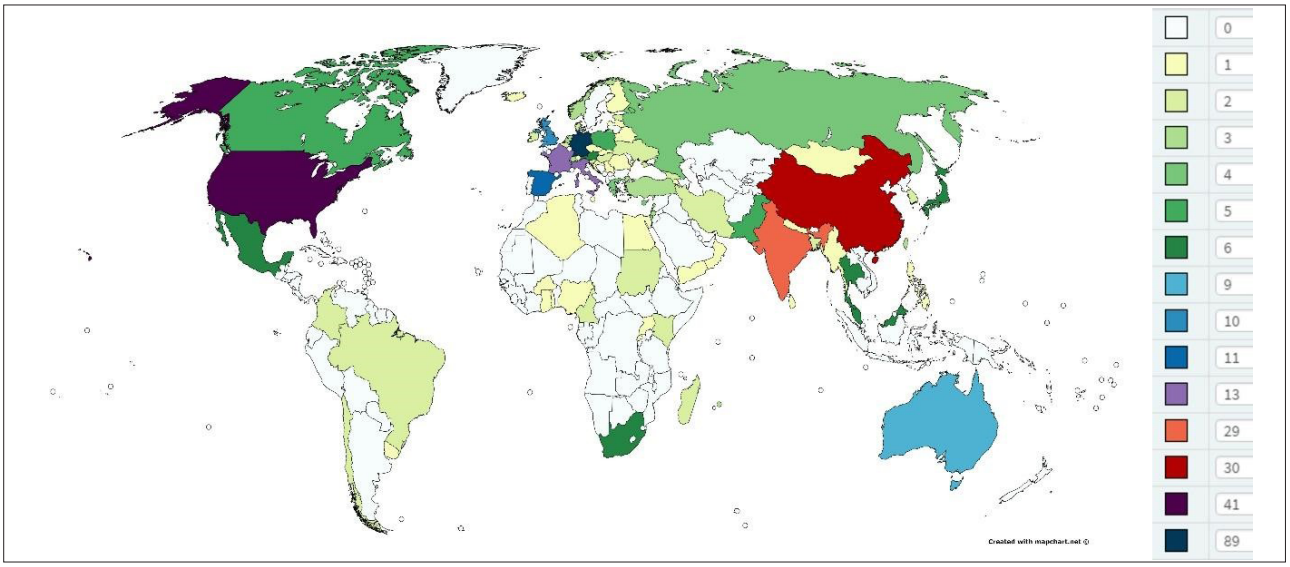


Figure 1: The 80 countries that were represented at the 66th Lindau Nobel Laureate Meeting. The total number of young participants is indicated by the colour and number on the chart to the right of the figure.

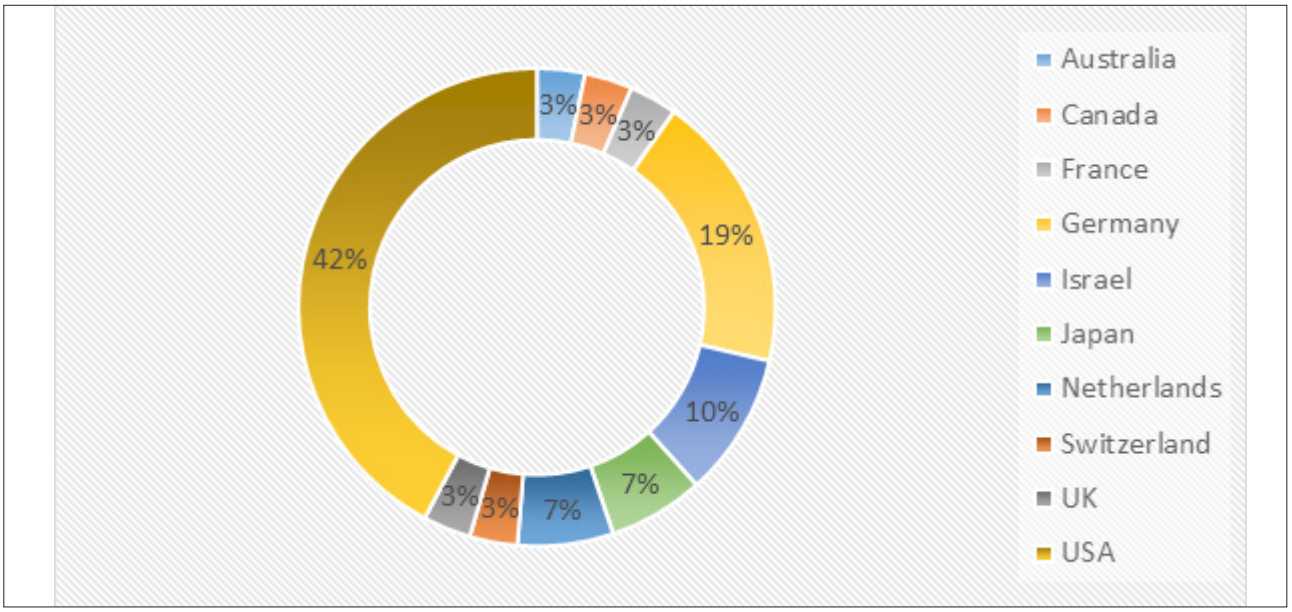


Figure 2: Graphical representation of the total number of Nobel Laureates (by country, %) who attended the 66th Lindau Nobel Laureate Meeting.

Prize in the Natural Sciences, this speaks more to the lack of diversity in the Nobel Prize awards themselves than the meeting. Countries such as the USA and Germany with their large numbers of Nobel Laureates (more than 15% each; Figure 2) were represented by only men.[4]

Analysis

It is evident from the data on the number of Nobel Laureates and young scientists who participated in the 66th Lindau Nobel Laureate Meeting (Table 2), that countries with Nobel Laureates who attended the meeting had a higher number of young scientist participants. This was also verified by the data on individual governments' expenditure on education. It was also observed that most countries with higher expenditure on education (primary to tertiary) were mostly represented at the meeting (Table 3). The young scientist attendees from Germany (and possibly from the other EU countries) could also be strongly linked to the fact that they had a greater awareness of the meeting, which is, after all hosted in Germany. Countries like China and India had a total number of young participants of 29 and 30, respectively. Both these countries have a population of more than 1 billion people; the number of their young scientists at the meeting could therefore be considered relatively inadequate.

A total of 10 countries which were represented by their Nobel Laureates had a total of 46% young scientists who participated in the meeting. This number of participants was almost half of the total number of young scientists at the meeting yet they comprise only 12% of countries that were represented in the meeting. Of the 10 countries that were represented at the 66th Lindau Nobel Laureate Meeting, only Sweden, Russia and Italy were absent. These three countries fall under the top 10 countries with the most number of Nobel Laureates in the fields of Peace, Literature, Economics, Chemistry, Physics and Physiology or Medicine. The increase in the number of US Nobel Laureates only began after World War II – before which the European nations were much more successful.[12] The dominance of Western countries is because of the huge socio-political power these nations have held over the last century as well as their bigger investment in STEM research when compared to their developing world counterparts.

Of the 400 young scientists, 30 were selected by the Lindau Nobel Laureate Council to present their work as posters during the meeting. Only 26.7% of these presenters were women; the majority originated from or were studying in developed countries (Table 4). The highest number of poster presenters originated from and/or were studying in Germany, followed by the USA. China, Norway, Thailand and Slovenia were also each represented by one young scientist. No data were available on the total number of young scientists who submitted abstracts for poster presentations but were unsuccessful.

Table 2: The relationship between the countries of origin of the Nobel Laureates and young scientists who participated in the 66th Lindau Nobel Laureate Meeting

Country	Nobel Laureates	Young scientists
Australia	1	9
Canada	1	5
France	1	13
Germany	6	89
Israel	3	4
Japan	2	6
Netherlands	2	3
Switzerland	1	3
UK	1	10
USA	13	41
Total	31	183

Table 3: Government expenditure on education (GEE; in % of GDP, 1980–2009) of countries represented at the LiNo16 meeting

Country	GEE (GDP %)	Year	Country	GEE (GDP %)	Year
Denmark	7.8	2007	Germany	4.5	2007
Iceland	7.4	2007	Hong Kong	4.5	2009
Kenya	7.0	2006	Uruguay	4.5	2011
Norway	6.8	2007	Algeria	4.3	2008
Malta	6.4	2007	China	4.2	2014
South Africa	6.0	2013	Italy	4.3	2007
Belgium	6.0	2007	Romania	4.3	2007
Finland	5.9	2007	Spain	4.3	2007
Israel	5.9	2007	Czech Republic	4.2	2007
Slovenia	5.7	2006	Cyprus	4.1	2007
France	5.6	2007	Malaysia	4.1	2008
Mongolia	5.6	2009	Rwanda	4.1	2008
United Kingdom	5.5	2007	Thailand	4.1	2009
USA	5.5	2007	Chile	4.0	2008
Austria	5.4	2007	Greece	4.0	2005
Ghana	5.4	2005	Oman	3.9	2006
Netherlands	5.3	2007	Russian Federation	3.9	2006
Ukraine	5.3	2007	Egypt	3.8	2008
Hungary	5.2	2007	Cameroon	3.7	2009
Switzerland	5.2	2007	Slovakia	3.6	2007
Yemen	5.2	2008	Japan	3.5	2007
Brazil	5.1	2007	Georgia	3.2	2009
Latvia	5.0	2007	Mauritius	3.2	2009
Canada	4.9	2007	Uganda	3.2	2009
Ireland	4.9	2007	India	4.9	2014
Poland	4.9	2007	Armenia	3.0	2007
Colombia	4.8	2009	Madagascar	3.0	2009
Estonia	4.8	2007	Sudan	28	1985
Mexico	4.8	2007	Philippines	2.8	2008
Iran	4.7	2009	Pakistan	2.7	2009
Serbia	4.7	2008	Sri Lanka	2.6	1985
Burkina Faso	4.6	2007	Turkey	2.9	2005
Croatia	4.6	2009	Bangladesh	2.4	2008
Nepal	4.6	2009	Liechtenstein	2.0	2007
Togo	4.6	2009	Nigeria	0.8	1995
Australia	4.5	2007	Myanmar	0.6	2000
Belarus	4.5	2009	Palestinian Territory	N/A	N/A

Table 4: The data (by gender) on the total number of young scientists who presented posters of their work during the meeting

Discipline	Women	Men
Atomic, Molecular, Optical and Fundamental Measurements	1	2
Quantum Optics, Quantum Information and Photonics	1	4
Astronomy, Astrophysics and Cosmology	1	5
Bio-Physics	1	2
Solid-State, Materials Science and Condensed Matter Physics	2	6
High Energy Physics	2	3
Total	8	22

Conclusion

We believe that the benefits of attending the 66th Lindau Nobel Laureate Meeting are significant, and that scientific development can be improved by increased diversity of participation. Young female scientists were under-represented but slightly better represented than the female Nobel Laureates. It was also observed that there are regions that were very under-represented, both in terms of the number of attendees and the number of young scientists who were given an opportunity to present their work. It was also observed that there was a correlation between the number of Nobel Laureates and young scientists who attended the 66th Lindau Nobel Laureate Meeting. Representatives from developed countries dominated in the meeting because of the socio-political power these nations have held over the last century. Underdeveloped countries, which were mostly under-represented at the meeting, are spending far less on education as a proportion of their GDPs[13] when compared to the government expenditure on education by developed countries.

Acknowledgements

The Academy of Science of South Africa and the South African Department of Science and Technology funded the South African delegates to attend the 66th Lindau Nobel Laureate Meeting in Germany.

References

1. Dirks C, Cunningham M. Enhancing diversity in science: is teaching science process skills the answer? CBE Life Sci Educ. 2006;5:218–226. https://doi.org/10.1187/cbe.05-10-0121
2. Espinosa L. Pipelines and pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence. Harv Educ Rev. 2011;81(2):209. https://doi.org/10.17763/haer.81.2.92315ww157656k3u
3. Guterl F. Diversity in science: Where are the data? Sci Am. 2014 October 01. https://doi.org/10.1038/scientificamerican1014-40
4. Banks JA, Banks CAM, Cortes CE, Hahn CL, Merryfield MM, Moodley KA, et al. Democracy and diversity. Principles and concepts for educating citizens in a global age [document on the Internet]. c2006 [cited 2016 Oct 30]. Available from: http://depts.washington.edu/centerme/DemDiv.pdf
5. Nowotny H, Scott P, Gibbons M. Re-thinking science: Knowledge and the public in an age of uncertainty. Cambridge: Polity Press; 2001.
6. 66th Lindau Nobel Laureate Meeting Programme; 2016 June 26 – July 01; Lindau, Germany.
7. Lindau Nobel Laureate Meeting Mediatheque [document on the Internet]. c2016 [cited 2016 Oct 31]. Available from: http://www.mediatheque.lindau-nobel.org/
8. Lindau Nobel Laureate Meeting participants directory; 2016 June 26 – July 01; Lindau, Germany.
9. World Bank. Expenditure on education, public (% of GDP) [document on the Internet]. c2013 [cited 2016 Oct 31]. Available from: http://hdr.undp.org/en/content/expenditure-education-public-gdp
10. The Nobel Prize in Physics [document on the Internet]. No date [cited 2016 Oct 31]. Available from: https://www.nobelprize.org/nobel_prizes/physics/laureates/1903/
11. Nobel Prize Facts [document on the Internet]. No date [cited 2016 Oct 01]. Available from: https://www.nobelprize.org/nobel_prizes/facts/
12. Kirk A. Nobel Prize winners: Which country has the most Nobel laureates? The Telegraph. 2015 October 12. Available from: https://www.telegraph.co.uk/news/worldnews/northamerica/usa/11926364/Nobel-Prize-winners-Which-country-has-the-most-Nobel-laureates.html
13. United Nations. World population projected to reach 9.7 billion by 2050 [document on the Internet]. c2015 [cited 2016 Oct 01]. Available from: http://www.un.org/en/development/desa/news/population/2015-report.html

AUTHORS:
Rosemary A. Dorrington[1,2]
Amanda T. Lombard[3]
Thomas G. Bornman[4,5]
Janine B. Adams[5]
Hayley C. Cawthra[6,7]
Shaun H.P. Deyzel[4,8]
Wayne S. Goschen[4]
Kenneth Liu[9]
Jacques Mahler-Coetzee[10]
Gwynneth F. Matcher[2,11]
Christopher McQuaid[12]
Shirley Parker-Nance[4,11]
Angus Paterson[2]
Renzo Perissinotto[13]
Francesca Porri[2,14]
Michael Roberts[15]
Bernadette Snow[16]
Patrick Vrancken[17]

AFFILIATIONS:
[1]DST/NRF SARChI Chair in Marine Natural Products, Department of Biochemistry, Microbiology and Biotechnology, Rhodes University, Grahamstown, South Africa
[2]South African Institute for Aquatic Biodiversity (SAIAB), Grahamstown, South Africa
[3]DST/NRF SARChI Chair in Marine Spatial Planning, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa
[4]South African Environmental Observation Network, Elwandle Coastal Node, Port Elizabeth, South Africa
[5]Department of Botany, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa
[6]Marine Geoscience Unit, Council for Geoscience, Bellville, South Africa
[7]Centre for Coastal Palaeoscience, Nelson Mandela University, Port Elizabeth, South Africa
[8]Department of Zoology, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa
[9]Department of Geology, University of Fort Hare, Alice, South Africa

Working together for our oceans: A marine spatial plan for Algoa Bay, South Africa

Southern Africa occupies a critical position within the southern hemisphere for the study of broadscale global change and the three oceans around South Africa (the Atlantic, Indian and Southern Oceans) play a vital role in determining local and regional climate and weather patterns. Oceans and coasts also provide various resources and services (e.g. food and carbon sequestration), but these services are threatened by human activities. Uncertainty of the impact and consequences of these anthropogenic activities makes it problematic to manage marine resources. Given the recent global emphasis on the development of 'ocean economies', the exploitation of living (fisheries, aquaculture and tourism) and non-living (oil and gas, minerals, energy) marine resources should be on a scale that is socially and economically justifiable and ecologically sustainable.

In 2014, 'Operation Phakisa' was launched in South Africa as an initiative to accelerate execution of the National Development Plan.[1] The primary focus of Phakisa is to unlock the economic potential of South Africa's oceans.[2] This will be achieved through the 'implementation of an overarching, integrated ocean governance framework for sustainable growth of the ocean economy that will maximise socio-economic benefits while ensuring adequate ocean environmental protection' by 2019. Marine spatial planning (MSP) is a key component of this integrated governance framework, and the development of MSP legislation during 2016 was prioritised as 'critical' to achieving the Operation Phakisa objectives. Accordingly, the Department of Environmental Affairs (DEA) published the Marine Spatial Planning Bill (2017)[3] 'to provide a framework for marine spatial planning in South Africa's waters; to provide for the development of the marine spatial plan; to provide for institutional arrangements for the implementation of the marine spatial plan and governance of the use of the ocean by multiple sectors; and to provide for matters connected therewith'.

A Marine Spatial Plan for the South African maritime domain

The MSP Bill provides for a multi-stage approach to develop a Marine Spatial Plan for the South African maritime domain, which may be divided into four bio-geographic marine areas: the East Coast, the South-East Coast, the West Coast and the Prince Edward Islands. According to the MSP Framework,[4] South Africa's marine area plans will be prepared sequentially. This framework will allow effort to be focused on one marine area at a time and the consequent experience gained from preparing each plan will be used to refine and improve the preparation (or review) of other plans. To determine the order in which the plans will be prepared, an initial assessment of the priorities for each marine area is required, taking into account the cross-sectoral state of knowledge of system processes and functioning, environmental pressures, human uses and development opportunities. A critical component of the MSP process will be a national data- and information-gathering exercise, in which existing knowledge on social-ecological systems and their interactions will be consolidated. Given that this will be South Africa's first experience of conducting MSP, the first plan will allow the MSP process to be fine-tuned for application in subsequent plans.

Algoa Bay as a case study for the first South African Marine Area Plan

Algoa Bay (Figure 1) is an ideal planning region for a case study, given the substantial body of biophysical data that exists for the area. Research on the physical oceanography began in the 1980s with short-term, fine-scale projects, which led to an increase in knowledge of the Bay system.[5-9] The establishment of the South African Environmental Observation Network (SAEON) Algoa Bay Sentinel Site for Long-Term Ecological Research in 2007 has led to the collection of a wealth of data[10-12] that has resulted in Algoa Bay being considered the best-monitored coastal area in Africa and in the southern hemisphere. The diversity of habitats, oceanographic processes and the socio-economic reliance on the marine environment in Algoa Bay has provided dynamic natural laboratories for conducting multidisciplinary, multi-institutional field-based research that can be applied to other regions.[13] Building on this solid biophysical base, we intend to generate new knowledge on the social-ecological systems in the Bay. Although our planning products will be focused on Algoa Bay, social-ecological systems extend well beyond the Bay (e.g. larval transport, sea-level rise, fisheries benefits).

The distribution and ecology of the large fauna of Algoa Bay have been extensively studied since the first description of bird species around the Bird Island complex in 1946 by Courtenay-Latimer and Gibson-Hill[14]. The focus of research has been mostly on top predators[15-21], pelagic fish species[22-28], squid[29], and the impact of global change on the local fisheries[30]. Knowledge of the diversity of subtidal benthic macrofauna is less extensive and focused research is more sporadic and mostly limited to observational data obtained by SCUBA at depths of less than 30 m during collection of invertebrate specimens (sponges, tunicates, soft corals) for marine natural product chemistry.[31-34] Recent data collected by a remotely operated vehicle through the African Coelacanth Ecosystem Programme suggest that the Bay may support extraordinary invertebrate diversity including numerous novel, undescribed taxa (Parker-Nance and Dorrington, unpublished data).

Information on the biology and ecological dynamics within the shallow subtidal and intertidal zones has two focus areas: sandy beach[35-36] and rocky shore ecosystems. Research in both systems has targeted resources used for human consumption and, in the case of rocky shores, also biological invasions.[37] These studies mostly have provided information on the genetic and spatial structure of marine assemblages,[38,39] and on the early life stage dynamics that characterise the Bay in comparison to the adjacent stretch of open coast. The structure and function of marine species in the Bay differ from those in the adjacent open coast[40] such that the dynamics of stock replenishment differ between the two[41]. Recent developments also suggest the importance of the Bay as an area of larval retention, although behavioural and oceanographic features can further shape population connectivity within this complex system.[42,43]

[10]Nelson Mandela School of Law, University of Fort Hare, East London, South Africa

[11]Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, South Africa

[12]DST/NRF SARChI Chair in Marine Ecosystem Research, Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa

[13]DST/NRF SARChI Chair in Shallow Water Ecosystems, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa

[14]Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa

[15]Department of Oceanography, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa

[16]Department of Development Studies, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa

[17]DST/NRF SARChI Chair in the Law of the Sea and Development in Africa, Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Rosemary Dorrington

EMAIL:
r.dorrington@ru.ac.za

KEYWORDS:
community of practice; systematic biodiversity plan; oceanography; social-ecological system; system dynamics; marine law; ocean governance

HOW TO CITE:
Dorrington RA, Lombard AT, Bornman TG, Adams JB, Cawthra HC, Deyzel SHP, et al. Working together for our oceans: A marine spatial plan for Algoa Bay, South Africa. S Afr J Sci. 2018;114(3/4), Art. #a0247, 6 pages. http://dx.doi.org/10.17159/sajs.2018/a0247

PUBLISHED:
27 Mar. 2018

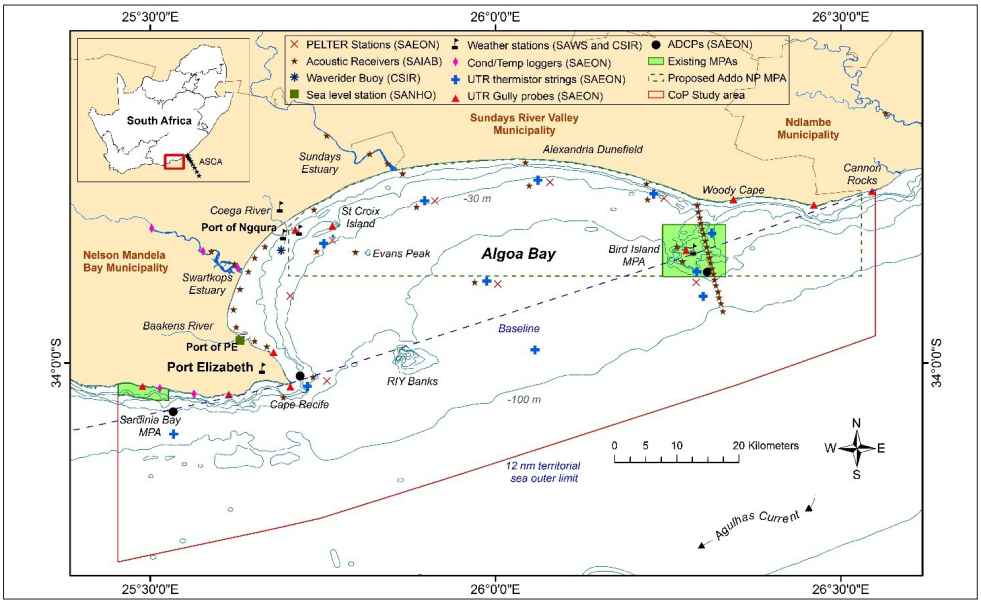


Figure 1: Map of the Algoa Bay Project study area.

Other studies have focused on the diversity of the phytoplankton microbial communities in the Algoa Bay system including diatoms and bacterial communities[34,44-46], but knowledge of the true extent of microbial diversity in the Bay and elsewhere along the South African coast is very limited[47,48]. There have been a number of studies on the Sundays and Swartkops (the two largest estuaries that enter the Bay), including a recent investigation of estuarine microbial communities.[49] A preliminary study was undertaken on the sedimentary deposits of the Algoa Bay seafloor in the 1970s.[50]

The legal regime in the Bay is dominated by the *Maritime Zones Act, 1994* (*Act 15 of 1994*). The Bay has the status of internal waters and because the waters are part of the territory of South Africa, all domestic legislation must be assumed to apply including, for instance, the *Marine Traffic Act, 1981* (*Act 2 of 1981*), the *National Ports Act, 2005* (*Act 12 of 2005*), the *Marine Living Resources Act, 1998 (Act 18 of 1998)*, *the National Environmental Management Act (NEMA), 1998 (Act 107 of 1998)* and the significant amount of environmental legislation that stems from the NEMA, for example, the *Integrated Coastal Management Act, 2008 (Act 24 of 2008*). The development of the Marine Spatial Plan also needs to take place in accordance with any future MSP legislation (The MSP Bill, 2017[3]). Adding to the legal complexity is the need to take into account the scope of the powers and duties of the provincial legislative and executive bodies and the adjacent coastal municipalities e.g. the Nelson Mandela Bay Metropolitan Municipality. Also relevant is legislation impacting on relevant coastal activities on land including, for instance, the *Spatial Planning and Land Use Management Act, 2013 (Act 16 of 2013)*, as well as maritime safety and security legislation, and their application along the adjacent stretch of coast. The maritime and coastal surface space also articulates with both that which is below the water and that which is above it. Consequently, legislative provisions governing air space (such as the *Civil Aviation Act 13 of 2009*), as well as defined aspects of the coastal and sub-aqua space (such as underwater cultural heritage under the *National Heritage Resources Act 25 of 1999*), must be considered and integrated into marine governance frameworks including MSP. We propose here that Algoa Bay should be used as a case study to develop appropriate methods and products to inform the first regional marine area plans. The Bay is an appropriate pilot site because it includes complex scales of governance, human use and biophysical environments, and MSP for this area will likely be scalable to broader areas.

The 'Algoa Bay Project'

In September 2017, we established a South African Research Chair Initiative (SARChI)-led community of practice, including 6 SARChI Professors and 12 senior researchers representing three universities (Rhodes University, Nelson Mandela University and the University of Fort Hare) and three national research institutions – SAEON, the South African Institute for Aquatic Biodiversity (SAIAB), both located within the Eastern Cape, and the Council for Geoscience Marine Geoscience Unit. All members of the new community of practice are actively involved in research in the broad field of marine science, with many studies focusing on the Algoa Bay system. These researchers, together with their research groups and collaborators, provide the transdisciplinary consortium with the critical mass required for the Algoa Bay Project (Figure 2).

MSP is dependent on understanding and managing the complexity of legal and socio-economic requirements on the one hand and environmental considerations (including physical, chemical and biological factors) on the other. Broadly, the Algoa Bay Project is based on three fundamental pillars: (1) a bioregional plan that includes a detailed biophysical map of the region together with data on biodiversity patterns and processes of the system; (2) a governance framework that integrates the complex legal regimes that govern activities in the region, including local, provincial, national and international legal instruments, and (3) a socio-economic plan that addresses stakeholder needs, human use, impacts on ocean and human health, economic value of ocean resources and activities, the valuation of ecosystem services, and the mapping of ecological infrastructure (Figure 2).

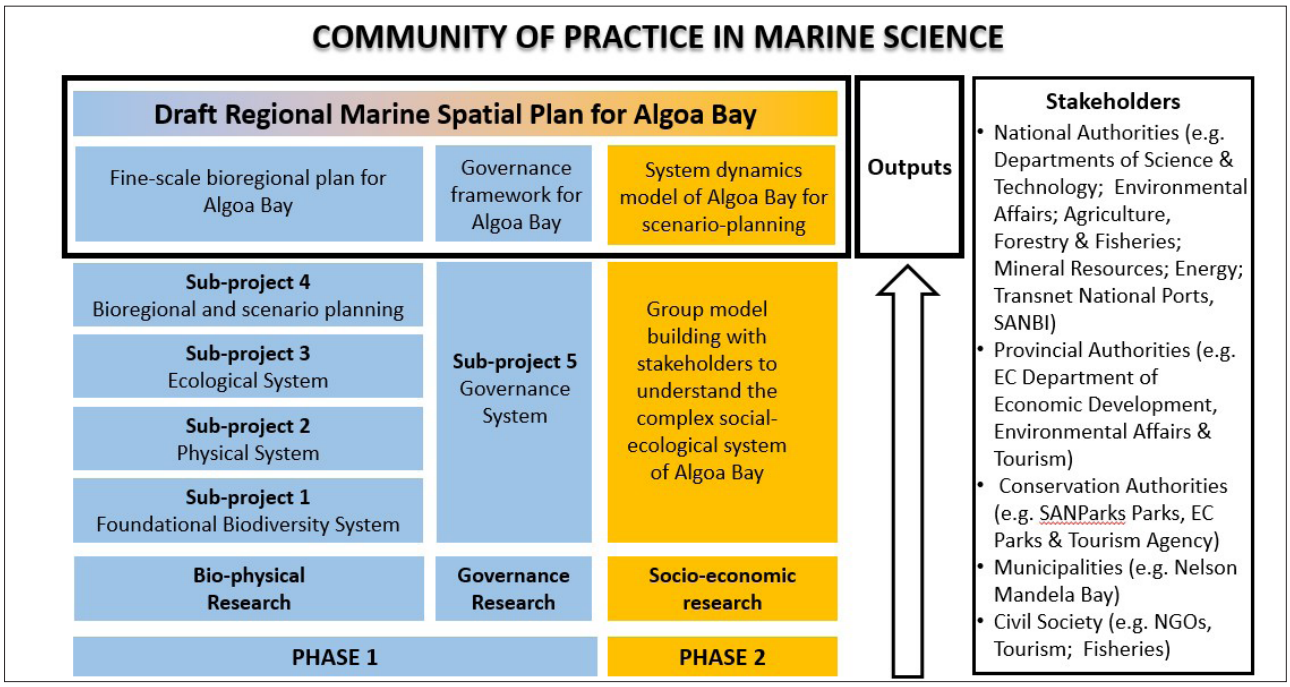


Figure 2: Research plan for the Algoa Bay Project.

In the first phase of the Algoa Bay Project, we will focus on the development of a bioregional plan (using systematic biodiversity planning methods[51-53]) and a governance framework, while the second phase will be the development of a socio-economic plan, with the ultimate goal being an integrated marine spatial plan.

To embrace novelty, creativity and meaningful transdisciplinarity in this study, we plan to address a major knowledge gap – the development of a system dynamics modelling approach to MSP – within the Algoa Bay social-ecological system. The biophysical, socio-economic and legal systems within the Bay are complex,[54] and our approach will be to understand subsystems, and then link them to one another in an integrated system dynamics model. Such a (causal-loop) model will define the relationships between factors in both the coastal and ocean domains across the traditional shoreline boundary, and will allow scenario planning to evaluate the outcomes of different management and global change scenarios. All three pillars of the project (bioregional, governance and socio-economic) will be combined in the final MSP model allowing users (stakeholders) to understand the causal relationships between factors, and to model the impacts of changes in any factor on the system. For example, one could model the impacts of introducing a new ocean zoning system on the total catch of the small pelagic fishery; or the impact of an increase in fishing effort on the prey base of penguins.

A fine-scale bioregional plan and spatial data for Algoa Bay

A critical component of a bioregional plan is the fundamental knowledge of the extent and distribution of biodiversity, the ecosystem processes that sustain this biodiversity, and the anthropogenic factors that use and threaten it. While South African marine biota comprises more than 13 000 species, the majority of which are large bodied, knowledge of microbial (viruses, archaea, bacteria, fungi) and protozoan diversity is considered to be poor to non-existent. 'Moderate knowledge' of benthic macrofauna diversity is confined mainly to species collected from the intertidal and shallow coastal waters.[35,55] This knowledge gap will be addressed for the Algoa Bay system by providing high-resolution information on the diversity of microbial communities (archaea, bacteria and phytoplankton) and macrofauna (e.g. sponges, tunicates and bryozoans) as well as invertebrate and fish larvae in habitats extending from the intertidal to shallow coastal waters and the continental shelf. The result will be a fine-scale biodiversity map of the Algoa Bay system (Figure 1) that will be used to assess the effectiveness of existing and proposed Marine Protected Areas, and to recommend how knowledge gaps can be addressed. The study will also investigate the extent and intensity of anthropogenic impacts (e.g. aquaculture, shipping and used-water inputs) in the Bay, and produce a human-use map for the bioregional planning and systems modelling exercises. This work will build on the broadscale marine protected area planning already undertaken by the South African National Biodiversity Institute (SANBI)[56] and South African National Parks (SANParks) in the design of the proposed Addo Elephant National Park Marine Protected Area.

Current knowledge of the oceanography of Algoa Bay is insufficient to fully understand this complex system, and many questions still need to be answered. The multidisciplinary Algoa Bay Project seeks to provide a broad overview of the system by integrating existing data on the oceanography, bathymetry, coastal and marine geology and geomorphology. Knowledge gaps will be filled by targeted studies making use of research platforms available through SAIAB and SAEON. Areas of special interest include upwelling, Agulhas Current influences, swell, long-period waves and internal waves, storm surges, current dynamics of open bays and around capes, thermohaline structures and dynamics, physical and biological properties of fronts, rips, palaeo sea-level calculations and projected future and mega rips, sea-level rise, wind and weather patterns, bathymetric and marine geophysical seafloor mapping, and sediment dynamics and transport.

Also important to the development of a marine bioregional plan is the fundamental understanding of biodiversity movement patterns and ecosystem functioning. Algoa Bay is uniquely influenced by oceanographic features including warm tropical waters from the Agulhas Current and upwelling of cool, nutrient rich bottom waters. In addition, fresh water is fed into the Bay by rivers (largely the Swartkops and Sundays) as well as being fringed by the Alexandria Dunefield (Figure 1). These features, together with the physical geography, combine to provide habitats for a diversity of benthic invertebrate species (Figure 3). Ecosystems of special interest in the region include permanently and temporarily open/closed estuaries, rocky shores, islands, sandy beaches and coastal dunefields, coastal wetlands, peritidal tufa stromatolites as well as the pelagic and benthic ecosystem of the Bay and continental shelf.

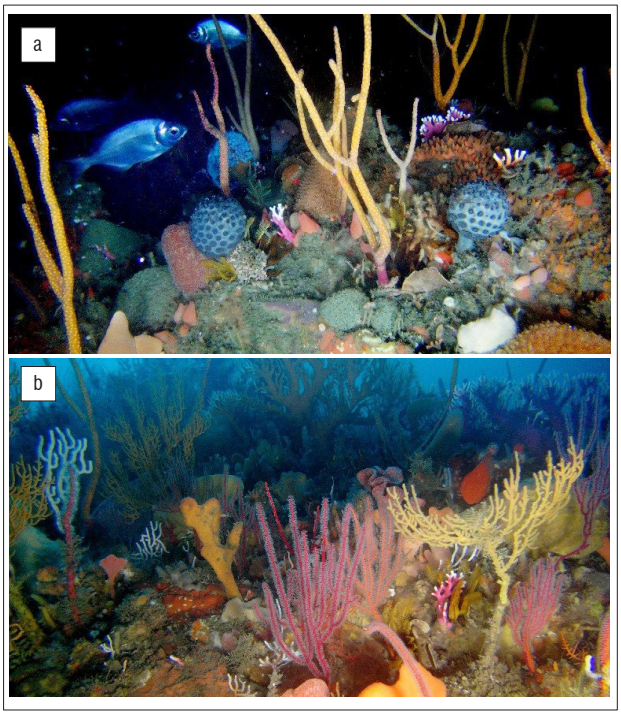


Figure 3: Algoa Bay is extraordinarily rich in invertebrate biodiversity. These images of benthic habitats were collected by a remotely operated vehicle at a depth of 50 m from (a) Evan's Peak reef and (b) Riy Banks reef.

The aims of the project are: (1) to map patterns of biodiversity distribution, abundance and composition of Algoa Bay; (2) to identify drivers of habitat segregation, species dispersal and connectivity and (3) to model interactions between indigenous and alien species. Research will focus on the dynamics of changing community structures and ecosystem processes in response to climate change and human-induced factors, such as alien species introductions, land-use change, residential/industrial development and associated pollution. The expected outcomes should include an updated, state-of-the-art assessment of each ecosystem, with high-resolution data across multiple scales of taxa.

The process of developing a marine bioregional plan will proceed in parallel, integrating existing and new knowledge as it becomes available from the biophysical mapping studies. We will apply best-practice systematic planning approaches using available data on biophysical patterns and processes. Data will be sourced from sub-projects 1–3 (Figure 2), spanning all habitats (intertidal to coastal to pelagic) and all levels of biodiversity (genes to ecosystems). Products will include a bioregional plan, but the collation of all spatial information in one format will provide added value. Once spatially explicit socio-economic data become available (Phase 2), the bioregional plan can be iteratively rerun, allowing the quantitative evaluation of trade-offs between human and environmental needs should they conflict, and to develop a transparent and sustainable marine zoning for the Bay. Novel approaches will include cross-realm planning (land to sea), the incorporation of tracking (animal movement), connectivity (e.g. larval transport) and risk (e.g. sea level rise, anthropogenic pollution) data into the plan, and scenario planning within a system dynamics modelling framework.

Legal landscape and development of legislative framework for Algoa Bay

The development of a Marine Spatial Plan for Algoa Bay must take place against the background of the complex legal regime governing activities in Algoa Bay (*The Maritime Zones Act, 1994 [Act 15 of 1994]*) with respect to internal waters and the territorial sea. Because the waters are part of the territory of South Africa, all domestic legislation applies. The *National Environmental Management: Integrated Coastal Management Act, 2008 (Act 24 of 2008*) requires that Algoa Bay be managed in an integrated manner. Research will focus on a systematic evaluation of the consequences of the enactment of this Act on legislation and related activities in a specific geographic area. The development of the Marine Spatial Plan will need to take into account the legislative developments regarding the MSP Bill.[3] Because the new Act will be the first of its kind in South Africa, a crucial component of this sub-project is to critically discuss the provisions of the Bill, both within the whole internal economy of the Bill and with regard to its relationship with other pieces of legislation, including the Constitution (especially Chapters 2 and 3), as well as with their implementing subordinate legislation.

The development of the Marine Spatial Plan also needs to take into account the scope of the legislative and executive jurisdictions of the Eastern Cape Province and the adjacent coastal municipalities as well as the extent to which they have made use of those jurisdictions. That is in addition to pieces of legislation impacting on relevant coastal activities on land. This research will constitute the first attempt at assessing to what extent both sets of legislation are synchronised and if the steps taken to enforce them are compatible. Finally, research will focus on the legal regime governing coastal and marine tourism and recreation in Algoa Bay (including related underwater and over water space as these spaces articulate with the maritime domain[57]), as well as the extent to which the Marine Spatial Plan ought to take into account the range of relevant factors affecting good governance of tourism activities on both sides of the low-water mark.

Conclusions

The Algoa Bay Project will develop the first Marine Spatial Plan in South Africa, and feed directly into the process for developing a national Marine Spatial Plan as set out by the MSP Bill (2017).[3] The primary recipients of project outputs will be the National MSP Working Group, which has been tasked with producing the Marine Area Plans as set out in the MSP Framework.[4] The project will also serve to support refinement and improvement of the process of developing these area plans. In Phase 1, the Algoa Bay Project aims to inform the biophysical and governance components of the overall system dynamics model, while Phase 2 will focus on the socio-economic component. All three components will then be combined in the overarching system dynamics model that can be used as a decision-support tool to design efficient strategies for the sustainable use of the Bay.

National stakeholders involved in Operation Phakisa that will benefit from this project include the Departments of Science and Technology (Marine and Antarctic Research Strategy), Environmental Affairs (the MSP process; Marine Protected Areas), Agriculture, Forestry and Fisheries (sustainable use of ocean resources); Mineral Resources; and Energy (mining, oil and gas exploration). Other important stakeholders include SANBI (Marine Programme), SANParks and the Transnet National Ports Authority which manage the Port Elizabeth and Ngqura harbours. Locally, the main beneficiaries of the project will be the adjacent municipalities (for example, the Nelson Mandela Bay Metropolitan Municipality) as well as civil society[58] (which includes the private sector such as fisheries and tourism).

Finally, an important objective of the Algoa Bay Project is to raise the profile of marine science and the potential benefits of the 'Blue Economy' in the Eastern Cape, which is not traditionally known as a centre for marine research and development. This in turn should make the region more attractive for young scientists and entrepreneurs who in turn will help to develop the economic potential of the oceans and coasts of the Eastern Cape in a sustainable manner, to the benefit of all who live there.

References

1. National Planning Commission. National Development Plan 2030: Our future – make it work. Pretoria: National Planning Commission; 2012. Available from: https://www.gov.za/sites/www.gov.za/files/Executive%20Summary-NDP%202030%20-%20Our%20future%20-%20make%20it%20work.pdf

2. Zuma J. Address by President Zuma at the Operation Phakisa: Unlocking the economic potential of the ocean economy open day; 2014 October 15; International Convention Centre, Durban, South Africa. Available from: http://www.operationphakisa.gov.za/cc/Documents/Open%20Day%20Operation%20Phakisa%20President%20Speech.pdf.

3. Marine Spatial Planning Bill [B 9—2017]. Department of Environmental Affairs. Government Gazette no. 40726 of 28 March 2017.

4. Draft Marine Spatial Planning Framework. Department of Environmental Affairs. Government Gazette no. 40219 of 19 August 2016.

5. Schumann EH, Ross GJ, Goschen WS. Cold water events in Algoa Bay and along the Cape south coast, South Africa, in March/April 1987. S Afr J Sci. 1988;84:579–583.

6. Goschen WS, Schumann EH. Ocean current and temperature structures in Algoa Bay and beyond in November 1986. Afr J Marine Sci. 1988;7(1):101–116. https://doi.org/10.2989/025776188784379198

7. Roberts MJ. Dispersion of a buoyant effluent discharged into the nearshore waters of Algoa Bay [MSc thesis]. Port Elizabeth: University of Port Elizabeth; 1990.

8. Goschen WS, Schumann EH. An Agulhas Current intrusion into Algoa Bay during August 1988. Afr J Marine Sci. 1994;14(1):47–57. https://doi.org/10.2989/025776194784286914

9. Lutjeharms JR, Cooper J, Roberts M. Upwelling at the inshore edge of the Agulhas Current. Cont Shelf Res. 2000;20(7):737–761. https://doi.org/10.1016/S0278-4343(99)00092-8

10. Roberts MJ. Coastal currents and temperatures along the eastern region of Algoa Bay, South Africa, with implications for transport and shelf–bay water exchange. Afr J Mar Sci. 2010;32(1):145–161. https://doi.org/10.2989/1814232X.2010.481153

11. Goschen WS, Schumann EH. The physical oceanographic processes of Algoa Bay, with emphasis on the western coastal region. IMT document number: PO106-10000-730002. Pretoria/Cape Town: South African Environmental Observation Network (SAEON)/ the Institute of Maritime Technology (IMT); 2011.

12. Goschen WS, Schumann EH, Bernard KS, Bailey SE, Deyzel SH. Upwelling and ocean structures off Algoa Bay and the south-east coast of South Africa. Afr J Mar Sci. 2012;34(4):525–536. https://doi.org/10.2989/1814232X.2012.749810

13. Goschen WS, Bornman TG, Deyzel SH, Schumann EH. Coastal upwelling on the far eastern Agulhas Bank associated with large meanders in the Agulhas Current. Cont Shelf Res. 2015;101:34–46. https://doi.org/10.1016/j.csr.2015.04.004

14. Courtenay-Latimer M, Gibson-Hill CA. A preliminary note on the Bird Island group in Algoa Bay. Ostrich. 1946;17(2):75–86. https://doi.org/10.1080/00306525.1946.9638902

15. Randall EW. Species diversity and size ranges of cephalopods in the diet of jackass penguins from Algoa Bay, South Africa. Afr Zool. 1981;16(3):163–166. https://doi.org/10.1080/02541858.1981.11447752

16. Batchelor AL, Ross GJ. The diet and implications of dietary change of Cape gannets on Bird Island, Algoa Bay. Ostrich. 1984;55(2):45–63. https://doi.org/10.1080/00306525.1984.9634757

17. Karczmarski L. Group dynamics of humpback dolphins (*Sousa chinensis*) in the Algoa Bay region, South Africa. J Zool. 1999;249(3):283–293. https://doi.org/10.1111/j.1469-7998.1999.tb00765.x

18. Karczmarski L, Cockcroft VG, Mclachlan A. Habitat use and preferences of Indo-Pacific humpback dolphins *Sousa chinensis* in Algoa Bay, South Africa. Mar Mammal Sci. 2000;16(1):65–79. https://doi.org/10.1111/j.1748-7692.2000.tb00904.x

19. Dicken ML, Booth AJ. Surveys of white sharks (*Carcharodon carcharias*) off bathing beaches in Algoa Bay, South Africa. Mar Freshwater Res. 2013;64(6):530–539. https://doi.org/10.1071/MF12336

20. Koper RP, Karczmarski L, Preez D, Plön S. Sixteen years later: Occurrence, group size, and habitat use of humpback dolphins (*Sousa plumbea*) in Algoa Bay, South Africa. Mar Mammal Sci. 2016;32(2):490–507. https://doi.org/10.1111/mms.12279

21. Melly BL, McGregor G, Hofmeyr GJ, Plön S. Spatio-temporal distribution and habitat preferences of cetaceans in Algoa Bay, South Africa. J Mar Biol Assoc UK. 2017;May:1–5. https://doi.org/10.1017/S0025315417000340

22. Lasiak TA. Structural aspects of the surf-zone fish assemblage at King's Beach, Algoa Bay, South Africa: Long-term fluctuations. Estuar Coast Shelf Sci. 1984;18(4):459–483. https://doi.org/10.1016/0272-7714(84)90084-2

23. Beckley LE. The ichthyoplankton assemblage of the Algoa Bay nearshore region in relation to coastal zone utilization by juvenile fish. Afr Zool. 1986;21(3):244–252. https://doi.org/10.1080/02541858.1986.11447990

24. Beckley L, Buxton C. Underwater observations of reef fish in and around Algoa Bay, South Africa. Trans R Soc S Afr. 1989;47(1):29–38. https://doi.org/10.1080/00359198909520148

25. Pattrick P, Strydom NA, Goschen WS. Shallow-water, nearshore current dynamics in Algoa bay, South Africa, with notes on the implications for larval fish dispersal. Afr J Mar Sci. 2013;35(2):269–282. https://doi.org/10.2989/1814232X.2013.798593

26. Pattrick P, Strydom NA. Larval fish variability in response to oceanographic features in a nearshore nursery area. J Fish Biol. 2014;85(3):857–881. https://doi.org/10.1111/jfb.12477

27. Strydom NA, Coetzer CJ, Pattrick P. The complex early life history of a marine estuarine-opportunist fish species, *Solea turbyneі* (Soleidae) from temperate South Africa. Sci Mar. 2015;79(2):169–178. https://doi.org/10.3989/scimar.04193.18A

28. McInnes AM, Ryan PG, Lacerda M, Deshayes J, Goschen WS, Pichegru L. Small pelagic fish responses to fine-scale oceanographic conditions: Implications for the endangered African penguin. Mar Ecol Prog Ser. 2017;569:187–203. https://doi.org/10.3354/meps12089

29. Lipiński MR, Van der Vyver JS, Shaw P, Sauer WH. Life cycle of chokka-squid *Loligo reynaudii* in South African waters. Afr J Mar Sci. 2016;38(4):589–593. https://doi.org/10.2989/1814232X.2016.1230074

30. Chalmers R, Oosthuizen A, Götz A, Paterson A, Sauer WH. Assessing the suitability of commercial fisheries data for local-scale marine spatial planning in South Africa. Afr J Mar Sci. 2014;36(4):467–480. https://doi.org/10.2989/1814232X.2014.979228

31. Samaai T, Govender V, Kelly M. *Cyclacanthia* n.g. (Demospongiae: Poecilosclerida: Latrunculiidae incertae sedis), a new genus of marine sponges from South African waters, and description of two new species. Zootaxa. 2004;725(1):1–8.

32. Antunes EM, Copp BR, Davies-Coleman MT, Samaai T. Pyrroloiminoquinone and related metabolites from marine sponges. Nat Prod Rep. 2005;22(1):62–72. https://doi.org/10.1039/b407299p

33. Davies-Coleman MT, Veale CG. Recent advances in drug discovery from South African marine invertebrates. Mar Drugs. 2015;13(10):6366–6383. https://doi.org/10.3390/md13106366

34. Matcher GF, Waterworth SC, Walmsley TA, Matsatsa T, Parker-Nance S, Davies-Coleman MT, et al. Keeping it in the family: Coevolution of latrunculid sponges and their dominant bacterial symbionts. MicrobiologyOpen. 2017;6(2), e00417, 13 pages. https://doi.org/10.1002/mbo3.417

35. McLachlan A, Jaramillo E, Donn TE, Wessels F. Sandy beach macrofauna communities and their control by the physical environment: A geographical comparison. J Coast Res. 1993:27–38.

36. Brown AC, McLachlan A. The ecology of sandy shores. Cambridge, MA: Academic Press; 2010.

37. Robinson TB, Griffiths CL, McQuaid CD, Rius M. Marine alien species of South Africa – status and impacts. Afr J Mar Sci. 2005;27(1):297–306. https://doi.org/10.2989/18142320509504088

38. Nicastro KR, Zardi GI, McQuaid CD, Teske PR, Barker NP. Coastal topography drives genetic structure in marine mussels. Mar Ecol Prog Ser. 2008;368:189–195. https://doi.org/10.3354/meps07607

39. Von der Meden CE, Porri F, Erlandsson J, McQuaid CD. Coastline topography affects the distribution of indigenous and invasive mussels. Mar Ecol Prog Ser. 2008;372:135–145. https://doi.org/10.3354/meps07731

40. Von Der Meden CE. Intertidal patterns and processes: Tracking the effects of coastline topography and settlement choice across life stages of the mussels *Perna perna* and *Mytilus galloprovincialis* [doctoral dissertation]. Grahamstown: Rhodes University; 2009.

41. Porri F, Jackson JM, Von der Meden CE, Weidberg N, McQuaid CD. The effect of mesoscale oceanographic features on the distribution of mussel larvae along the south coast of South Africa. J Mar Systems. 2014;132:162–173. https://doi.org/10.1016/j.jmarsys.2014.02.001

42. Weidberg N, Porri F, Von der Meden CE, Jackson JM, Goschen W, McQuaid CD. Mechanisms of nearshore retention and offshore export of mussel larvae over the Agulhas Bank. J Mar Systems. 2015;144:70–80. https://doi.org/10.1016/j.jmarsys.2014.11.012

43. Jackson JM, Rainville L, Roberts MJ, McQuaid CD, Lutjeharms JR. Mesoscale bio-physical interactions between the Agulhas Current and the Agulhas Bank, South Africa. Cont Shelf Res. 2012;49:10–24. https://doi.org/10.1016/j.csr.2012.09.005

44. Mbambo SW. Scales of variability of phytoplankton composition and biomass in Algoa Bay, South Africa [dissertation]. Cape Town: University of Cape Town; 2014.

45. Pitcher GC, Cembella AD, Krock B, Macey BM, Mansfield L, Probyn TA. Identification of the marine diatom *Pseudo-nitzschia multiseries* (Bacillariophyceae) as a source of the toxin domoic acid in Algoa Bay, South Africa. Afr J Mar Sci. 2014;36(4):523–528. https://doi.org/10.2989/1814232X.2014.983971

46. Waterworth SC, Jiwaji M, Kalinski JC, Parker-Nance S, Dorrington RA. A place to call home: An analysis of the bacterial communities in two *Tethya rubra* Samaai and gibbons 2005 populations in Algoa Bay, South Africa. Mar Drugs. 2017;15(4):95. https://doi.org/10.3390/md15040095

47. Pfaff MC, Flaviani F, Du Plessis G, Rybicki EP, Schroeder DC. The overlooked foundation: Marine microbes in the oceans surrounding South Africa. Mar Maritime Sect. 2014:100.

48. Griffiths CL, Robinson TB, Lange L, Mead A. Marine biodiversity in South Africa: An evaluation of current states of knowledge. PLoS ONE. 2010;5(8), e12008, 13 pages. https://doi.org/10.1371/journal.pone.0012008

49. Matcher, GF, Froneman PW, Meiklejohn I, Dorrington RA. Distinct responses of bacterial communities to agricultural and urban impacts in temperate southern African estuaries. Estuar Coast Shelf Sci. 2018;200:224–233. http://dx.doi.org/10.1016/j.ecss.2017.11.015

50. Bremner JM. Surficial sediments in Algoa Bay. Marine Geoscience Group Technical Report 10. Cape Town: University of Cape Town; 1979. p. 66–74.

51. Lombard AT, Reyers B, Schonegevel LY, Cooper J, Smith-Adao LB, Nel DC, et al. Conserving pattern and process in the Southern Ocean: Designing a Marine Protected Area for the Prince Edward Islands. Antarct Sci. 2007;19(1):39–54. https://doi.org/10.1017/S0954102007000077

52. Grantham HS, Game ET, Lombard AT, Hobday AJ, Richardson AJ, Beckley LE, et al. Accommodating dynamic oceanographic processes and pelagic biodiversity in marine conservation planning. PLoS ONE. 2011;6(2), e16552, 16 pages. https://doi.org/10.1371/journal.pone.0016552

53. Haupt PW, Lombard AT, Goodman PS, Harris J.M. Accounting for spatiotemporal dynamics in conservation planning for coastal fish in KwaZulu-Natal, South Africa. Biol Conserv. 2017;209:289–303. https://doi.org/10.1016/j.biocon.2017.02.009

54. Bowd R, Quinn NW, Kotze DC. Towards an analytical framework for understanding complex social-ecological systems when conducting environmental impact assessments in South Africa. Ecol Soc. 2015;20(1):41. https://doi.org/10.5751/ES-07057-200141

55. Schoeman DS, Wheeler M, Wait M. The relative accuracy of standard estimators for macrofaunal abundance and species richness derived from selected intertidal transect designs used to sample exposed sandy beaches. Estuar Coast Shelf Sci. 2003;58:5–16. https://doi.org/10.1016/S0272-7714(03)00038-6

56. Sink KJ, Attwood CG, Lombard AT, Grantham H, Leslie R, Samaai T, et al. Spatial planning to identify focus areas for offshore biodiversity protection in South Africa. Final report for the Offshore Marine Protected Area Project. Cape Town: South African National Biodiversity Institute; 2011.

57. Vrancken P. Introductory perspectives on marine tourism in South African law. Obiter. 2011;32:613–633.

58. Reed J, Lombard AT. The role of civil society in supporting marine spatial planning. Maritime Review Africa. 2017 May/June; 20–24. Available from: http://www.maritimesa.co.za/BackIssues/2017/MayJune2017/tabid/172/Default.aspx

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

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].

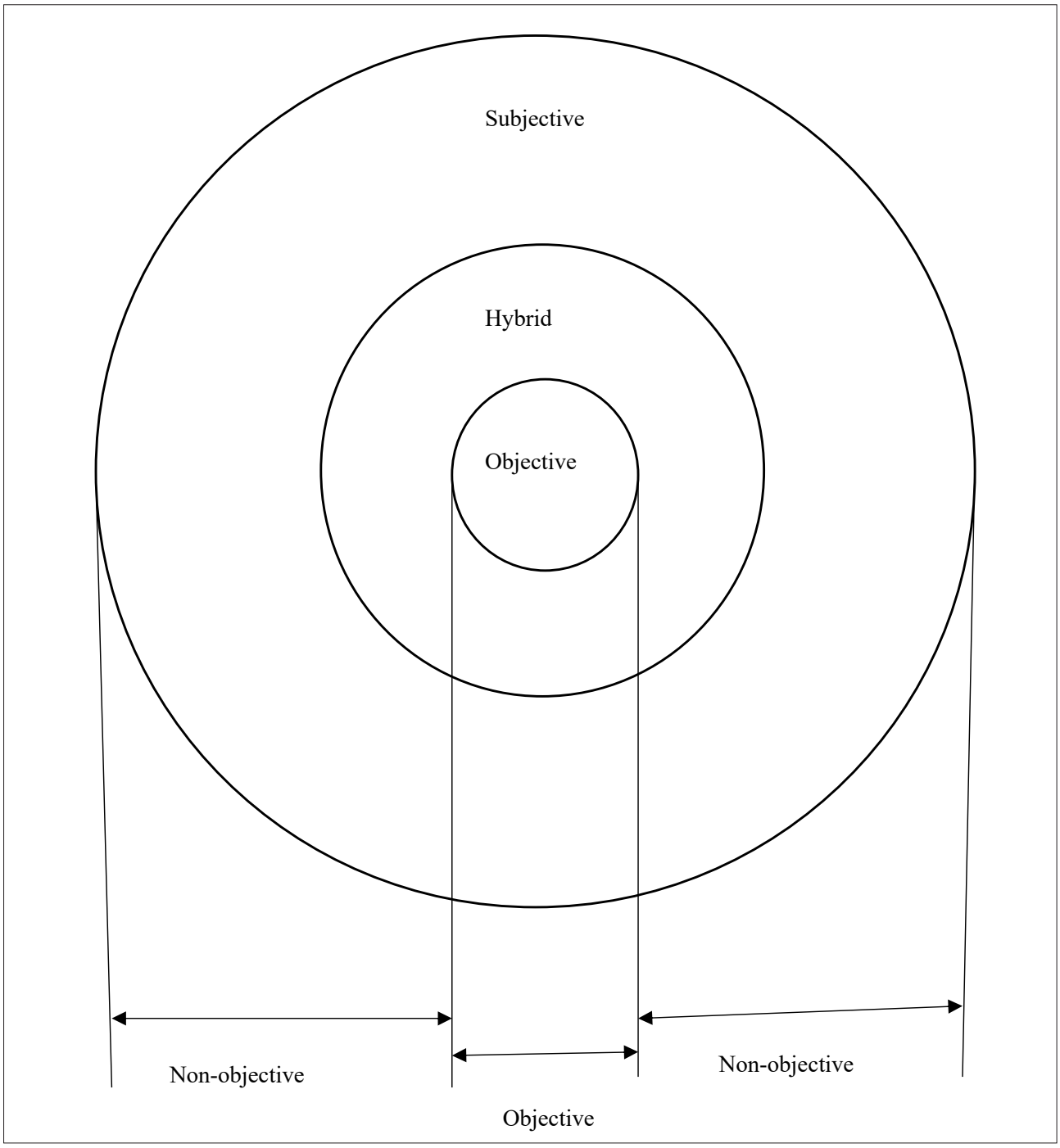


Figure 1: Error variance or 'potential for harm' of subjective versus objective NRF rating assessment.

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

Sclerotinia sclerotiorum disease prediction: A review and potential applications in South Africa

AUTHORS:
Lisa A. Rothmann[1]
Neal W. McLaren[1]

AFFILIATION:
[1]Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa

CORRESPONDENCE TO:
Lisa Rothmann

EMAIL:
coetzeeLA@ufs.ac.za

DATES:
Received: 11 May 2017
Revised: 08 Nov. 2017
Accepted: 30 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
Sclerotinia spp.; prediction models; plant disease epidemiology; fungicides; economic optimisation

HOW TO CITE:
Rothmann LA, McLaren NW. *Sclerotinia sclerotiorum* disease prediction: A review and potential applications in South Africa. S Afr J Sci. 2018;114(3/4), Art. #2017-0155, 9 pages. http://dx.doi.org/10.17159/sajs.2018/20170155

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

Sclerotinia sclerotiorum is a predominant plant pathogen, with host crops of agricultural and economic importance internationally. South African host crops of importance include canola, soybean and sunflower, which contribute significantly to the South African economy. This significance emphasises the importance of effective disease management strategies, including rotation with non-host crops, planting cultivars with a degree of tolerance, and using relevant cultural and chemical practices. The sporadic nature of disease outbreaks caused by *Sclerotinia* spp. can complicate fungicide application timing as a result of the pathogen's interaction with the host and environment. The use of prediction modelling for diseases caused by *Sclerotinia* spp. can contribute to increased fungicide application efficacy and a reduction in the number of unnecessary sprays. Predictive modelling is based upon the collection and statistical analysis of multi-locality and multi-seasonal, pathogen, disease and weather data. Incorporating the complexity of disease initiation and development into such models is dependent on selecting the correct statistical tools to interpret appropriate data, which can be used to develop a model that is accurate, precise and reliable. Internationally, forecasting models for diseases caused by *Sclerotinia* spp. exist and are applied commercially for multiple *Sclerotinia* spp. on important agricultural crops. The application of these models in a South African context has been limited but provides promise for effective disease intervention technologies. This review provides a platform to raise awareness of the potential applications of plant disease epidemiology and the use of statistics and mathematical modelling in agricultural systems. Plant disease forecasts are an important part of the future for sustainable and economically viable agronomic decisions.

Significance:

- Optimisation of plant disease management through ensuring that fungicide applications coincide with disease-favourable conditions, thus targeting the disease more strategically.
- The use of mathematical and statistical models to quantify the interactions among the host, pathogen and environment and predict future outbreaks of the disease.
- The study of temporal and spatial interactions among the host, pathogen and environment on plant disease behaviour.

Introduction

Sclerotinia sclerotiorum Lib. (de Bary) has the potential to cause disease in more than 400 plant species.[1] Internationally, economically important host crops of *S. sclerotiorum* include canola, dry bean, lettuce, peanut, soybean and sunflower.[2] In South Africa, primary risk crops are soybean and sunflower, with an increasing risk in canola.[3] The most prominent sign of *Sclerotinia* spp. is white cottony mycelium growing on infected plant tissues, with subsequent production of large melanised hyphal masses, called sclerotia. Symptoms associated with Sclerotinia diseases include a watery soft rot or wilting that can cause the collapse and death of host plants.[2]

Yield losses caused by *S. sclerotiorum*, particularly on canola in Sweden, the United Kingdom and Germany, may reach between 50% and 70%.[4,5] Canola is the third most valuable crop in Canada, where production has been severely affected by *S. sclerotiorum.*[6] In sunflowers, Sclerotinia head rot has been responsible for USD200 million damage annually in America.[1] In 2003, an estimated loss of USD1.9 million in the state of North Dakota (United States of America) was incurred as a result of Sclerotinia white mold on dry bean.[7] In South Africa during 2014, the effects of Sclerotinia stem rot of canola gained more attention as a consequence of the greater prevalence of the disease during that season compared to previous years.[8] Sunflower and soybean epidemics causing up to 60% severity and 65% yield losses, respectively, were recorded in South Africa over the 2013/2014 season.[9]

The economic importance of many of the crops affected by *Sclerotinia* spp. emphasises the importance of effective disease management strategies. Management of these diseases through rotation with non-host crops is limited by the extensive host range and duration of survival of sclerotia in soil. Agronomic management decisions – such as cultivar selection, proper irrigation management, planting dates and plant density – can also contribute to lowering disease severity.[4,10,11] The use of tillage practices in fields previously infested with *S. sclerotiorum,* to manage diseases caused by *S. sclerotiorum,* remains a point of contention.[9] C*oniothyrium minitans* is used as a biological control agent for fields previously infested with *S. sclerotiorum,* and is commercially available in South Africa.[9] However, timely fungicide applications, at critical host growth stages, can provide the most effective prevention measure for Sclerotinia diseases. Currently, there are a limited number of preventative registered active ingredients in South Africa for fungicide use on peas and sunflower; these include benomyl and procymidone.[12] However, the cost of fungicides and their application in addition to the potential need for multiple applications are constraints to the wider use of chemical control.[3,8,9]

The development of statistical or mathematical models for the management of plant diseases, through providing decision tools for prophylactic intervention and timely applications of fungicides, is based upon understanding the spatial and temporal aspects of plant disease epidemics. The sporadic nature and seasonal variation in Sclerotinia diseases complicates the development and validation of forecasting models. Current applications of disease models include forecasting disease initiation and severity, crop losses, pathogen dispersal and the timing of intervention technologies.[13] This review provides a platform to indicate the potential of predictive models for *S. sclerotiorum* in a global context and to promote economically viable Sclerotinia management strategies in South Africa.

Epidemiology of diseases caused by *Sclerotinia* spp.

Sclerotinia sclerotiorum infection of plants is largely driven by environmental variables which affect sclerotium survival and ascospore production.[2] In order to understand interactions within and among variables vital for the generation of effective disease prediction, a discussion of these determinants must precede a review of disease prediction models.

Formation and germination of sclerotia

A key life-cycle component of *S. sclerotiorum* is the sclerotium, the primary survival propagule. The prevention of formation, survival or germination of sclerotia is a primary disease control strategy.[14] Significant variation among *S. sclerotiorum* strains and, thus, varying sclerotial development and germination responses among strains, has been recorded.[15] However, in South Africa, low recombination frequencies have been recorded and strong similarities between isolates within and across provinces has been reported, suggesting that the population is young and developing.[16] Sclerotia can survive in the soil for 3 to 10 years. However, their survival is dependent on numerous factors, including *Sclerotinia species,* sclerotial size and shape, soil type and chemistry, soil microbial populations likely to degrade sclerotia, previous crops planted and climatic conditions.[4,16-19] Sclerotia have two germination mechanisms which initiate the processes of plant infection: carpogenic germination and myceliogenic germination.[1] These processes allow the pathogen to act as an air- and soil-borne pathogen, respectively.[20] The complex nature of influences causing either carpogenic or myceliogenic germination has not been determined absolutely and, thus, confounds progress in research and model development.

Carpogenic germination

Carpogenic germination of sclerotia results in stipe initiation followed by apothecia formation, responsible for ascospore release.[1] As a result, airborne inoculum is produced, facilitating the spread to, and the infection of, aerial tissues of host plants. Forcible discharge of ascospores is thought to be as a result of moisture tension, and atmospheric relative humidity (RH) fluctuations, allowing ascospores to travel distances ranging from 10 mm to over 100 m.[1,21] Thus, ascospores are not only found in the field in which the host crop is produced or infected.[16,18] Preconditioning of sclerotia is required to activate dormant sclerotia to germinate carpogenically. The duration of optimum conditions and variation among factors influencing preconditioning have resulted in the requirements for preconditioning being based on general assumptions. For example, it is assumed that sclerotium preconditioning usually occurs during winter or the non-cropping season.[2] In most cases, freezing is non-essential, although temperatures between 4 °C and 20 °C are sufficient to trigger carpogenic germination.[2]

The interaction between RH and air temperature is pivotal to the formation of apothecia, while soil temperature and moisture influence sclerotial germination responses.[1,11] Atmospheric temperatures ranging from 5 °C to 25 °C with high RH (not quantified in the literature cited) for approximately 10 days promote the development of apothecia.[2,18] When RH is at an optimum and air temperatures are ≥30 °C or ≤5 °C, increased incubation time and reduced carpogenic germination rate have been recorded, despite optimum RH.[19] Soil temperatures ≤10 °C and ≥35 °C significantly reduce sclerotial survival and carpogenic germination. Soil moisture ≥30% is favourable for apothecia formation.[2,11] Sclerotia in the upper 20 mm to 30 mm of the soil profile are regarded as competent epidemiologically; stipes at a depth >30 mm are rare under natural conditions, and stipes deeper than 30 mm cannot reach the sunlight needed for apothecial development.[11] Soil moisture in the upper 30 mm varies greatly as a result of surface weather conditions, notably air temperature, RH and wind as well as the growth stage of the crop. The latter is related to crop canopy density – a denser canopy increases RH and, therefore, soil moisture in the top 30 mm.[18] Cook et al.[22] found that sclerotia at >30 mm burial depths remained under exogenous dormancy, largely because of reduced soil moisture fluctuations which served as an anti-sclerotium germination stimulus.

The total time required for the formation of stipes to the initial dispersal of ascospores is approximately 3–4 weeks.[7] Once released, ascospores are viable for approximately 17 hours but, if conducive conditions are not present, host tissue infection will not occur.[23] Larger sclerotia produce greater numbers of apothecia and, thus, greater amounts of ascospores.[2] Sunflower head rot and canola and soybean stem rot are driven by carpogenic germination.[24,25]

Myceliogenic germination

Myceliogenic germination forms hyphae and mycelium from sclerotia, and is responsible for the direct sub-terrain infection of host plants, resulting in basal stem rots. The maximum distance between sclerotia and the host plant for infection to occur is 20 mm.[26] Colonisation of organic matter as an energy source, for the formation of mycelium and subsequent infection, is considered the germination trigger.[16,18]

Sclerotinia sclerotiorum regulates myceliogenic germination through sclerotial melanisation.[1] Incompletely melanised sclerotia are prone to reduced survival as a result of increased microbial degradation, sensitivity to adverse conditions and the absence of dormancy.[1,24] Myceliogenic germination will occur when a minimum of 12 hours of moderate temperatures (20–25 °C) with high RH (≥80%) occurs.[2,27] However, drying of sclerotia prior to higher soil moisture is known to trigger germination.[18] Soil moisture fluctuations between 30% and 100% field capacity promotes myceliogenic germination and host infection. Myceliogenic germination plays a major role in the disease cycle of Sclerotinia sunflower basal stalk rot and wilt and soybean basal stem rot.[24,25]

Infection process

Infection by ascospores occurs most frequently once flowering has commenced, as petals provide an exogenous source of nutrients and a path to infection. However, infection may take place before flowering on mechanically damaged or necrotic tissues.[28] Approximately 48–72 hours of continuous leaf wetness, RH ≥85% and temperatures of 20–25 °C are required for infection by ascospores.[5,18] However, infection by ascospores and subsequent disease development have been recorded at RH as low as 25% and soil moisture of 30%.[2,29] Mycelium is more tolerant to desiccation than ascospores and thus is more tolerant at lower RH.[7] Once infection has been initiated, *S. sclerotiorum* has the ability to remain inactive (latent) in host tissue in the absence of free moisture, resulting in delayed or arrested lesion development until favourable moisture levels re-occur.[30]

It is evident that survival, inoculum production and infection of plants by *S. sclerotiorum* involves numerous stimuli and pathogen responses to weather conditions, agronomic activities and host growth stages during critical phases of pathogen survival and growth. These variables need to be considered during model development to ensure the accuracy and precision of disease predictions.

Prediction models used for *Sclerotinia* spp.

In Europe, the management of Sclerotinia stem rot in winter canola is based on prophylactic fungicide sprays at full bloom as resistant cultivars are not available.[31] However, field experiments between 1981 and 2004 indicated that fungicide sprays were only 27–33% cost-effective.[31]

Bearing in mind the high cost of chemical applications, especially in view of the cost of procimydone within South Africa, this level of fungicide efficacy causes significant economic risk to producers.[12] The introduction of *Sclerotinia* risk assessment methods can benefit producers through the identification of intervention thresholds for optimising the timing of management strategies.[32] Models used to forecast diseases caused by *Sclerotinia* spp. vary in complexity from relatively low input point system models to more complex, multivariate systems. The following is a brief overview of the types of models, their applications and potential uses in a South African context.

Risk point systems

Risk point systems for diseases caused by *Sclerotinia* spp. have been developed for canola in Canada[33], Germany[34] and Denmark[35]. Prediction using the risk point system of Ahlers[34], for example, provided an accuracy of approximately 60% with a prevailing overestimation of disease incidence. An overestimation of disease risk will result in the unnecessary applications of fungicides and, thus, yield the model redundant. A similar Swedish risk point system was evaluated and calibrated to yield the model presented in Table 1.[4]

The Swedish risk point system was developed for spring sown canola in Sweden, using field-specific weather parameters and agricultural production practice data.[4] The only input requiring routine effort is the regional risk analysis for apothecium development, for which naturally produced sclerotia are placed in canola fields during spring, and monitored weekly to record the average number of apothecia developing from 100 sclerotia in each region.

Table 1: Forecasting model for Sclerotinia stem rot of canola[4]

Risk factor	Possible answers	Points
Number of years since last oilseed rape crop	More than 6 years	0
	3–6 years	5
	1–2 years	10
Disease incidence in last host crop	None	0
	Low (1–10%)	5
	Moderate (11–30%)	10
	High (31–100%)	15
Crop density	Low	0
	Normal	5
	High	10
Rain in the last 2 weeks	Less than 10 mm	0
	10–30 mm	5
	More than 30 mm	10
Weather forecast	High pressure	0
	Variable	10
	Low pressure	15
Regional risk for apothecium development (per 100 sclerotia)	0–5 apothecia	0
	6–10 apothecia	10
	11–100 apothecia	15

Producer will answer the risk factor questions and allocate the associated risk points; the sum of these points provides the producer with a fungicide application recommendation. If risk points are:

≤ 40, fungicide application not recommended

40–50 points, region of unknown risk

≥ 50, spraying recommended

The principle of risk point models is that potential answers are allocated weighted points based on the relationship between each risk factor and disease severity as observed in field evaluation trials. The Swedish risk point model evaluates fields at canola growth stage 65, i.e. flowering. The factors 'number of years since last oilseed rape crop' and 'disease incidence in last host crop' take sclerotial build-up and survival into consideration and thus primary inoculum pressure; 'crop density' takes into account the canopy microclimate; 'rain in the last 2 weeks' and 'weather forecast' take into account meso-climate; and 'regional risk for apothecium development (per 100 sclerotia)' provides the model with the risk of primary inoculum in specific regions. The points are totalled, and a fungicide application recommendation is made. If risk points are ≤40 then a fungicide application is not recommended. The area of unknown risk is between 40 and 50 points. If the risk points are ≥50 then a severe incidence of canola Sclerotinia stem rot is expected, and spraying is recommended. Twengström et al.[4] evaluated the accuracy of the risk point system at multiple sites over two seasons. The model correctly identified 75% of fields that required fungicide application. A 16% error occurred in which spraying was recommended but was not required.

During a 4-year trial, Foster et al.[36] monitored carrot crops, microclimate and pathogen variables in Canada to identify variables associated with white mold onset. Based on the relationships among variables and the number of apothecia/ascospores, the percentage of soil surface coverage by canopy growth (CG), mean daily soil moisture potential (soil matrix potential, SMP), and daily maximum or mean soil temperature (ST) were selected to develop algorithms to predict inoculum and, thus, disease risk. Crop closure at 100%, when one or two plants had collapsed leaves and 70–80% of plants were senescing, activated the forecasting system as a critical crop threshold was met. Initial apothecia observations were always after canopy closure, explaining the percentage soil surface shade in Table 2. The soil matrix potential and ST were included as microclimate factors as they were highly correlated ($p=0.01$) with initial ascospore occurrence (Table 2). The contribution of each level of selected crop and climate factors for any given day was described on a risk point system on a scale of zero to three – the higher the value, the higher the contribution to risk.

Table 2: Crop and microclimate risk factors, factor sub-ranges and corresponding multiplier values of points used to calculate risk for the occurrence of apothecia and ascospores of *Sclerotinia sclerotiorum* in carrot crops in Ontario, Canada[36]

Risk factor	Factor sub-range	Risk points
Crop Canopy growth (CG)	<95% of soil surface shaded 95–100% of soil surface shaded	0 1
Microclimate Soil matrix potential (SMP)	<-40 kPa -40 to -30 kPa -30 to -20 kPa > -20 kPa	0 1 2 3
Soil temperature (ST) Mean/maximum	≥25 / ≥ 30 °C 19 to 25 / 27 to 30 °C 19 to 22 / 24 to 27 °C ≤19 /≤ 24 °C	0 1 2 3

The producer will measure the risk factors and allocate the associated risk points. Each of these risk factors is multiplied and the sum of the multiplication over 7 days renders a 7-day severity index (SDSI) value, as illustrated by:

SDSI = ∑ (day 1-7) (CG × SMP × ST).

SDSI range interpretations:

0 to 32 risk points, ascospores < critical threshold = low risk present and management not required

33 to 44 risk points, ascospores > critical threshold = risk present and intervention could be required

45 to 63 risk points, ascospores > critical threshold = high risk and disease management practices should be applied

The risk algorithms were constructed to generate 7-day severity index values (SDSI). The models were as follows:

Model 1: $SDSI = \sum(day\ 1\text{-}7)(CG \times SMP \times ST)$

Model 2: $SDSI = \sum(day\ 1\text{-}7)(CG \times SMP)$

Once the crop threshold had been satisfied, the critical microclimate thresholds were the mean daily SMP of -40 kPa and mean ST of <25 °C. Model 2 is ideal for situations in which soil temperature data cannot be readily obtained, as they are excluded from the model.

Foster et al.[36] were able to reduce the number of fungicide applications, without compromising the degree of control provided, by using the calendar-based applications using Model 1. Model 1 was made available to Ontario carrot producers in 2006, therefore only Model 1 risk levels are discussed. Risk levels were as follows: (1) those in which the SDSI ranged from 0 to 32, ascospores were present but less than critical threshold for disease development, thus risk was present but management was not required; (2) SDSI ranged from 33 to 44 and ascospores were present and greater than critical threshold, thus the risk of disease was present and intervention could be required; (3) SDSI ranged from 45 to 63 and ascospore numbers were greater than critical threshold, indicating disease management practices should be applied as disease was likely to develop. Within-field ascospore numbers were measured on plates containing amended potato dextrose agar (PDA) selective to *Sclerotinia* species, as spore traps. The plates were placed on the bed of the carrot crop canopy for three consecutive hours in the mornings at each sampling site. The critical threshold was set at 10 colony forming units (CFU) per plate.[37]

The model developed by Foster et al.[36] is based on the same risk point premise as that of the Swedish risk point system; however, it takes into account more specific microclimate quantifications and the use of ascospore monitoring. The development of the algorithm was dependent on environmental variables being measured within a distance of 4.7 km from each field, and the authors suggested a limit of 5 km for the effective application of the models. A limitation, however, of both Models 1 and 2 is that of the labour-intensive and time-consuming process (approximately 3 days delay) of collecting and counting the CFUs, potentially resulting in the window of opportunity for fungicide applications being missed.[37]

The advantages of using relatively simple risk point tables to forecast when to apply fungicides are that they are easy to utilise by growers, have a quick and effective result distribution, are an inexpensive system and this method does not require the use of elaborate scientific equipment. This forecasting method allows for planning and cost-effective management in the agricultural industry.[4]

Prediction models based on petal infestation

Gugel and Morrall[32] identified a positive linear relationship between percentage petal infestation (PPI) of early bloom canola crops by ascospores of *S. sclerotiorum* and disease incidence (DI), although the regression R^2 values, an indicator of the goodness of fit of the model, were generally low. The model was based on the assumption that petal infestation reflects inoculum potential. Infected petals were found even when apothecia were not observed, indicating that extrinsically produced inoculum could also be recorded using this method.[32,38]

A forecast was derived from a prediction curve associated with the above regression model, where $<45\%$ mean percentage petal infestation (MPPI) corresponded with a low risk of disease, i.e. $<20\%$ mean disease incidence (MDI); 45–90% MPPI indicated moderate disease risk associated with a 20–40% MDI, and high disease risk correlated with MPPI $\geq$ 90% i.e. MDI $\geq$40%.[6] In subsequent trials, Turkington et al.[6] re-evaluated this relationship and the feasibility of using PPI as a forecasting tool in canola fields in Canada from 1985 to 1990. During the study, crop canopy density and timing of flowering were identified as valuable variables to determine MPPI, and thus MDI. They used arcsine-transformed MDI (TMDI) to meet the assumptions required for the regression analysis:

$$Y=\beta_0 + \beta_1 X_1 + e$$

where Y=TMDI, β_0 = intercept, β_1 = slope, X_1 = MPPI and e = error.

Canola canopy density was assessed qualitatively by rating crops as light, moderate or heavy in canopy. The inclusion of the canopy density correctly forecasted 73% of the disease risk over a 6-year period. However, the general tendency was that forecasting for disease risk below 45% was the most accurate.[41] The assumption of moisture being related to crop canopy density proved limiting, and the inclusion of weather data and sampling throughout the flowering period was to be followed in future studies to allow for more reliable and accurate forecasting of disease risk.[6]

Turkington and Morrall[39] speculated that petal infestation and disease risk may fluctuate in relation to changing conditions during flowering as a result of canopy density. Canopy density was quantified by measuring crop height, stem thickness, percentage light penetrated and leaf area index. Weather variables included temperature, RH and leaf wetness. Each of the canopy density variables correlated significantly with MDI. Assessments of petal infestation were conducted during early, full and late bloom; the MDI was significantly positively correlated with petal infestation at full and late bloom, and accounted for 98% of the variation in TMDI. This was associated with the increase in RH, daily leaf wetness, and the amount and pattern of rainfall during canopy closure, indicating that the relationship between MPPI and MDI is defined by the weather conditions during flowering. The addition of these variables to the existing model correctly forecasted disease incidence in 73% of 343 crops studied over a 6-year period. Disease incidence that was underestimated or overestimated in any field could be explained by the changes in PPI in below or above average canopy densities over the flowering period. The inclusion of petal infestation levels over the duration of the flowering period, as well as canopy density, increased the forecasting accuracy to 99%. Research conducted by Del Rio[40] supported the above model as weather conditions and the effect of crop microenvironment ultimately determine the relationships between disease incidence and petal infestation. This is correlated to the presence of sclerotia in the upper 20–30 mm of the soil profile, which are affected by the canopy microclimate.[11]

Canola flowers are most susceptible to infection by *S. sclerotiorum* when diurnal fluctuations – in RH, temperature and wind – occur and flowers have been fully opened for at least 2 days but are not near the point of wilting.[6] In the evening RH increases, while RH decreases during the day; increasing temperature and wind speed also result in a decrease in RH. These fluctuations affect the release of ascospores, but if the reduction in RH is too great, the 'drought effect' could decrease ascospore release. Wet petals or petals collected prior to heavy rain could result in an underestimation of disease because of the requirement of moisture for the release of ascospores.[6] Late bloom infections could prove a limitation to the model, as disease would have advanced and caused significant yield losses, thus fungicide applications would not be an advantage to producers. Furthermore, the abscission of leaves and resultant changes in the canopy density and microenvironment could halt stem infection. However, fungicides applied after early bloom may not be economically viable and, thus, late bloom infections may be associated with the risk of reduced fungicide efficacy. Further limitations may include equipment cost and expertise for monitoring equipment at multiple assessments, the variation in the ability of farmers to monitor variables associated with canopy density accurately and weather variables at infection. Thus, the use of a qualitative checklist including 'overcast, frequent rainfall and moderate temperatures' provided an indication of the risk of canola stem rot and thus the application of fungicides, but not a definitive forecast of when disease would be present in the field.[39]

Bom and Boland[41] found that the inclusion of soil moisture (SM) with petal infestation provided a stronger prediction model than petal infestation alone. The best prediction was achieved when a critical soil moisture content (SMc') of -10 or -15 kPa was used in combination with

petal infestation data. Fields with greater disease incidence (>20%) and a lower critical SMc' had the greater number of correct predictions of disease risk, in contrast to fields with limited disease incidence (<20%) in which disease risk was predicted correctly in fewer fields.[36,41]

Petal infestation, in canola crops, was recorded in various regions of Switzerland over 10 years, but severe Sclerotinia stem rot was observed only in years in which air temperature ≥12 °C between early and full bloom (growth stages BBCH 61 to 65).[42] This temperature threshold was believed to account for the poor correlation between the infection of petals with *S. sclerotiorum* and the severity of disease recorded in previous studies. Gladders et al.[43] evaluated the incidence of *Sclerotinia* on petals of canola in England. Inoculum was considered limiting for disease development. In other areas, petal infestation was much greater (>50% was the threshold) and subsequent Sclerotinia stem rot incidence warranted disease control measures. Analysis of data (personal observation, Gladders et al.[43]) indicated an exponential relationship between petal infestation (X) and Sclerotinia stem rot severity (Y) where $Y = 0.44X^{1.71}$ ($R^2 = 0.78$).

Initial studies to determine the incidence of petal infestation were dependent on grow-out tests on amended *Sclerotinia* selective PDA medium.[44,45] More recently, PPI quantification has been based on serological[46,47] and polymerase chain reaction (PCR) methods[48], and commercial test kits are available. The PPI model developed by Gugel and Morall[32] and amended by Turkington and Morrall[6,39] is a practical on-site model which farmers can conduct and apply themselves as *Sclerotinia* selective PDA kits and instructions are provided.

Prediction models based on carpogenic germination

Sclerotinia sclerotiorum causes lettuce drop which, in England, accounted for GBP9.5 million in losses at the farmgate in 2001.[11] A model to forecast disease risk using carpogenic germination of sclerotia was developed and evaluated.[49] This model is dependent on soil temperature and soil water potential as independent variables. The temperature components take into account temperature effects on sclerotia pre-conditioning for carpogenic germination. The assumption is that sclerotial pre-conditioning is a prerequisite for carpogenic germination and thus implied disease risk.[11]

The rates of pre-conditioning and germination were represented as follows:

$$R_c = a + be^{-kT} \quad \text{(exponential equation)}$$

where R_c is the rate of conditioning; T is soil temperature; and a, b and k are constants.

$$R_g = \exp\left(\frac{d_0+d_1}{T+273}\right) \quad \text{(Arrhenius equation)}$$

where R_g is the rate of sclerotium carpogenic germination; d_0 and d_1 are constants; and T is soil temperature.

The model developed by Clarkson et al.[11] supported research completed by Phillips[50] which stated that pre-conditioning of sclerotia by lower temperature resulted in more rapid sclerotium germination. Sclerotia not exposed to 4 °C conditioning, took more than 75 days to germinate, and less than 50% germinated. This indicates that temperature fluctuations can affect disease predictions. Optimum temperatures of 18 °C to 20 °C are required for the completion of the germination process within 30 to 50 days, when the soil has a water potential threshold of -100 kPa. A decrease in germination frequency occurs as temperatures approach 25 °C, with no germination at temperatures >26 °C and soil water potentials >-300 kPa.

Disease forecasting models, particularly those based on predicting critical inoculum potential, can have many sources of variation, including the pathogen isolates as well as the environment. However, the above models (exponential and Arrhenius equations) accurately predicted the germination or non-germination of sclerotia at approximately 80% of 64 sites. Errors associated with 'carpogenic germination' models, especially those associated with soil water potential, have limited the reliability of outcomes, because of the model's sensitivity to soil water potential. The above model is not yet complete as numerous issues must still be resolved, although Clarkson et al.[49] believe that the approach of predicting initial inoculum present in the field is promising for future effective prediction of Sclerotinia lettuce drop.

Probability regression models

Logistic and Poisson regression models were developed by Mila et al.[51] to calculate the prevalence and incidence, respectively, of Sclerotinia stem rot on soybeans. The percentage of infested fields in a region and the percentage of infected plants per field defined prevalence and incidence, respectively. Logistic regression can be applied when the target variable is qualitative with two categories, e.g. disease and no disease. Poisson regression is used in the study of rare events when the responses take on the form of counts. Sclerotinia stem rot was only found in 85 of the 1545 fields sampled, thus the occurrence of the disease can be considered a rare event.[51] The input variables for the models include soil conditions, agronomic practices and weather variables, i.e. temperature and rainfall. Tillage practices (i.e. conservation, minimum and conventional tillage) and soil texture were also recorded.

The development of the logistic prevalence model was divided into two parts. Model I dealt with the mid-western USA spring conditions (April) and Model II dealt with summer (July and August). The logistic regression analysis was used to identify significant relationships among environmental and disease variables:

$$P(Y=1) = exp\ (\sum biXi)/1 + exp\ [\sum biXi]),$$

where b_i are parameters to be estimated and X_i are covariates or predictors.

The logistic regression for Model I was:

$$P(Y=1) = exp\ (bo + b2State + b3AprilTxTillage + b4AprilPr)/(1+exp\ [bo + b2State + b3AprilTxTillage + b4AprilPr])$$

where:

Y (no, 0; yes, 1) = disease presence in a field

AprilT = mean temperature in April (°C) and AprilPr = total precipitation in April (cm)

Tillage (conventional, 0; no tillage, 1; minimum tillage, 2) = indicator variable for tillage practices

State (Illinois, 0; Iowa, 1; Minnesota, 2; Ohio, 3) = indicator variable for regional effects.

The logistic regression for Model II was:

$$P(Y=1) = exp\ (bo + b2State + b3AveragexTillage + b4JulyPr)/(1+exp\ [bo + b2State + b3AveragexTillage + b4JulyPr])$$

where:

Y (no, 0; yes, 1) = disease presence in a field

Average = mean temperature for July and August (°C)

JulyPr = total precipitation in July (em)

Tillage (conventional, 0; no tillage, 1; minimum tillage, 2) = indicator variable for tillage practices

State (Illinois, 0; Iowa, 1; Minnesota, 2; Ohio, 3) = indicator variable for regional effects.

Soil tillage influenced carpogenic germination, with conservation tillage promoting a greater DI of Sclerotinia stem rot, and ploughing decreasing the frequency of sclerotial germination through the relationship between tillage and soil moisture. Surface residues retained through conservation, minimum and conventional tillage were 30%, 15–30% and <15%, respectively.[48]

Strong R^2 values indicated that the models had good predictive accuracy with $R^2 = 0.65$ and $R^2 = 0.71$ for Models I and II, respectively. The models suggested that pre-season regional prevalence prediction of the risk could be feasible. The prediction values with spring (April) temperatures, where low temperature and precipitation increased the regional risk, were related to sclerotia pre-conditioning necessary for subsequent apothecial development.[11,49] However, based on previous epidemiological studies (see above) and the significant positive correlation between spring and summer temperatures with DI, Mila et al.[51] suggested that Model II was more reliable scientifically and could be used to explain the Sclerotinia stem rot epidemics on a regional scale. The restricted number of observations used to develop the models was a limitation as the Sclerotinia stem rot prevalence was overestimated in regions in which the observed prevalence was low.

The Poisson regression analysis was:

$P(y) = \lambda^y e^{-\lambda}/y!$,

where $\lambda = \exp(\sum b_i X_i)$ and b_i and X_i are the parameters to be estimated and the covariates or predictors, respectively.

This equation was used to quantify the relationship between independent variables and in-field *S. sclerotiorum* incidence. More agronomic practices were also incorporated in the Poisson regression analysis, i.e. the use of fungicide-treated seed, weed management, use and type of manure and fertiliser applications (Table 3).

R^2-values for this model were low, illustrated when estimated incidences were plotted against the observed incidences and yielded a $R^2=0.065$. The authors speculated that field specific information was not sufficiently integrated into the model, i.e. cultivars may have differed in susceptibility, microclimate variables were not quantified, and ascospore inoculum was not quantified. Under South African conditions, variation in stability of soybean cultivar responses to the disease have been reported and therefore a genetic coefficient to quantify genotype response to potential disease risk may be required when developing a South African Sclerotinia prediction model (*sensu* McLaren and Craven[52]). Nonetheless, the procedures and principle of using logistic and Poisson regression provide a useful approach to modelling Sclerotinia stem rot.[51]

Table 3: Parameter estimates of a Poisson regression analysis used to predict the incidence of soybean Sclerotinia stem rot (caused by *Sclerotinia sclerotiorum*) using mid-western USA summer variables and agricultural management practices[51]

	Poisson Model I		Poisson Model II	
Variable	**Parameter estimate**	**Standard error**	**Parameter estimate**	**Standard error**
Intercept	6.88[a]	1.3	13.28[a]	2.08
Average temperature in July & August	-0.46[a]	0.055	-0.68[a]	0.09
Iowa[b]	0.71[a]	0.28	0.48	0.35
Minnesota[b]	0.98[a]	0.29	0.46	0.37
Ohio[b]	-1.67[c]	0.75	-2.1[a]	0.77
July total precipitation	0.028[a]	0.01	0.033[d]	0.02
August total precipitation	0.037[a]	0.014	0.021	0.02
Average temperature in July & August x no tillage[c]	-0.022[d]	0.013	-0.027[d]	0.01
Average temperature in July & August x minimum tillage[c]	0.024[a]	0.006	0.008	0.01
Clay	***	***	-0.03[c]	0.01
June total precipitation	***	***	0.044[d]	0.02
Fungicide seed treatment	***	***	1.63[d]	0.87
Manure application	***	***	-18.63[a]	5.44
Weed cultivation	***	***	2.42[c]	1.14
Fertiliser application	***	***	-16.5[a]	3.63
Seed treatment x August total precipitation	***	***	-0.24[c]	0.11
Manure x average temperature	***	***	0.83[a]	0.23
Weed cultivation x June total precipitation	***	***	-0.37[a]	0.14
Fertiliser x clay	***	***	0.63[a]	0.15
Fertiliser x average temperature	***	***	0.082[a]	0.03
Deviance	1.195	***	628	***
Log (likelihood)	-528.7	***	-286.8	***

[a]*Level of significance, 1%*

[b]*Indicator variable for regional effect (Illinois, 0; Iowa, 1; Minnesota, 2; Ohio, 3)*

[c]*Level of significance, 5%*

[d]*Level of significance, 10%*

[e]*Indicator variable for tillage practices; conventional = 0; no tillage = 1; minimum-tillage = 2)*

Poisson regression model: $P(y) = \lambda y e^{-\lambda}/y!$, where y is the diseased soybean plants (1, 2,…,20) and $\lambda = \exp(\Sigma b_i X_i)$, with X_i as variables and b_i as parameter estimates presented in the table.

Harikrishnan and Del Rio[7] repeated the logistic regression method to develop a probability model for white mold of dry bean in North Dakota. They increased the resolution of disease measurement (250 fields surveyed from 2003 to 2005) and site-specific environmental data (maximum distance from a weather station was 20 km). Disease incidences ranged from 0 to 96%. Weather parameters recorded from May to August yearly included monthly total precipitation, number of rainy days and mean minimum air temperature (°C). A spray threshold of ≥20% disease incidence was used. Regression parameters are presented in Table 4. Although several forms of the model were evaluated and yielded R^2 values ≥75%, epidemiological significant predictors were selected. These included total rainfall (mm) in the first half of June, average minimum surface temperature (°C) in July and number of rainy days in August. The final singular model was highly sensitive (97%), specific (85%) and had a strong R^2 (0.85) associated with the model's overall statistical accuracy.

Table 4: Parameter estimates of a logistic regression model used to explain white mold incidence (≥20%) in dry bean crops in North Dakota, USA, during 2003 to 2005[7]

Variable	Parameter estimate	Standard error	Variance inflation	Pr>X²
Intercept	-26	8.85	0	0.0008
TRFJ1	-0.05	0.02	5.01	0.03
MINTJU1	1.5	0.64	4.49	0.01
RDAUG1	1.7	0.4	1.5	<0.0001

TRFJ1, total rainfall (mm) during first half of June; MINTJU1, average minimum temperature during first half of July; RDAUG1, number of rainy days during first half of August

A limitation of the model[7] was the overestimation of white mold incidence when disease incidence was <20%, although the model resulted in good accuracy when disease levels were >20%. The authors speculated that the addition of more variables or still greater resolution of disease assessment could be considered to improve accuracy of the model.

Crop-loss-related forecasting model

Koch et al.[5] developed a crop-loss-related model known as *SkleroPro* for winter sown canola in Germany. Although the model was developed in Germany, it has been applied widely in other parts of Europe. Interactions between DI and weather parameters, between DI and agronomic practices, and between DI and the changing price of the commodity in relation to yield gains associated with the cost of fungicide applications were evaluated.[5]

The *SkleroPro* structure and components are shown in Figure 1. The model runs from GS55 (mid-bud stage) to GS68 (end of flowering, initiation of pod formation), when it was too late for the probability of yield gain from fungicide application.[31] Hourly air temperature (T), precipitation (Precip), RH (measured hourly 2 m above the ground) and sunshine hours (Sun) per day are collected from the nearest weather station. These data are used in the computation of within-canopy temperature (T-Cpy) and RH (RH-Cpy). In order for the crop-loss model to give site-specific recommendations, the date of bud stage (GS55) and crop rotation information, i.e. 2, 3 or >3-year cycles (CropR), need to be provided by the canola producers. The inclusion of crop rotation cycles was pivotal to *SkleroPro* as it proved to be the strongest site-specific factor influencing the disease. The crop development stage (ONTO) is calculated by the multiplication of the growth function (ESF) by the temperature function (TES). This calculation is initiated at the GS55 stage and predicts the GS58 to GS68 using prevailing weather data. This model was developed by the German Meteorological Service.[5] The equations are illustrated below:

ONTO=TES X ESF; where

$$TES = \begin{cases} \dfrac{1}{1 + e^{0{,}198.(25-T)}} & \text{if } 55 \le GS \le 66 \\ \dfrac{1}{1 + e^{0{,}03.(80-T)}} & \text{if } 66 < GS < 66 \end{cases} \quad ESF = \begin{cases} \dfrac{0{,}6}{0{,}02.e^{(GS-55)^2}} & \text{if } 55 \le GS < 61 \\ \dfrac{0{,}15}{0{,}05.e^{(GS-61)^2}} & \text{if } 61 \le GS < 65 \\ \dfrac{0{,}15}{0{,}05.e^{(GS-65)^2}} & \text{if } 65 \le GS < 75 \end{cases}$$

where T = temperature (°C) and GS = growth stage provided by the producer.

The regional risk model is initiated when late-bud stage (GS58) is predicted by the ONTO function, indicating that senescing tissue is present and, thus, initiating a risk of Sclerotinia stem rot. The primary determinant for stem rot initiation is the accumulation of infection hours (Inh). An index (INFEST) is calculated to ensure that the infection hours occur during weather conditions conducive for disease development, i.e. 7 to 11 °C and 80–86% RH, where the optimum temperature for disease development is 18 °C. If the INFEST index indicated the Inh are during suitable conditions and >23 hours after the critical crop growth stage, GS58, a regional risk for stem rot is assumed.[5] The INFEST equation is illustrated below:

$$INFEST = \begin{cases} \dfrac{1}{e^{0{,}003.(18-T)^{2{,}5}} + 0{,}04.(94-RH)^2} & \text{if } T \le 18^{o}C \\ \dfrac{1}{e^{0{,}003.(T-18)^{2{,}5}} + 0{,}04.(94-RH)^2} & \text{if } T > 18^{o}C \end{cases}$$

where T = temperature and RH = relative humidity.

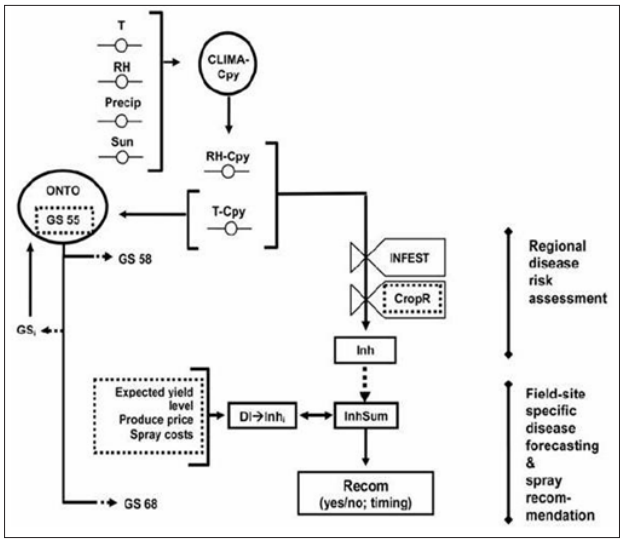
If the regional risk warning is delivered, the hours greater than the 23-hour threshold are accumulated (InhSum), which delivers the site-specific risk forecasts and fungicide application recommendations by calculating the economic damage threshold (Inh_i), illustrated below:

$$Inh_i = \frac{2{,}2 - \dfrac{C.100}{P.E} - 4{,}76}{\dfrac{-0{,}35}{0{,}21}}$$

where Inh_i = economic damage threshold; C= spray costs; P= produce price; E= expected yield; which are provided by the producer.

Site-specific information required to calculate the Inh_i includes the expected yield, commodity price and fungicide application costs. If the Inh_i value is greater than the InhSum value, a 'yes' recommendation to spray is delivered to the producer. Routine fungicide applications at GS65 are effective for disease management; however, site-specific recommendations during GS61 to GS69 add economic value for producers by taking advantage of the window of opportunity for maximising efficacy against stem rot infections.[5]

In a retrospective evaluation of the model, Koch et al.[5] reported a 70% accuracy, with 24% of predictions overestimating DI and 6% underestimating DI. Only 53% of sprays made to routinely sprayed (comprising five applications) fields were justified economically. The model reduced unnecessary fungicide sprays by 39% compared with routine spray applications. During the 2005 season, only 9% of the routine sprays proved to be justified economically, representing a saving of 75% of sprays had the model been applied commercially.[5] This model proved reliable as a crop-loss forecast model with a field, site and time specificity with regard to fungicide application recommendations. *SkleroPro* became available online in 2006, through the Information System for Integrated Plant Production (ISIP).[5]



T, temperature; RH, relative air humidity (hourly, 2 m above ground); Precip, hourly rainfall (mm); Sun, daily sunshine duration; GS, growth stage; CLIMA-Cpy, canopy microclimate; RH-Cpy, relative air humidity in the canopy; T-Cpy, temperature in the canopy; ONTO, simulation model of oilseed rape (OSR) development; CropR, frequency of OSR in crop rotation; DI, disease incidence; InhSum, sum of infection hours; Inhi, number of infection hours corresponding to the level of DI at the economic damage threshold; Recom, spray recommendation (yes or no and when).

Figure 1: Functional scheme of steps and components included in the forecasting procedure of *SkleroPro* from Koch et al.[5]

Conclusion

Although the predictive models described here focused on canola, dry bean, lettuce and soybean, the models form a basis for the development of risk assessment for diseases caused by *Sclerotinia sclerotiorum* on crops important to South African agriculture. Most of the models are built from the same basic driving variables that are restructured and recalibrated for the relevant crops, localities and purposes. The models demonstrate the importance and potential value of crop disease prediction models in the optimisation of timing intervention practices for disease management. Some of the above models have contributed to optimising fungicide application efficiency and have reduced the number of fungicide applications, thus mitigating environmental damage and reducing fungicide resistance build-up.[53,54] An important consideration is the number of variables included in the models and the need for a balance between what is practical in the field versus disease risk prediction accuracy. Evident from this review is that this balance has not yet been reached and additional improvements are required for many models to meet these desired characteristics for adoption by farmers. The sporadic nature of diseases caused by *Sclerotinia* spp. complicates the validation of models. As a result, many producers have been reluctant to implement such decision-support tools, because of a lack of reliability, the need for too many input variables, and a lack of infrastructure to deliver such tools to the agricultural industry.

The generation of *S. sclerotiorum* predictive risk models for South African circumstances will require local monitoring of the disease and the collection of weather and agronomic data from multiple localities and seasons. The Crop Estimates Committee (from the Department of Agriculture, Forestry and Fisheries) has been conducting *Sclerotinia* surveys with South African producers over the last decade and weather data can be provided by multiple reliable sources. The use of local *S. sclerotiorum* epidemiological knowledge and data could serve towards development of a forecasting model for South African canola, soybean and sunflower producers.[16,52]

The epidemiology of diseases caused by *Sclerotinia* species is still not fully understood on all possible host crops; therefore, more predictive models can potentially be developed and made available commercially. Predictive forecasts are potentially an important part of the future for economically viable agronomic decisions. Furthermore, it is clear that the models need to be locality-specific with adaptions required before a model can be applied to an area. Therefore, this review reveals opportunities for future research in South Africa for developing *S. sclerotiorum* forecasting models on susceptible crops of importance grown through the diverse production practices in this region of the world.

Acknowledgements

The financial assistance of the South African National Research Foundation (NRF) towards this research is hereby acknowledged. The work is based on the research supported in part by the NRF (grant number 111288). Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Authors' contributions

L.A.R. was responsible for the conceptualisation, funding acquisition, data curation, data collection and writing the initial draft. N.W.M. was responsible for the conceptualisation, funding acquisition, critically reviewing the writing and student supervision.

References

1. Bolton MD, Thomma BPHJ, Nelson BD. *Sclerotinia sclerotiorum* (Lib.) de Bary: Biology and molecular traits of cosmopolitan pathology. Mol Plant Pathol. 2006;7:1–16. https://doi.org/10.1111/j.1364-3703.2005.00316.x
2. Saharan GS, Mehta N. Sclerotinia diseases of crop plants: Biology, ecology and disease management. Dordrecht: Springer; 2008. https://doi.org/10.1007/978-1-4020-8408-9
3. Ramusi M, Flett B. Sclerotinia disease of sunflower: A devastating pathogen. Pretoria: GrainSA; 2015 [updated 2017]. Available from: http://www.grainsa.co.za/sclerotinia-disease-of-sunflower:-a-devastating-pathogen
4. Twengström E, Sigvald R, Svensson C, Yuen J. Forecasting Sclerotinia stem rot in spring sown oilseed rape. Crop Prot. 1998;17:405–411. https://doi.org/10.1016/S0261-2194(98)00035-0
5. Koch S, Dunker S, Kleinhenz B, Röhrig M, Tiedemann A. A crop loss model for Sclerotinia stem rot in winter sown oilseed rape. Phytopathology. 2007;97:1186–1194. https://doi.org/10.1094/PHYTO-97-9-1186
6. Turkington TK, Morrall RAA, Gugel RK. Use of petal infestation to forecast Sclerotinia stem rot of canola: Evaluation of early bloom sampling, 1985–90. Can J Plant Pathol. 1991;13:50–59. http://dx.doi.org/10.1080/07060669109500965
7. Harikrishnan R, Del Rio LEA. Logistic regression model to predict the risk of white mold incidence on dry bean in North Dakota. Plant Dis. 2008;92:42–46. https://doi.org/10.1094/PDIS-92-1-0042
8. GrainSA. Canola – Cut Sclerotinia stem rot down to size. Pretoria: GrainSA; 2016 [updated 2017]. Available from: http://www.grainsa.co.za/canola-cut-sclerotinia-stem-rot-down-to-size
9. Crave M, Ramusi M, Flett B. *Sclerotinia* – a disease of note in numerous crops. Oilseeds Focus. 2016 March;6–7. Available from: https://www.proteinresearch.net/imgs/newsletters/oilseeds-focus/2016/oilseeds-focus-vol-2-no-2-march-2016.pdf
10. Boland GJ, Hall R. Epidemiology of Sclerotinia stem rot of soybean in Ontario. Phytopathology. 1988;78:1241–1245. https://doi.org/10.1094/Phyto-78-1241
11. Clarkson JP, Phelps K, Whipps JM, Young CS, Smith JA, Watling M. Forecasting Sclerotinia disease on lettuce: Toward developing a prediction model for carpogenic germination of *Sclerotinia*. Phytopathology. 2004;94:268–279. https://doi.org/10.1094/PHYTO.2004.94.3.268
12. CropLife. South African Agricultural Remedies Database: Registered fungicides. Centurion: CropLife South Africa; 2015. Available from: http://croplife.co.za/images/croplife/home/Fungicides.pdf
13. Campbell CL, Madden, LV. Introduction to plant disease epidemiology. New York: Wiley; 1990.
14. Le Tourneau D. Morphology, cytology and physiology of *Sclerotinia* spp. in culture. Phytopathology. 1979;69:887–890. https://doi.org/10.1094/Phyto-69-887

15. Foley ME, Dogramaci M, Underwood W. A re-evaluation of myceliogenic germination of sclerotia for *Sclerotinia sclerotiorum* strain Sun-87 [abstract]. Paper presented at: National Sclerotinia Initiative 2016 Annual Meeting; 2016 January 20–22; Bloomington, MN, USA. Minneapolis, MN: United States Department of Agriculture; 2016.

16. Steyn C. Soybean response to rust and Sclerotinia stem rot under different biotic and abiotic conditions [thesis]. Bloemfontein: University of the Free State, 2015. Available from: http://scholar.ufs.ac.za:8080/xmlui/bitstream/handle/11660/2321/SteynC.pdf?sequence=1

17. Purdy LH. *Sclerotinia sclerotiorum*: History, disease and symptomatology, host range, geographic distribution and impact. Phytopathology. 1979;69:875–880. https://doi.org/10.1094/Phyto-69-875

18. Abawi GS, Grogan RG. Epidemiology of diseases caused by *Sclerotinia* species. Phytopathology. 1979;69:899–904. https://doi.org/10.1094/Phyto-69-899

19. Dillard HR, Grogan RG. Relationship between sclerotial spatial pattern and density of *Sclerotinia* minor and the incidence of lettuce drop. Phytopathology. 1985;75:90–94. https://doi.org/10.1094/Phyto-75-90

20. Hao JJ, Subarao KV, Duniway JM. Germination of *Sclerotinia* minor and *S. sclerotiorum* sclerotia under various soil moisture and temperature combinations. Phytopathology. 2003;93:443–450. https://doi.org/10.1094/PHYTO.2003.93.4.443

21. Steadman JR. White mold – A serious yield limiting disease of bean. Plant Dis.1983;67:346–350. https://doi.org/10.1094/PD-67-346

22. Cook GE, Steadman JR, Boosalis MG. Survival of *Whetzelinia sclerotiorum* and initial infection of dry edible beans in western Nebraska. Phytopathology. 1975;65:250–255. https://doi.org/10.1094/Phyto-65-250

23. Bečka D, Prokinová E, Šimka J, Cihlář P, Bečková L, Bokor P, et al. Use of petal test in early-flowering varieties of oilseed rape (*Brassica napus* L.) for predicting the infection pressure of *Sclerotinia sclerotiorum* (Lib.) de Bary. Crop Prot. 2016;80:127–131. https://doi.org/10.1016/j.cropro.2015.11.006

24. Huang HC, Dueck J. Wilt of sunflower from infection by mycelia-germinating sclerotia of *Sclerotinia sclerotiorum*. Can J Plant Pathol. 1980;2:47–52. http://dx.doi.org/10.1080/07060668009501437

25. Huang HC. Factors affecting myceliogenic germination of sclerotia of *Sclerotinia sclerotiorum*. Phytopathology. 1985;75:433–437. https://doi.org/10.1094/Phyto-75-433

26. Ferreira SA, Boley RA. *Sclerotinia sclerotiorum*. Hawaii: Crop Knowledge Master; 1992. Available from: www.extento.hawaii.edu/kbase/crop/Type/s_scler.htm

27. Dow RL, Porter DM, Powell NL. Effect of environmental factors on *Sclerotinia* minor and *Sclerotinia* blight of peanut. Phytopathology. 1988;78:672–676. https://doi.org/10.1094/Phyto-78-672

28. Hegedus DD, Rimmer SR. *Sclerotinia sclerotiorum*: When 'to be or not to be' a pathogen? FEMS Microbiol Lett. 2005;251:177–184. https://doi.org/10.1016/j.femsle.2005.07.040

29. Harikrishnan R, Del Río LE. Influence of temperature, relative humidity, ascospore concentration, and length of drying of colonized dry bean flowers on white mold development. Plant Dis. 2006;90:946–950. https://doi.org/10.1094/PD-90-0946

30. Tu JC. Management of white mold of white beans in Ontario. Plant Dis. 1989;73:281–285. https://doi.org/10.1094/PD-73-0281

31. Dunker S, Von Tiedemann A. Disease yield loss analysis for Sclerotinia stem rot in winter oilseed rape. IOBC. 2004;19:59–65.

32. Gugel RK, Morrall RAA. Inoculum-disease relationships in Sclerotinia stem rot of rapeseed in Saskatchewan. Can J Plant Pathol. 1986;8:89–96. http://dx.doi.org/10.1080/07060668609501848

33. Thomas P. Sclerotinia stem rot checklist. In: Canola growers manual. Winnipeg: Canola Council of Canada; 1984. p. 1053–1055.

34. Ahlers D. Integrated plant protection for fungus diseases in oilseed rape. Gesunde Pflanzen. 1989;41:306–311.

35. Jakobsen HL. Erfaringer med varsling for knoldbaegersvamp i Danmark 1986–91. [Experiences with monitoring of Sclerotinia in Denmark]. Nordiska Jordbruksforskares Forening (NJF) Seminar; 1991; Copenhagen, Denmark. p. 225–230.

36. Foster AJ, Kora C, McDonald MR, Boland GJ. Development and validation of a disease forecasting model for Sclerotinia rot of carrot. Can J Plant Pathol. 2011;33:187–201. http://dx.doi.org/10.1080/07060661.2011.563753

37. Kora C, McDonald MR, Boland GJ. Epidemiology of *Sclerotinia* rot of carrot caused by *Sclerotinia sclerotiorum*. Can J Plant Pathol. 2005;27:245–258. http://dx.doi.org/10.1080/07060660509507222

38. Morrall RAA, Dueck J. Epidemiology of Sclerotinia stem rot of rapeseed in Saskatchewan. Can J Plant Pathol. 1982;4:161–168. http://dx.doi.org/10.1080/07060668209501319

39. Turkington TK, Morrall PAA. Use of petal infestation to forecast Sclerotinia stem rot of canola: The influence of inoculum variation over the flowering period and canopy density. Phytopathology. 1993;83:682–689. https://doi.org/10.1094/Phyto-83-682

40. Del Rio LE. Development of a forecasting model to establish risk of Sclerotinia stem rot development on canola in North Dakota. Phytopathology. 2010;100:184.

41. Bom M, Boland GJ. Evaluation of disease forecasting variables for Sclerotinia stem rot (*Sclerotinia sclerotiorum*) of canola. Can J Plant Sci. 2000;80:889–898. https://doi.org/10.4141/P99-071

42. Gindrat D, Frei P, Vullioud P, Pellet D. Prediction of Sclerotinia stem rot on oilseed rape in Switzerland. Rev Suisse Agric. 2003;35:225–231.

43. Gladders P, Ginsburg D, Smith JA. Sclerotinia in oilseed rape: A review of the 2007 epidemic in England. Cambridge: Home-Grown Cereals Authority; 2008. Available from: http://cereals-2.ahdb.org.uk/publications/documents/cropresearch/PR433_Final_Project_Report.pdf

44. Morrall RAA, Rogers RB, Rude SV. Improved techniques of controlling Sclerotinia stem rot of canola (oilseed rape) with fungicides in western Canada. Mededelingen van de Faculteit Landbouwwetenschappen Rijksuniversiteit, Gent (Belgium). 1989;54:643–649.

45. Morton JG, Hall R. Factors determining the efficacy of chemical control of white mold in white bean. Can J Plant Pathol. 1989;11:297–302. http://dx.doi.org/10.1080/07060668909501116

46. Jamaux I, Spire D. Development of a polyclonal antibody-based immunoassay for the early detection of *Sclerotinia sclerotiorum* in rapeseed petals. Plant Pathol. 1994;43:847–862. https://doi.org/10.1111/j.1365-3059.1994.tb01629.x

47. Lefol C, Morrall RAA. Immunofluorescent staining of Sclerotinia ascospores on canola petals. Can J Plant Pathol. 1996;18:237–241. http://dx.doi.org/10.1080/07060669609500618

48. Qinab L, Fua Y, Xiea J, Chenga J, Jiangab D, Liab G, et al. A nested-PCR method for rapid detection of *Sclerotinia sclerotiorum* on petals of oilseed rape (*Brassica napus*). Plant Pathol. 2011;60:271–277. https://doi.org/10.1111/j.1365-3059.2010.02372.x

49. Clarkson JP, Phelps K, Whipps JM, Young CS, Smith JA, Watling M. Forecasting *Sclerotinia* diseases on lettuce: A predictive model for carpogenic germination of *Sclerotinia sclerotiorum*. Phytopathology. 2007;97:621–631. https://doi.org/10.1094/PHYTO-97-5-0621

50. Phillips AJL. Carpogenic germination of sclerotia of *Sclerotinia sclerotiorum*: A review. Phytophylactica. 1987;19:279–283.

51. Mila AL, Carriquiry AL, Yang XB. Logistic regression modelling of prevalence of soybean Sclerotinia stem rot in north-central region of United States. Phytopathology. 2004;94:102–110. https://doi.org/10.1094/PHYTO.2004.94.1.102

52. McLaren NW, Craven M. Evaluation of soybean cultivars for resistance to Sclerotinia stalk rot in South Africa. Crop Prot. 2008;27:231–235. https://doi.org/10.1016/j.cropro.2007.05.010

53. Fry WE. Integrated control of potato late blight-effects of polygenic resistance and techniques of timing fungicide applications. Phytopathology. 1977;67:415–420. https://doi.org/10.1094/Phyto-67-415

54. Prandini A, Sigolo S, Filippi L, Battilani P, Piva GF. Review of predictive models for *Fusarium* head blight and related mycotoxin contamination in wheat. Food Chem Toxicol. 2009;47:927–931. https://doi.org/10.1016/j.fct.2008.06.010

Water for sustainable development in the Berg Water Management Area, South Africa

AUTHORS:
Megan J. Cole[1,2]
Richard M. Bailey[1]
James D.S. Cullis[3]
Mark G. New[2]

AFFILIATIONS:
[1]School of Geography and the Environment, University of Oxford, Oxford, United Kingdom
[2]African Climate and Development Initiative, University of Cape Town, Cape Town, South Africa
[3]Aurecon South Africa, Cape Town, South Africa

CORRESPONDENCE TO:
Megan Cole

EMAIL:
meganjcole@gmail.com

DATES:
Received: 28 Apr. 2017
Revised: 24 Aug. 2017
Accepted: 29 Oct. 2017
Published: 27 Mar. 2018

KEYWORDS:
SDGs; water-dependent jobs; Berg WMA; water-use efficiency

HOW TO CITE:
Cole MJ, Bailey RM, Cullis JDS, New MG. Water for sustainable development in the Berg Water Management Area, South Africa. S Afr J Sci. 2018;114(3/4), Art. #2017-0134, 10 pages. http://dx.doi.org/10.17159/sajs.2018/20170134

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
University of Oxford

Water is fundamental to human well-being and economic growth. Measuring how water contributes to sustainable development is an important aspect of the United Nations Sustainable Development Goal (SDG) 6, 'Water and sanitation for all'. This importance is especially significant for water-scarce developing countries such as South Africa. Appropriate indicators can support decision-making and highlight key issues on inequality, unemployment and sustainability. In this paper, additional indicators for SDG 6.4 on water-use efficiency are proposed that focus on how individuals and households benefit, both directly and indirectly, from the allocations and use of water resources. The Berg Water Management Area (WMA) in the southwest corner of South Africa is used as a case study to illustrate the results. Residential per capita water use and municipal water losses were determined for all towns in the area. Figures for jobs and income per unit of water use were calculated for the heavily water-dependent industries, namely, agriculture, agriprocessing, freshwater aquaculture, mining and steel processing. This approach to measuring the socio-economic benefits of water use are relevant for other countries seeking to measure the role that water plays in achieving inclusive sustainable development, and could be included in the final SDG 6 indicator suite.

Significance:

- New measures of water-use efficiency based on jobs and income are proposed.
- New indicators are proposed for SDG 6.
- Water use, jobs and annual income are estimated for all heavily water-dependent sectors in the Berg WMA.

Introduction

Since the approval of 'Agenda 2030 for Sustainable Development' and its Sustainable Development Goals (SDGs) by the United Nations (UN) General Assembly in September 2015[1], the UN Inter-Agency and Expert Group (IAEG) on SDGs has developed 241 global indicators across the 169 targets and 17 goals[2]. The SDGs build upon the Millennium Development Goals but have a much broader vision, seeking to end poverty and hunger, protect the planet from degradation, ensure prosperity for all, and foster peaceful, just and inclusive societies.[1] Equity is at the heart of the SDGs and the aim is to 'leave no-one behind'. SDG 6 is summarised as 'Water and sanitation for all'. The six main targets cover access to safe drinking water (6.1), access to sanitation (6.2), water quality (6.3), water-use efficiency and scarcity (6.4), water resources management (6.5) and water-related ecosystems (6.6), while the means of implementation targets cover official development assistance (6.a) and participation of local communities in management (6.b). Nine global indicators have been defined by the IAEG for the main targets and IAEG, together with UN-Water, are developing indicator guidelines and methodologies to support national governments and to facilitate the sharing and comparison of data both within and across countries and over time.[3] The methodologies are undergoing pilot testing to incorporate lessons learnt at the country level[4] and the global indicators will be complemented by regional and national indicators developed by Member States[1].

Achieving SDG 6 is important for other SDGs as water is a fundamental enabler of sustainable development and economic growth, is essential for food security, health and sanitation, and is a basic human right.[5] Water resources are under pressure as a result of population growth, water-intensive economic growth, pollution and a changing climate.[6] According to the International Food Policy Research Institute, under a business-as-usual scenario, 45% of the global gross domestic product (GDP), 52% of the world's population and 40% of grain production could be at risk as a consequence of water stress by 2050.[7] The UN World Water Development Report 2016[7] estimated that 1.35 billion jobs are heavily water-dependent while 1.15 billion jobs are moderately water-dependent; thus 78% of the current global workforce depends on water for their livelihood.

SDG 6 is particularly important for South Africa. The water supply sector has become increasingly important for economic growth over the past three decades[8] as surface water resources (72% of supply) are rapidly approaching full utilisation[9]. South Africa is one of the driest countries in the world with low and highly variable rainfall, erratic run-off and high evaporation.[9] More than 60% of river flow arises from only 20% of the land area requiring large-scale inter-basin transfers.[9] At the same time, poverty, unemployment and inequality are significant problems in South Africa: almost 46% of the population are unable to meet their basic needs[10], the broad unemployment rate is 37%[11] and the Gini index of income inequality is 0.65 – one of the highest in the world[10]. The relationships between water and development are therefore very important.

Table 1 shows current reporting in South Africa on the main six SDG 6 targets. The obvious gap appears in SDG target 6.4, which is, by 2030, substantially increase water-use efficiency across all sectors and ensure sustainable withdrawals and supply of fresh water to address water scarcity and substantially reduce the number of people suffering from water scarcity. Indicator 6.4.1 – which measures change in water-use efficiency over time – is of specific concern, which is not unsurprising as there is a lack of data globally in this area.[7,12] In the most recent version of the step-by-step methodology for indicator 6.4.1, UN-Water define the indicator as 'the value added per water withdrawn, expressed in USD/m^3 over time of a given major sector (showing over time the trend in water-use efficiency)'[12].

Table 1: Draft Sustainable Development Goal 6 global indicators[4,12]

Target	UN-Water detailed indicator description	South Africa reporting
6.1 Access to safe, affordable drinking water	6.1.1 Proportion of population using improved drinking water sources located on premises, with water available when needed, and free from contamination (can be disaggregated by service level (none/basic/safe))	Annual General Household Survey (StatsSA)
6.2 Access to adequate sanitation and hygiene	6.2.1 Proportion of population using an improved sanitation facility at household level that is not shared with other households and excreta is safely disposed of, including a handwashing facility with soap and water (can be disaggregated by service level (none/basic/safe))	Annual General Household Survey (StatsSA) Note: access to shared/not shared facility not specified
6.3 Improve water quality	6.3.1 Proportion of wastewater safely treated (can be disaggregated by treatment level (primary/secondary/tertiary), source (household/economic activity) and recipient (fresh water/sea/soil))	Annual Green Drop report (DWS)
	6.3.2 Proportion of bodies of water with good ambient water quality (electrical conductivity/total dissolved solids; percentage dissolved oxygen; dissolved inorganic nitrogen/total nitrogen; dissolved inorganic phosphorus/total phosphorus; and faecal coliform/*Escherichia coli* bacteria) (can be disaggregated by water body and water quality parameter)	National monitoring programme (DWS)
6.4 Increase water-use efficiency and address water scarcity	6.4.1 Value added per water withdrawn, expressed in USD/m^3 over time of a given major sector (showing over time the trend in water-use efficiency)	
	6.4.2 Level of water stress: fresh water withdrawal as a proportion of available freshwater resources (total renewable freshwater resources less environmental water requirements)	Water reconciliation strategies (DWS)
6.5 Water resources management	6.5.1 Degree of integrated water resources management implementation (0–100) – assessing four components: policies, institutions, management tools and financing	DWS
	6.5.2 Proportion of transboundary basin area with an operational arrangement for water cooperation	DWS
6.6 Ecosystems	6.6.1 Change in time of the spatial extent, volume of water and health of water-related ecosystems	National Biodiversity Assessment (SANBI) every 7 years

StatsSA, Statistics South Africa; DWS, Department of Water and Sanitation; SANBI, South African National Biodiversity Institute

Three major sectors are defined by the International Standard Industrial Classification of All Economic Activities (ISIC): agriculture, forestry and fishing (ISIC A), mining and quarrying, manufacturing, construction and energy (ISIC B,C,D,F) and all service sectors (ISIC 36-39, 45-99). The indicator assesses the impact of economic growth on water resource utilisation rather than water productivity.[12] The South African Department of Water and Sanitation (DWS) does not measure water-use efficiency across sectors, although steps are being taken in this direction. In a recent study for the Berg Water Management Area (WMA) Resource Classification project[13], the gross value add (GVA) was calculated for eight high-level industry sectors and for sub-sectors that benefit from aquatic ecosystems. However, GVA per volume of water used only covers one aspect of water-use efficiency, and focuses on businesses and industries. As people and equity are at the heart of the SDGs, it is important to also measure the degree to which water use benefits individuals and households and contributes to their development.

In this paper, additional indicators for SDG 6.4 on water-use efficiency are proposed, based on a case study of the social and economic benefits of water in the Berg WMA in South Africa. These are 'residential per capita water use', 'municipal water losses', 'heavily water-dependent jobs per volume of water used' and 'heavily water-dependent income per volume of water used'.

Context – the Berg Water Management Area

South Africa's national DWS previously divided the country into 19 Water Management Areas (WMAs), each containing a large river system.[14] The Berg WMA, in the southwestern corner of the Western Cape Province, is named after the 285-km-long Berg River.[15] Although it is now part of the Olifants–Berg WMA (one of nine new WMAs established in 2013), it is still referred to as such in government assessments.[13] Rainfall ranges from over 3000 mm/a in the southeast mountains to less than 300 mm/a in the west and northwest coastal plain (Figure 1). The natural mean annual run-off (MAR) is 1094 million cubic metres per year (Mm^3/a) – 2.2% of the national total, although the present day mean annual run-off is 784 Mm^3/a; 65 Mm^3/a of groundwater resources are being utilised.[13] The Falkenmark indicator for the Berg WMA is 193 cubic metres per person per year ($m^3/c/a$) compared with 921 $m^3/c/a$ for South Africa (using Census 2011 population figures), thus classifying both areas as water scarce. The Berg WMA is home to the Cape Fynbos, a unique floral kingdom with World Heritage status, and the Berg River Estuary, which is of major conservation importance for birdlife.[16]

The Berg WMA includes the City of Cape Town metropole, the southern part of the West Coast district municipality and the western part of the Cape Winelands district municipality. There are 45 towns in the Berg WMA situated within six local municipalities (LM): Saldanha Bay LM (97% urban), Swartland LM (72% urban), Drakenstein LM (85% urban), Stellenbosch LM (77% urban), the southern half of the Bergrivier LM

(60% urban) and the western corner of Witzenberg LM (54% urban) (Figure 2). Over 4.5 million people live in the Berg WMA, with 87% being in the City of Cape Town.[17] In the Berg WMA, 79% of people have access to piped water in their dwelling and 92% have access to a flush toilet, compared with national figures of 46% and 57%, respectively.

The City of Cape Town contributes 71% of the Western Cape Province's GDP and is an important contributor to the economic growth in the area.[18] Economic activity in the Berg WMA is diverse; it includes tourism, agriculture, wine-making, food processing, manufacturing, fisheries, commercial forestry, financial services, ITC, nuclear power generation, hydropower generation and port operations.[13] The Saldanha Bay Industrial Development Zone, a Special Economic Zone, aims to be a world-class marine engineering hub focused on the international oil and gas sector and related services and support industries.[19]

The Berg River is a heavily utilised system with an annual water demand of 690 Mm^3/a, which is shared between residential and industrial users (52%), irrigation (43%) and afforestation and alien vegetation (5%).[13] The Western Cape Water Supply System serves 74% of the Berg WMA requirements (12)[13] and facilitates water transfers between rivers, including from the Breede–Gouritz WMA to the Berg River. Demand in the Western Cape Water Supply System is expected to exceed supply in the year 2022, although plans are in place to increase supply through the Berg River–Voëlvlei Dam augmentation, large-scale water re-use, large-scale Table Mountain Group Aquifer development, and desalination of seawater.[20] The increasing cost of water may constrain economic growth in the future.[21]

Methods

The proposed SDG 6.4 indicators 'residential per capita water use', 'municipal water losses', 'heavily water-dependent jobs per cubic metre of water used' and 'heavily water-dependent income per cubic metre of water used' are described below.

Residential per capita water use

Measuring residential per capita water use addresses both water-use efficiency and equity. The *National Water Act of 1998*[22] established the basic human needs reserve, which has been defined as 25 litres per capita per day (L/c/d) or 6000 litres per household per month. The World Health Organization recommends a minimum of 50 L/c/d for basic health, and 100 L/c/d in a country like South Africa with high TB and HIV/Aids rates.[23] Residential water-use data are not publicly available, but municipal water accounts for all towns in the Berg WMA[24-28] and the City of Cape Town were obtained for this study. The municipal accounts report bulk water supply, treated water supply, water losses, and water use by sector. The water use excludes the water losses so is a fairly accurate measure of end-use. The residential per capita water use was calculated using residential use data for the audit year July 2011 to June 2012 and Census 2011 population data, collected by Statistics South Africa (StatsSA) in October 2011, and stored in the SuperCross database.[17] In the case of Jamestown and Raithby, the DWS All Town Studies[29,30] were used as these towns do not have their own municipal accounts. For the privately owned De Hoek mine village, the Riebeek West mine village residential per capita use figure was used.

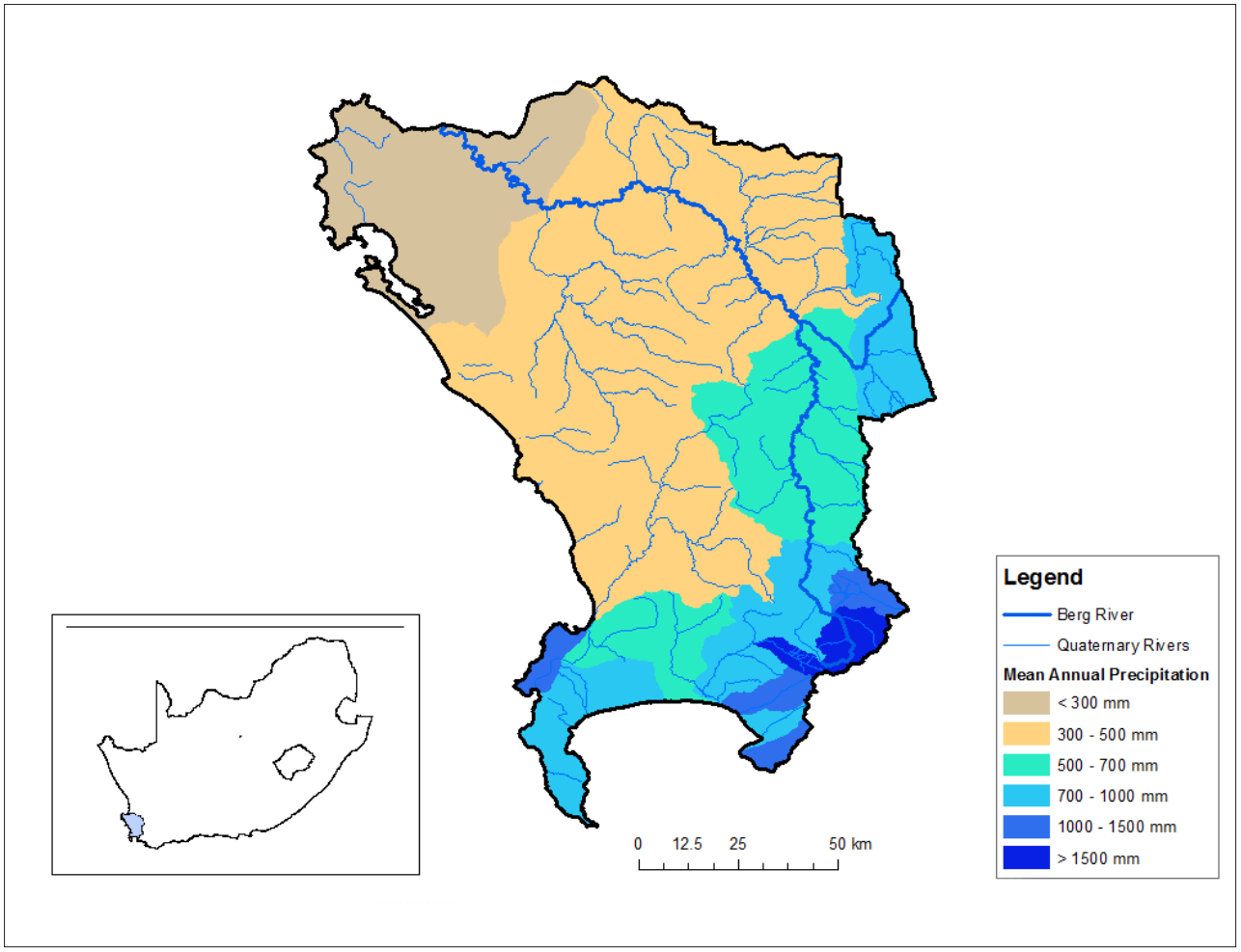


Figure 1: Mean annual precipitation and quaternary rivers in the Berg Water Management Area.

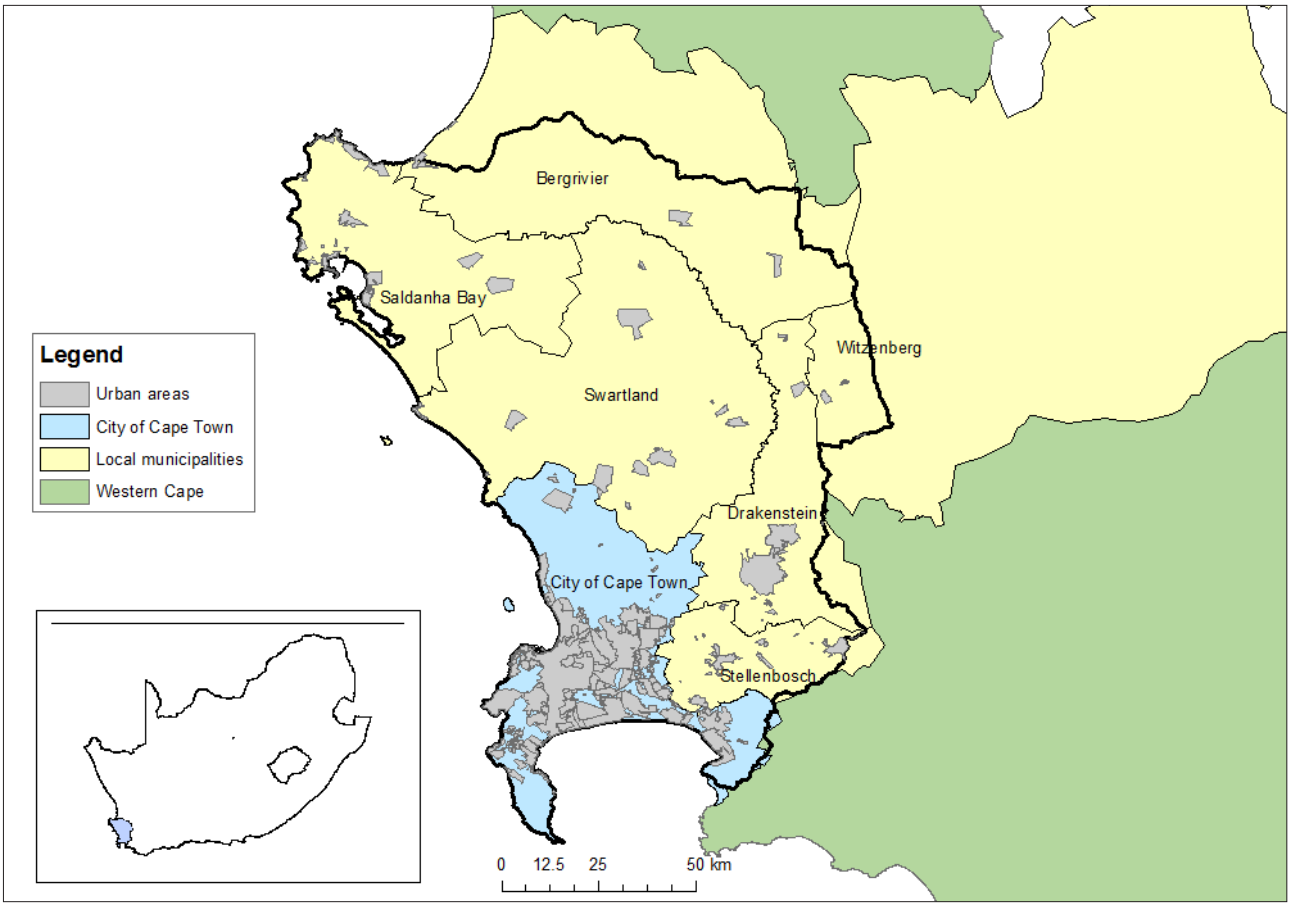


Figure 2: Urban areas, local municipalities and the City of Cape Town within the Berg Water Management Area.

Municipal water losses

Reducing water losses is a focus area for water-use efficiency strategies. Municipal water use accounts record the physical losses of raw water through bulk distribution and water treatment (i.e. the difference between raw water and treated water supply) and the losses of treated water through internal distribution (i.e. the difference between treated water supply and billed metered consumption). Most towns in the Berg WMA receive treated water, therefore the water losses quoted are for internal distribution only.

Water-dependent jobs and income

Water-use efficiency implies maximising output per unit of water. This output could be measured as gross value added, income or jobs. A few attempts have been made to determine water-use efficiency in the Western Cape and the Berg WMA.[13,31,32] These attempts have been limited to a single sector (agriculture or agriprocessing) or to general industry categories such as 'Agriculture, Hunting, Forestry and Fishing' and 'Manufacturing', which include jobs with a wide range of water dependence. Using the general industry categories does not allow analysis of heavily water-dependent jobs, moderately water-dependent jobs, and jobs in other sectors. This challenge is addressed in this paper by using the microdata provided by StatsSA in the Census 2011 '10% sample' which provides employment figures for 263 different industry variables. The census data also provide the income bracket for each job. By combining water-use data and census data, 'jobs per million cubic metres water used in heavily water-dependent sectors' and 'total annual income per cubic metre water used in heavily water-dependent sectors' can be calculated.

The national census of 2011 was conducted in October 2011 for all 51 770 560 South African residents.[33] All persons aged 15 and older were asked about their employment status, employment industry and occupation. Unfortunately, the industry and occupation data are not included in StatsSA's SuperCross database. Instead, StatsSA has provided the Census 2011 '10% sample', which includes anonymised data for individuals in about 10% of all households and collective quarters. StatsSA do not provide publicly available information on how the sampling was done but it is understood that they did systematic sampling on the households and then included individuals from the selected households. The '10% sample' includes (among many other variables) the employment industry, occupation, income and local municipality for individuals.

In the Berg WMA, the '10% sample' includes 132 564 employed people – 8.4% of the actual employed population. For this study, the 263 industry variables were grouped into sub-sectors related to their dependence on water as defined by the UN.[7] Agriculture, forestry, inland fisheries, mining and resource extraction, energy and water supply, agriprocessing, and manufacturing of textiles are defined as heavily water-dependent sectors. Manufacturing/transformation of wood products, rubber, plastics, metals and chemicals, and construction and transport are defined as moderately water-dependent sectors. All other sectors have low water dependence. Using the 10% sample, the percentage of jobs in each heavily water-dependent and moderately water-dependent sector were calculated per municipality. Here, the ratio of sector jobs to total jobs in the '10% sample' and the total number of jobs in the full Census 2011 were used to estimate the actual number of jobs per sector and municipality. The income of each worker was aggregated to calculate a total annual income per sector and municipality. As income is reported as an income bracket, not actual income, the median income of the bracket specified per person was used. The minimum and maximum total annual income were also calculated using the lower and upper ranges of the specified bracket. The income results will be underestimates as they exclude 6% of jobs with 'unspecified' income.

Further analysis of water-dependent jobs is limited by water-use data. The municipal accounts record water use for industrial/business/commercial, municipal, farm and 'other' sectors, but the exact industries and sub-sectors are unknown. The DWS database of all registered water use – called Water Authorisation Registration and Management System (WARMS) – specifies 11 water-use sectors: irrigation, livestock watering, aquaculture, industry urban, industry non-urban, other urban, mining, power generation, water supply service, recreation and schedule 1. It also records customer name, water source and geographical location, which enables identification of the sector in most cases and the municipality. In this study, the latest WARMS database was obtained from DWS in April 2016. A total of 6089 registered water uses in the Berg WMA were analysed and the total WARMS water volumes for the heavily water-dependent sectors (agriculture, freshwater aquaculture, mining and agriprocessing) were calculated for each municipality. Power generation was excluded as the operations do not use fresh water. In addition, grey literature on heavy water users in the Berg WMA was reviewed to determine the water use and workforce. This review identified steel processing in Saldanha[34] as a major water user.

The jobs and income analysis from the '10% sample' was combined with the water use analysis to calculate best estimates for the five heavily water-dependent sectors for the two indicators 'jobs per volume water used' and 'total annual income per volume water used'. The details for each sector are described below.

Agriculture

There are 2540 active registered water uses for irrigation and 89 for livestock watering in WARMS, with total water allocations of 387 650 971 m^3/a and 3 715 023 m^3/a, respectively. In addition, towns supply 1 304 352 m^3/a to farms. The total annual water use for agriculture is therefore 392 670 356 m^3/a. There are 4314 jobs in the '10% sample' with industry variables for 'crop farming', 'animal farming', 'mixed farming', 'agricultural and animal husbandry services' and 'organic fertiliser production'. As this is 3.3% of jobs in the 10% sample, it was estimated that there were 40 465 jobs in agriculture in 2011. Jobs and income were analysed for the six LMs and the City of Cape Town.

Freshwater aquaculture

There are 16 active registered water uses for freshwater aquaculture in WARMS with a total of 2 404 565 m^3/a. They are located near Paarl, Franschhoek and Tulbagh and are likely to be for trout farming in mountain streams. There are 19 jobs in the '10% sample' with the industry variable for 'fish hatcheries and fish farms' located in the Drakenstein and Witzenberg LMs. As this figure is 0.1% of jobs in the 10% sample, it was estimated that there were 42 690 jobs in freshwater aquaculture in 2011. Jobs and income were analysed for the two LMs.

Mining

There are only two mines in the Berg WMA, both operated by PPC Ltd: De Hoek Mine in the Bergrivier LM and Riebeek West Mine in the Swartland LM. Together they have a registered water use of 1 276 526 m^3/a in WARMS and a workforce of 400 in their limestone mines and cement factories.[35] The PPC mines have associated mine villages, which are recorded as separate main places (towns and suburbs defined for census collection) in Census 2011 with 324 employed people. One can assume that all these people work on the mine sites. There are 61 quarries in the Berg WMA[36], but there is very limited information on their water use and jobs. In WARMS, there are seven registered water uses for quarries, ranging from 6000 m^3/a to 134 000 m^3/a, although the designation is usually 'industry urban' and not 'mining' so it is uncertain what the water is used for. The '10% sample' under-samples mining and quarrying with only two employed persons in Bergrivier LM and eight in Swartland LM. It is assumed that the two mine villages were not included in the sample. Given these constraints, the jobs and income analysis was limited to the two PPC mine villages, but extrapolated for the additional PPC mine employees who live off-site.

Steel processing

Saldanha Works, part of ArcelorMittal South Africa, is an export-focused steel plant near the Saldanha port that produces 1.2 Mt/a of hot rolled coil from iron ore. The plant has 568 employees.[37] It does not appear in WARMS so it was assumed that it uses municipal water from Saldanha Bay LM. It consumes 8 million litres of water per day[34] – 2920 million litres per year or 2 920 000 m^3/a. Each employee therefore benefits from 5157 m^3/a or 14 128 litres per day. Therefore, there are 195 jobs per million cubic metres per year. As the employees cannot be identified in the census records, no income analysis could be done.

Agriprocessing

Agriprocessing is a combination of different activities that transform agricultural products into useable items. The Western Cape Department of Agriculture provides the location and name of all agriprocessing facilities in the province in its online tool, CapeFarmMapper.[38] The Berg WMA has 12 red meat abattoirs, 8 white meat abattoirs, 129 agriprocessing plants, 137 dairies, 5 fruit packers, 4 cool chain facilities, 117 packhouses and 591 cellars – a total of 1003 agriprocessing facilities. In 2015, the Western Cape Department of Environmental Affairs and Development Planning (DEA&DP) assessed the water use of seven priority agriprocessing sectors – fruit juice, wine and brandy, olive oil, essential oils, dairies, poultry abattoirs and livestock abattoirs – using CapeFarmMapper, a literature review and interviews.[32] Based on a limited sample size, they estimated the water use of each sector in litres per unit. The WARMS database also provides information on water use, and for this study 57 active registered water uses were assessed, totalling 5 896 885 m^3/a for 1 brewery, 1 tannery, 7 food and beverage producers and 23 wine cellars. As neither source covers all agriprocessing facilities, a literature survey was undertaken to fill in any gaps to estimate total annual water use per sub-sector per municipality.

For breweries, the single brewery WARMS licence and an estimated average water use of 385 m^3/a for 60 craft breweries[39] based on production figures[40] and SABMiller's figure of 8.55 litres of water required to produce one litre of beer[41] were used. For wine cellars, production figures[42] and DEA&DP's estimate of 3.9 litres of water use per litre of wine produced were used. For dairies, production figures[43] and DEA&DP's estimate of 3.6 litres of water per litre of dairy product were used. For food and beverage processing, the average water volume for the six plants in WARMS was multiplied by the number of plants in CapeFarmMapper. For tanneries, the single WARMS licence of 55 000 m^3/a was used. For white meat abattoirs, the number of abattoirs in CapeFarmMapper, an average number of birds per facility, and DEA&DP's estimate of 16.7 litres of water per bird were used. For red meat abattoirs, the number of abattoirs in CapeFarmMapper, an average number of animals per facility, and an estimate of 50 litres of water per animal were used. All these figures exclude the water required to farm or transport the required agricultural inputs – both food and pastures. All the figures have significant uncertainty because of the very limited data on actual water use in individual agriprocessing facilities and therefore averages were used extensively.

Overall, 3594 agriprocessing jobs with the industry variables 301 to 306 in the six LMs and the City of Cape Town in the census '10% sample' were analysed. These jobs make up 3.9% of jobs in the 10% sample; therefore, there were an estimated 40 340 jobs in agriprocessing in 2011 in the Berg WMA.

Results

Table 2 provides summarised data for the four indicators described above for the Berg WMA and its municipalities, while Figure 3 plots residential water use and water losses at the town level, highlighting local variability. Figure 4 and Figure 5 show comparisons for estimated 'jobs per million cubic metre of water use' and 'median income per cubic metre of water use', respectively, among heavily water-dependent sectors across municipalities in the Berg WMA. Figure 5's error bars show that income could range from the lower to the upper end of the census income brackets. The underlying town- and municipal-level data used in the figures are provided in the supplementary material (Supplementary tables 1–6).

Table 2: Indicators of water-use efficiency in the Berg Water Management Area (WMA)

Municipality	Residential per capita water use (L/c/d)	Municipal water losses (%)	Jobs per million cubic metre in heavily water-dependent sectors (Mm³/a)	Total income per cubic metre in heavily water-dependent sectors (ZAR/m³/a)
Bergrivier	138	13	114	79 ± 26
Saldanha Bay	112	18	786	239 ± 80
Swartland	107	15	234	102 ± 32
Drakenstein	106	11	41	18 ± 6
Stellenbosch	152	18	7	9 ± 3
Witzenberg	107	10	112	15 ± 5
City of Cape Town	125	14	932	187 ± 62
Berg WMA	124	16	177	184 ± 61

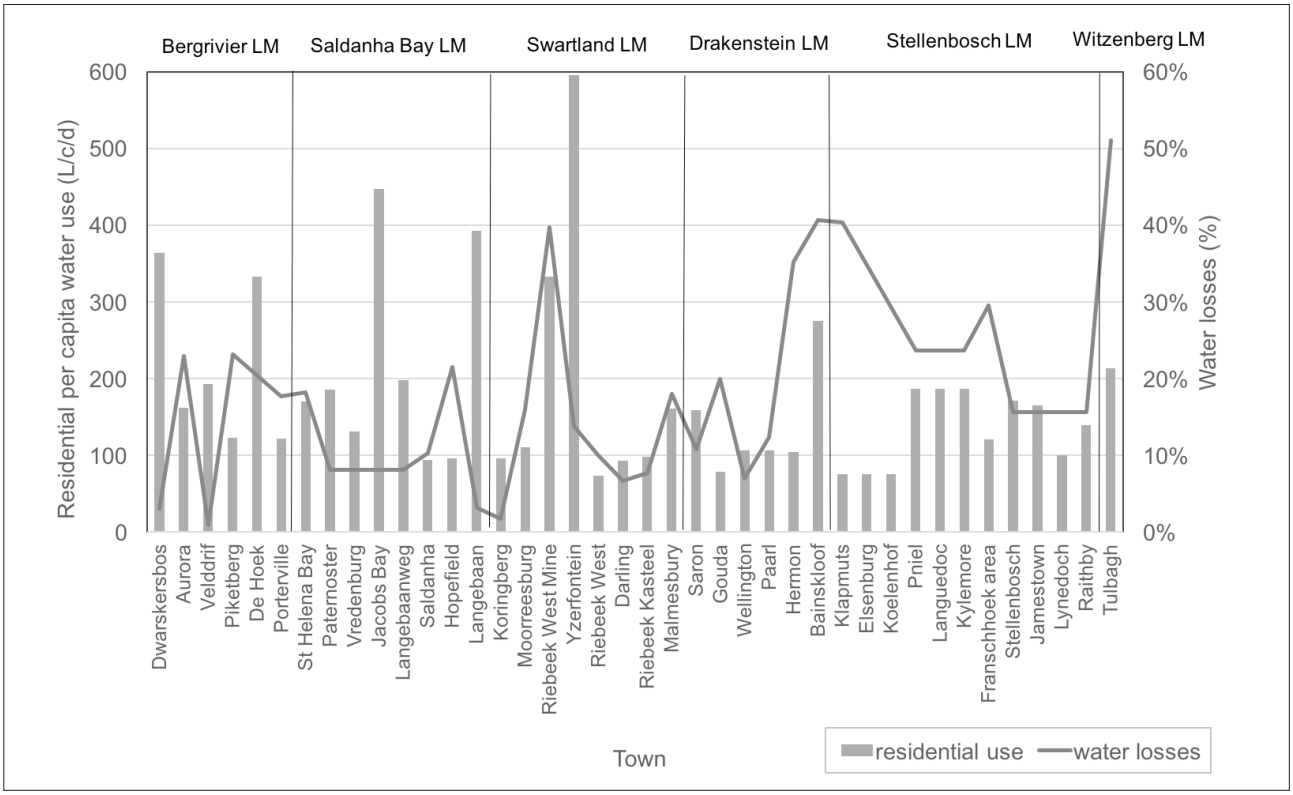


Figure 3: Residential per capita water use and municipal water losses (bulk water supply less end-use) in towns in the Berg Water Management Area.

In addition, Supplementary table 7 provides municipal-level data for moderately water-dependent jobs per sector. These results show that, although at the municipal level residential water use is within the targeted range, 11 towns are below the recommended minimum target of 100 L/c/d and 6 towns are above the proposed maximum target of 300 L/c/d. Similarly, although municipal-level water losses are all below 20%, they can reach over 50% at the town level. Residential per capita water use ranges significantly from 73 L/c/d (in Riebeek West) to 596 L/c/d (in Yzerfontein) and water losses range from 1% (in Velddrif) to 51% (in Tulbagh) at town level. The jobs and income vary significantly across municipalities and sectors but, overall, agriprocessing produces the most jobs (634 jobs/Mm³) while mining produces the most income per unit of water used (ZAR57.52/m³). The City of Cape Town and Saldanha Bay LM produce the most jobs and income per unit of water used.

Discussion

In this study, data on residential water use, municipal water losses, and jobs and income per unit of water used have been determined for the Berg WMA, its six local municipalities and its metropole, the City of Cape Town. These data have provided a case study for potential additional indicators to measure water-use efficiency in SDG 6.4. They have also provided an insight into the variability that can be found at different spatial scales and levels. This approach supports the SDG requirement to disaggregate data by geographical location and other factors.

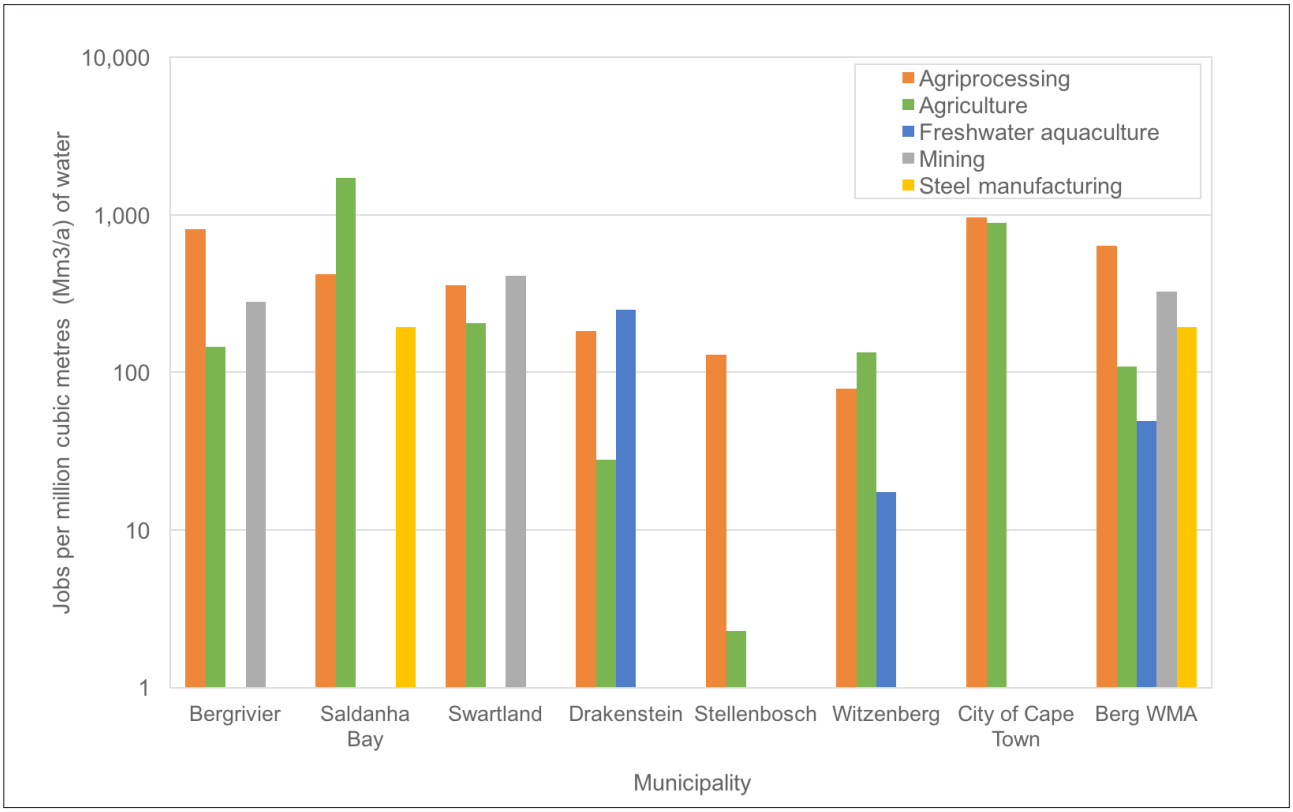


Figure 4: Total jobs per million cubic metres of water used for heavily water-dependent sectors in the Berg Water Management Area (WMA).

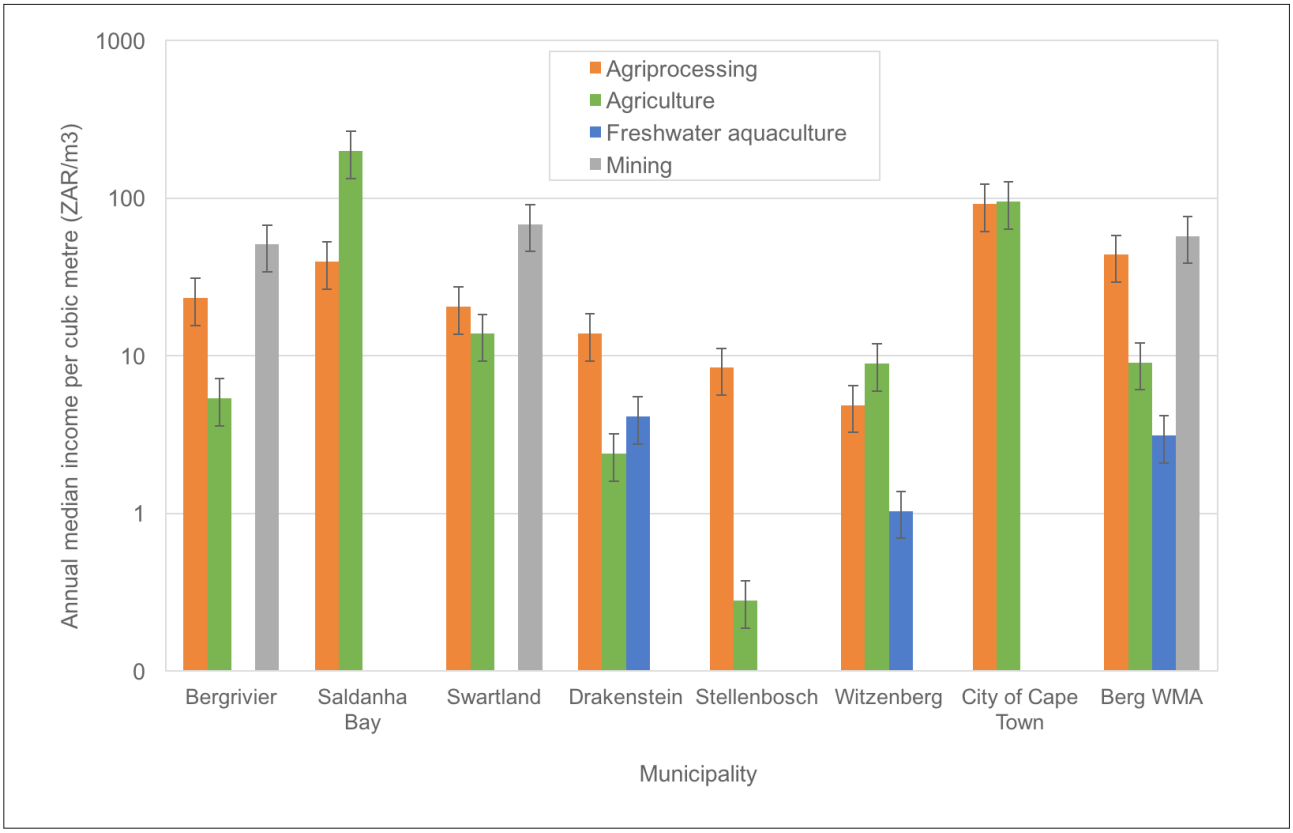


Figure 5: Total annual median income per cubic metre of water used for heavily water-dependent sectors in the Berg Water Management Area (WMA).

Water-use efficiency and equity through jobs and income

South Africa faces the 'triple challenge' of poverty, unemployment and inequality. Despite significant progress on social development since 1994, there has been little improvement in these three areas.[44] One of the main aims of this paper was to expand the concept of water-use efficiency to include jobs and income, and thus include individuals and households into efficiency considerations. Unfortunately, currently available water data are not sufficient to calculate accurate figures for jobs and income per unit of water, although the results provide the most comprehensive assessment to date and illustrate how these figures can be calculated in the future. The analysis would be greatly improved if StatsSA released the full industry Census 2011 data set (rather than only the 10% sample) and if heavy water users were required to report their annual water use figures. Accurate water-use data are most needed for agriculture and agriprocessing.

Water allocations to heavily water-dependent sectors create jobs and provide an income for the employees. The results show that 9% of jobs in the Berg WMA are heavily dependent on water and 16% are moderately dependent on water. This is largely influenced by the City of Cape Town, which accounts for 87% of the population, and water-dependence is much higher in the more rural Bergrivier LM (50% heavily dependent, 9% moderately dependent) and Swartland LM (31% heavily dependent, 11% moderately dependent). The biggest heavily water-dependent user by far is irrigated agriculture, followed by food processing, although there is significant uncertainty in the water-use data for agriprocessing. The jobs analysis shows that agriprocessing produces the most jobs per volume water used in the Berg WMA at 633 jobs/Mm^3, followed by mining (326 jobs/Mm^3), steel processing (195 jobs/Mm^3), agriculture (109 jobs/Mm^3) and aquaculture (49 jobs/Mm^3). The figures can vary significantly across municipalities within agriculture (from 2 to 1712 jobs/Mm^3) and agriprocessing (from 79 to 965 jobs/Mm^3). This variation is as a result of the diversity in the type of farming and crops farmed, and diversity in the type of agriprocessing facility. Water allocation decisions therefore should be made for specific sub-sectors and regions rather than for agriculture or agriprocessing as whole sectors. The two sectors are also directly linked so that the water allocated to agriculture supports jobs in both agriculture and agriprocessing.

The income analysis shows that mining produces the highest income per cubic metre of water used in the Berg WMA at ZAR57.52/m^3, followed by agriprocessing at ZAR43.73/m^3, agriculture at ZAR9.07/m^3 and freshwater aquaculture at ZAR5.20/m^3. Again, there is significant variation across the municipalities and significant uncertainty in the agriprocessing figures. Comparing the income analysis to the jobs analysis shows that some sectors, like mining, generate a few high-paying jobs, while other sectors, like agriculture, generate many lower-paying jobs. The water allocated to each sector therefore has different benefits and outcomes for the socio-economic well-being of the area. While jobs and income measure the benefit of water allocation to individual people, it is important to also measure the gross value add of different sectors as this measures the benefit to the region and the country through economic growth and indirectly to poverty reduction through social grants and service delivery.

The allocation of water resources to different areas and sectors will influence the potential for job creation and economic growth in the region. As the population grows and water resources come under increasing pressure, creating jobs with low dependence on water may become a priority or necessity. This change may have a positive relative impact on the City of Cape Town, as it is dominated by the low water use service industry, but a negative relative impact on small towns that rely on heavy water users like agriculture and mining. Water allocation decisions therefore must consider the long-term future of all residents. This consideration will require much better data collection on water use, employment and income in different sectors, GIS-based tools to facilitate spatial analysis, scenario development to analyse possible effects of different water allocation decisions, and stakeholder engagement with affected communities and businesses.

SDG 6 indicators and data disaggregation

While SDG 6.4.1 on water-use efficiency is described by the IAEG and UN-Water as 'the value added per water withdrawn in USD/m^3 over time of a given major sector (showing over time the trend in water-use efficiency)', this description only covers one aspect of water-use efficiency and focuses on businesses and industries. In this paper, the focus is on people and social development – examining how the allocation and use of water resources benefit individuals and households. For this reason, additional SDG 6.4 indicators are proposed on residential water use, and jobs and income per unit of water used by different industries. An indicator on water losses is included as it affects the water supply available to households. SDG indicator 6.4.1 is a new indicator with no pre-existing data or experience, hence no target has been set by the IAEG, and UN-Water argues that 'its strongest meaning appears when its values are compared over time'[12]. Comparing the data for the four additional indicators described in this paper would produce useful insights into the change in allocation and use of water in the Berg WMA over time. This comparison would require a similar analysis of Census 2001 and the annual town water use accounts for the 2001/2002 period.

A requirement of the SDGs is that data are disaggregated by gender, race, geographical location and other context-specific factors so that inequality is measured in all its forms. The IAEG propose that SDG 6.1 and 6.2 could be disaggregated by service level (safe/basic/none) (see Table 1), but they could also be disaggregated by geographical location, race and gender. The SDG 6.4 indicators proposed in this study can be disaggregated by geographical location, as illustrated. The jobs and income indicators can be disaggregated by race, gender and income bracket, if the individual census data records in the 10% sample are used. This analysis would be very useful in showing who in society is benefitting most from water used by agriculture and industry.

Data visualisation of water for sustainable development

The Sustainable Development Knowledge Network has developed and published a SDG Index to visualise the annual SDG status of each country.[45] They use a radar plot for each country to compare progress on the 17 goals, and to compare countries. The outer limit of the radar plot represents the highest score achieved by any country measured. SDG 6 is included as a single wedge in the radar plot. There would be value in developing a similar radar plot for SDG 6 itself to show comparative progress towards the individual targets. The data gathered in this case study are used to illustrate what this plot may look like. Figure 6 plots drinking water access in dwellings (SDG 6.1), household access to a flush toilet (SDG 6.2), drinking water quality (Blue Drop Score, 0–100), household ownership of a washing machine, jobs per million cubic metres of water used (proposed SDG 6.4), and income per cubic metre of water used (proposed SDG 6.4) in the municipalities of the Berg WMA. The underlying data are provided in Supplementary table 8. All these indicators relate to the direct and indirect individual benefits of water resources and their allocation and use in the region. This plot is a powerful visual tool that can quickly highlight variations and inequalities in social well-being and access to natural resources. Figure 6 shows that there is very little variation in access to drinking water and sanitation but significant variation in water-use efficiency across the Berg WMA. Thus, although most people have access to piped water, they do not all have access to the benefits of water resources. This variation can entrench social inequalities and hinder sustainable development.

Conclusion

Measuring how water contributes to sustainable development is an important goal of the SDGs, particularly for a water-scarce developing country such as South Africa. Appropriate indicators can support decision-making and highlight key issues on inequality, unemployment and sustainability. In this paper, additional indicators for SDG 6.4 on water-use efficiency are proposed that focus on how individuals and households benefit from allocations and the use of water resources. The results for the Berg WMA, in the southwest corner of South Africa, show a significant range in water-use efficiencies which have important implications for water allocation decisions. By making water-use efficiency about people as well as industries, the question of the equity of water allocations is raised.

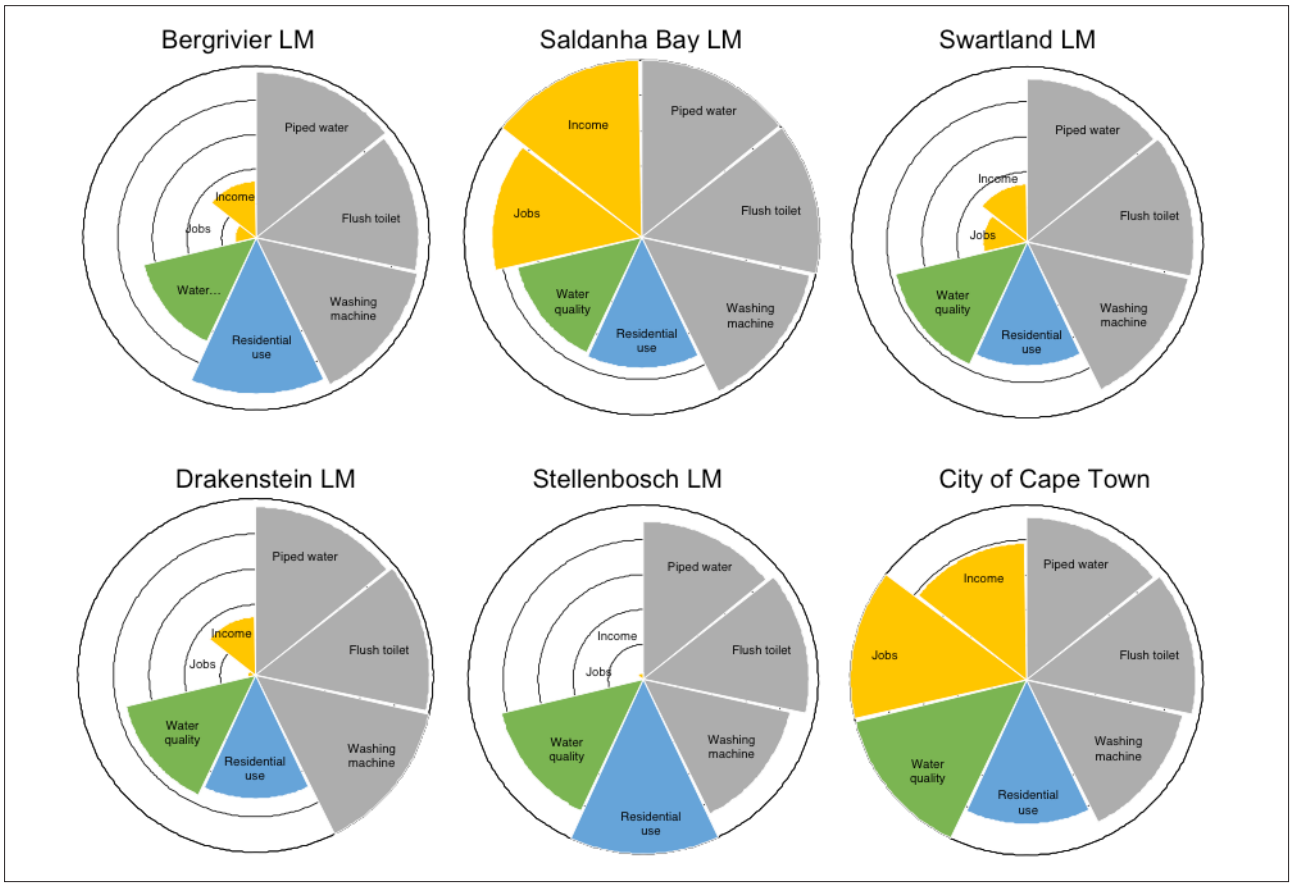


Figure 6: Barometers for water benefits for local municipalities (LM) in the Berg Water Management Area.

This approach to measuring the socio-economic benefits of water use is relevant for other countries seeking to measure the role that water plays in achieving inclusive sustainable development, and could be included in the final SDG 6 indicator suite.

Acknowledgements

This research was funded by the School of Geography and the Environment, University of Oxford. We thank the following people for their assistance with data collection: Jaco Human at iX engineers; Cornelius Riemann at Umvoto; Claire Pengelly and Pieter Janse van Vuuren at GreenCape; Helen Seyler at delta-h; Lynn Woolfrey and Takwanisa Machemedze at DataFirst UCT; Jane Turpie, Marthinus van der Walt and Jacob Muller at UCT; Jacyntha Twynam and Julius Okinor at WesGro; and Kersti Hickley and Leandre September at the City of Cape Town. We also thank John Dearing and the two anonymous reviewers for their helpful review.

Authors' contributions

M.J.C. was responsible for the conceptualisation, methodology, data collection, data analysis, data curation, writing the initial draft and writing revisions; R.M.B. was responsible for the conceptualisation, student supervision and writing revisions; J.D.S.C. was responsible for the conceptualisation, methodology and writing revisions; and M.G.N. was responsible for the conceptualisation, student supervision and writing revisions.

References

1. United Nations General Assembly. Transforming our world: The 2030 agenda for sustainable development. New York: United Nations General Assembly; 2015.
2. IAEG-SDGs. Tier classification for global SDG indicators [document on the Internet]. c2016 [cited 2017 Mar 27]. Available from: https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-05/Tier_Classification_of_SDG_Indicators_21_Dec_2016.pdf
3. United Nations Water. Integrated monitoring guide for SDG 6: Targets and global indicators. Geneva: United Nations Water; 2016.
4. United Nations Water. How to use the Integrated Monitoring Guide for SDG 6 (draft 20 April 2016). Geneva: United Nations; 2016.
5. United Nations General Assembly. Resolution adopted by the General Assembly on 28 July 2010 64/292. The human right to water and sanitation [document on the Internet]. c2010 [cited 2016 Nov 02]. Available from: https://www.un.org/es/comun/docs/?symbol=A/RES/64/292&lang=E
6. United Nations High Level Panel on Water. UN High Level Panel on Water action plan [document on the Internet]. c2016 [cited 2016 Nov 02]. Available from: https://sustainabledevelopment.un.org/HLPWater
7. United Nations World Water Assessment Programme. The United Nations World Water development report 2016: Water and jobs. Paris: UNESCO; 2016.
8. Burrows LR, Botha AP. Explaining the changing input-output multipliers in South African: 1980–2010. In: Proceedings of the Biennial Conference of the Economic Society of South Africa; 2013 September 25–27; Bloemfontein, South Africa. Pretoria: Economic Society of South Africa; 2013. p. 1–33.
9. South African Department of Water Affairs (DWA). National Water Resource Strategy June 2013. 2nd edition. Pretoria: DWA; 2013.
10. South African Department of Planning, Monitoring and Evaluation (DPME). Development indicators 2014. Pretoria: DPME; 2015.
11. Statistics South Africa. Statistical release P0211: Quarterly Labour Force Survey – Quarter 4: 2016. Pretoria: Statistics South Africa; 2017.

12. United Nations Water. Step-by-step monitoring methodology for Indicator 6.4.1 on water-use efficiency. Version 2017-03-10 [homepage on the Internet]. c2017 [cited 2017 June 28]. Available from: http://www.sdg6monitoring.org/news/indicators/641
13. South African Department of Water and Sanitation (DWS). Determination of water resource classes and associated resource quality objectives for the Berg Catchment. Status quo. Pretoria: DWS; 2016.
14. South African Department of Water Affairs and Forestry (DWAF). National Water Resource Strategy. Pretoria: DWAF; 2004.
15. South African Department of Water Affairs and Forestry (DWAF). National Water Resource Strategy: Appendix D. Pretoria: DWAF; 2004. p. D.1–D19.4.
16. South African Department of Water Affairs and Forestry (DWAF). Proposal for the establishment of the Berg Catchment Management Agency. Pretoria: DWAF; 2007.
17. Statistics South Africa. SuperCross census 2011. Version 1. Pretoria: Statistics South Africa; 2016.
18. Western Cape Provisional Treasury. Municipal economic review and outlook 2016. Cape Town: Western Cape Provincial Treasury; 2016.
19. Saldanha Bay Industrial Development Zone [homepage on the Internet]. c2017 [cited 2017 Feb 20]. Available from: http://www.sbidz.co.za
20. South African Department of Water and Sanitation (DWS). Support to the continuation of the Water Reconciliation Strategy for the Western Cape Water Supply System (WCWSS): Status report October 2014. Pretoria: DWS; 2014. Available from: https://www.dwa.gov.za/Projects/RS_WC_WSS/Docs/WCWSS Status Report Oct2014 Final.pdf
21. Seyler H. Water as a constraint on economic development: 2014-2015 Research project progress report. Cape Town: GreenCape; 2015.
22. Republic of South Africa. National Water Act. Act No 36 of 1998.
23. Howard G, Bartram J. Domestic water quantity, service level and health. Geneva: World Health Organisation; 2003.
24. Worley Parsons. Bergrivier Municipality annual WSDP performance and water services audit report for 2014/15. Cape Town: Bergrivier Municipality and Worley Parsons; 2015.
25. Worley Parsons. Drakenstein Municipality annual WSDP performance and water services audit report for 2014/15. Cape Town: Drakenstein Municipality and Worley Parsons; 2015.
26. Worley Parsons. Saldanha Bay Municipality annual WSDP performance and water services audit report for 2011/2012. Cape Town: Saldanha Bay Municipality and Worley Parsons; 2012.
27. Worley Parsons. Stellenbosch Municipality annual WSDP performance and water services audit report for 2010/211. Cape Town: Stellenbosch Municipality and Worley Parsons; 2012.
28. Worley Parsons. Swartland Municipality annual WSDP performance and water services audit report for 2014/15. Cape Town: Swartland Municipality and Worley Parsons; 2015.
29. South African Department of Water Affairs (DWA), Umvoto. Reconciliation strategy for Jamestown. Pretoria: DWA; 2008.
30. South African Department of Water Affairs (DWA), Umvoto. Reconciliation strategy for Raithby. Pretoria: DWA; 2008.
31. South African Department of Water Affairs (DWA). First level monitoring and status quo assessment for monitoring water allocation reform in South Africa. Pretoria: DWA; 2013.
32. Western Cape Government. Water infrastructure and opportunities for agriculture and agri-processing in the Western Cape. Cape Town: Western Cape Government Department of Environmental Affairs and Development Planning; 2015.
33. Statistics South Africa. Census 2011 metadata report no. 03-01-47. Pretoria: Statistics South Africa; 2012.
34. ArcelorMittal. ArcelorMittal Saldanha Works Analyst visit [document on the Internet]. c2011 [cited 2016 Dec 02]. Available from: http://www.arcelormittalsa.com/InvestorRelations/Presentations.aspx
35. PPC. PPC Investor site visit Western Cape 25 March 2011 presentation [document on the Internet]. c2011 [cited 2016 Oct 12]. Available from: http://ppc.co.za/investors/Investor-westernCape1.pdf
36. South African Department of Minerals and Energy (DME). Operating mines and quarries and mineral processing plants in the Republic of South Africa, 2006. Pretoria: DME; 2006.
37. ArcelorMittal. Saldanha Works overview [document on the Internet]. c2016 [cited 2016 Dec 01]. Available from: http://www.arcelormittalsa.com/Operations/SaldanhaWorks/Overview.aspx
38. Western Cape Department of Agriculture. CapeFarmMapper 2.0 [homepage on the Internet]. c2016 [cited 2016 Nov 09]. Available from: http://gis.elsenburg.com/apps/cfm/
39. CraftBru. Map of SA craft breweries [homepage on the Internet]. c2017 [cited 2017 Feb 04]. Available from: http://craftbru.com/breweries/south-africa/map/
40. Brown J. The rise of the craft breweries. City Press [Internet]. 2016 Feb 05; Available from: http://city?press.news24.com/Business/the?rise?of?the?craft?breweries?20160205
41. SABMiller, WWF-UK. Water footprinting: Identifying and addressing water risks in the value chain. Woking, UK: SABMiller Plc and WWF-UK; 2009.
42. South African Wine Industry Statistics (SAWIS). 2016 SA wine industry statistics Nr 40. Paarl: SAWIS; 2016.
43. South African Department of Agriculture, Forestry and Fisheries (DAFF). A profile of the South African dairy market value chain. Pretoria: DAFF; 2014.
44. Cole MJ, Bailey RM, New MG. Tracking sustainable development with a national barometer for South Africa using a downscaled "safe and just space" framework. Proc Natl Acad Sci USA. 2014;111(42):E4399–4408. http://dx.doi.org/10.1073/pnas.1400985111
45. Sachs J, Schmidt-Traub G, Kroll C, Durand-Delacre D, Teksoz K. SDG Index & dashboards – Global report. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN); 2016.

System usability scale evaluation of online banking services: A South African study

AUTHORS:
Mathias Mujinga[1]
Mariki M. Eloff[2]
Jan H. Kroeze[1]

AFFILIATIONS:
[1]School of Computing, College of Science, Engineering and Technology, University of South Africa, Johannesburg, South Africa
[2]Institute for Corporate Citizenship, University of South Africa, Pretoria, South Africa

CORRESPONDENCE TO:
Mathias Mujinga

EMAIL:
mujinm@unisa.ac.za

DATES:
Received: 02 Mar. 2017
Revised: 22 June 2017
Accepted: 03 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
security; privacy; ebanking usability; website usability; South Africa

HOW TO CITE:
Mujinga M, Eloff MM, Kroeze JH. System usability scale evaluation of online banking service: A South African study. S Afr J Sci. 2018;114(3/4), Art. #2017-0065, 8 pages. http://dx.doi.org/10.17159/sajs.2018/20170065

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

Online banking is a critical service offered by financial institutions to their clientele to facilitate easier and faster access to financial services and transactions. Banks currently spend huge amounts of money on development and maintenance of websites and backend systems that offer online banking facilities to clients. Here we address the effect of moderating factors on online banking usability assessment in South Africa. Using statistical analysis techniques that included *t*-tests, ANOVA and correlation, we investigated whether there are statistically significant mean differences in system usability scale (SUS) scores based on a variety of moderating factors in South Africa. Findings based on a sample of 540 respondents show that SUS scores differ significantly based on factors such as age, experience and income, whereas factors such as gender, use frequency and employment did not affect the mean SUS scores. Given the individual SUS scores for a variety of users based on different demographics, the financial institutions might improve service usability to target specific user groups and realise their return on investment in digital banking channels. Therefore improving service usability might go a long way in encouraging online banking adoption in South Africa.

Significance:

- The overall assessment of online banking service by users based on a SUS measurement tool was investigated.
- The effect of moderating variables on the mean SUS scores of different user groups was established.
- An insight into areas of improvement with regard to usability based on demographic information of users is provided.

Introduction

Online banking is a critical service offered by financial institutions to their clientele to facilitate easier and faster access to financial services and transactions. The service also allows banks to reduce client in-branch visits, thereby combating operational costs. Banks currently spend huge amounts of money on the development and maintenance of websites that offer online banking facilities to clients. To that end, there is a need to address both the security and privacy of the customers' financial and personal information while at the same time providing a system that users can actually use and enjoy. Usability of the channels that provide online banking is significantly important; the main channel is the website user interface, which is usually accessible through a plethora of devices.

This paper addresses the following research question: What is the effect of moderating factors on online banking usability assessment in South Africa? We aim to contribute to the body of knowledge by, firstly, investigating the overall assessment by users of online banking services based on a system usability scale (SUS) measurement tool, and, secondly, establishing the effect of moderating variables on the mean SUS scores of different user groups in the sample.

We investigated the possible effect of service usability as a contributing factor in service adoption based on different population groups. Given the individual SUS scores for a variety of users based on different demographics, the financial institutions might improve service usability to target specific user groups and realise their return on investment in digital banking channels.

Online banking

Since the late 1990s, the financial landscape has been changing through the use of Internet technologies.[1] Financial institutions are now providing clients with self-service options based on the Internet. These services are optimised for both mobile devices and desktop computers, considering limited bandwidth and the cost of Internet access in less developed communities. The emergence of electronic banking (ebanking) services such as virtual banking, home banking and online banking, which provide various banking activities through digital channels, has revolutionised the industry.[2]

The number of online banking users has been growing throughout the world, as the convenience of using online banking to perform banking transactions throughout the day has an edge over previous delivery channels, mainly a visit to a brick-and-mortar bank branch. Nonetheless, online banking has problems that still need to be addressed to achieve the full benefits of the service, such as usability[3] and security[4]. Firstly, as much as adoption is increasing, the rate of adoption is not the same across different economies. For instance, the rate of uptake in developed economies is significantly higher than in developing economies. Secondly, there are still security and privacy risks associated with conducting financial activities online that need to be addressed, as cited by previous studies.[4]

Online banking in South Africa

The population of South Africa is approximately 55.9 million[5], with an adult population of around 42.9 million[6], based on June 2016 estimates. The country has an Internet penetration rate of 52%, meaning that 28.5 million South Africans are Internet users.[7] Of the 36.8 million adults, 24.9 million (58%) have formal bank accounts.[6] Based on 2014 estimates, 14 million banking adults have Internet access, but only 2.3 million, that is, just more than 9% of bank account holders, use online banking services[8], which, in turn, accounts for only 4% of the total population. This represents a worryingly low level of adoption of the service. In comparison, the online banking adoption rate in the European Union for 2016 was 49% of the total population, with Norway having the highest rate of 91%.[9] Meanwhile, the global uptake of online banking was 28.7% of the total Internet audience, with the Middle East and Africa at 8.8%, based on the latest available data from 2012.[10]

The majority of South African banks currently provide free-of-charge online banking service in bundled banking packages to encourage clients' uptake of the service. Even individual transactions through digital channels incur no additional charges, as opposed to high fees for conducting the same transactions inside the branch. Regardless of these cost-effective initiatives, the use of online banking in South Africa is still low compared with that in other countries. Internet access is not a significant barrier to online banking adoption in South Africa, as a significant proportion of bank account holders already have access to the Internet, contrary to previous findings by Karjaluoto et al.[11] This suggests that, in addition to Internet access, there are other factors at play in low online banking adoption.

Online fraud has since been identified as a major deterrent to online banking adoption, with a Kaspersky Lab and B2B International survey[12] showing that, globally, 64% of people worry about online banking fraud, and yet only 21% believe they are a target for cyberattacks. Troublingly, only 60% use security solutions for any connected devices they own and use online.[12] With regard to adoption, usability of the user interface of the service might also play a role in inhibiting adoption and continued use.

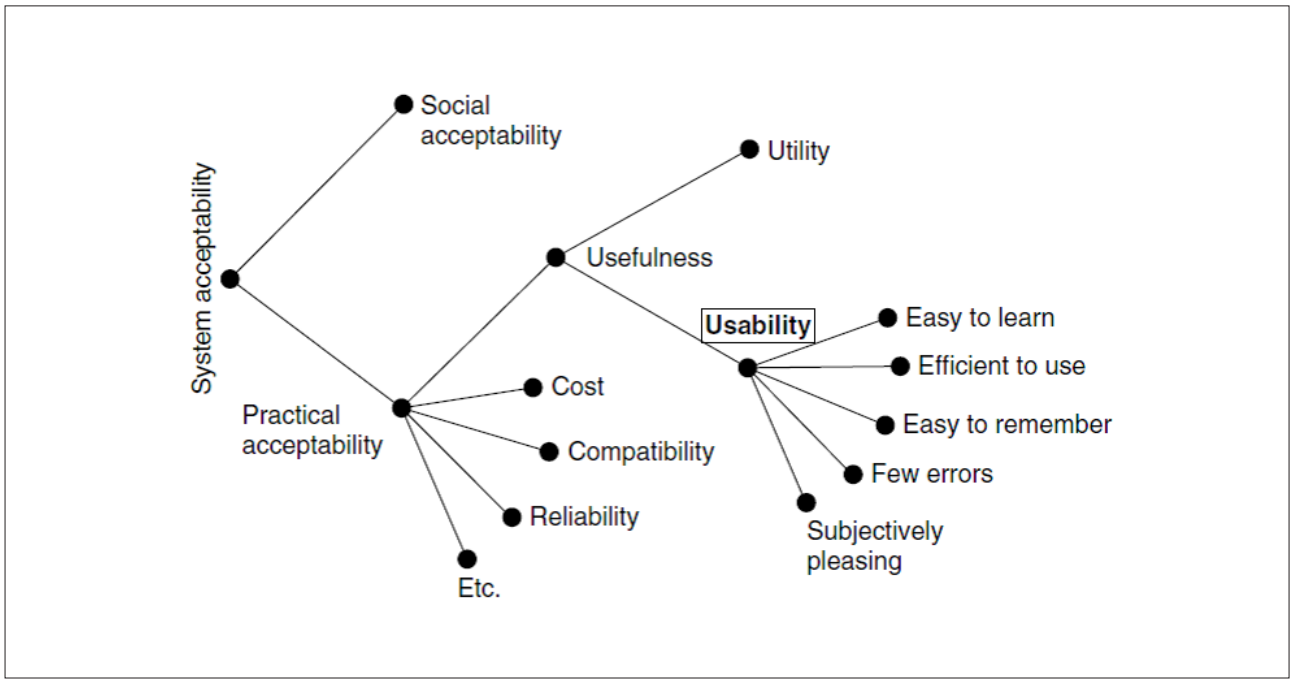
Usability

Usability – the reason why users love certain products or services that they use daily – is invisible. As Barnum[13] puts it: 'When usability is inherent in the products we use, it's invisible. We don't think about it. But we know it's there.' The absence of usability in a product or service brings about frustration; in extreme cases, users decide not to bother using the product or service. Preece et al.[14] note that most gadgets are engineered to work effectively, while neglecting usability aspects from the users' perspective. The same can be said of most software applications, even more so in information security systems. With this concern in mind, the resulting phenomenon has been the introduction of usability early in the design process, leading Mitnick and Simon[15] to allude to the notion that attackers are exploiting the human factors neglected by designers to gain access to computer systems.

The subjective nature of what constitutes usability and a wide variety of artefacts that apply the concept of usability have given rise to many definitions of the term 'usability'. One definition of usability that has become a standard is the usability process-oriented approach from ISO 9241-11[16], succinctly summarised by Bevan et al.[17] as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use'. The definition highlights three key aspects: specific users, specified goals and a specified context of use. As scholars and practitioners have become familiar with numerous aspects of usability, it has become apparent that usability is not a one-dimensional property of a user interface, but that it incorporates attributes such as learnability, efficiency, memorability, errors and satisfaction.[18]

Usability goals

There are numerous usability goals suggested by scholars, and some of them are prescribed for specific information technology systems and services. Rubin and Chisnell[19] argue that what makes a system *usable* is the 'absence of frustration' while interacting with it. The authors go on to state that when a product or service is truly usable, 'the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation, or questions'.[19] Figure 1 illustrates the position of usability in a wider system acceptability context.



Source: Nielsen[18]

Figure 1: System acceptability attributes.

Social acceptability and practical acceptability form part of system acceptability. Practical acceptability also considers aspects such as reliability, compatibility, cost and usefulness, to name but a few. Usefulness is explained in terms of utility and usability, where utility assesses whether the system provides the intended functionality while usability considers how well users can use the system functionality.[18] Figure 1 lists easy to learn, efficient to use, easy to remember, few errors and subjectively pleasing as the main usability goals. It is important to note that a product or service has to first be useful in completing a specific task before looking at how usable it is to the user.

Website usability

With regard to websites, usability is an important aspect in the design of successful online portals. To quote Nielsen[20], one of the main web usability authors: 'Usability rules the Web. Simply stated, if the customer can't find a product, then he or she will not buy it.' The same sentiment applies to services offered through online channels, including online banking. If users cannot find online banking functions provided by the website, chances are high that they will simply not use the website and will visit the branch or make a telephone call instead. Given huge investments by financial institutions in digital channels, it is important to use any means possible to help users migrate to online banking, and making online banking websites and interfaces usable is a major step in the right direction.

Website usability is critical in attracting new visitors and has a direct influence on customer satisfaction.[21] Users are unlikely to revisit a website that exhibits poor usability, and hence, customer loyalty to a website is significantly influenced by website usability.[22] Given the potential risk of financial loss, services such as online banking need to meet both usability and security goals. Measuring usability is often difficult, as usability of a product or service is highly subjective. As such, there are quite a number of ways to measure usability. One such measurement tool is the standardised SUS developed by Brooke[23]. We used the SUS tool to measure the usability of online banking services in South Africa. We report on the results of survey data collected from online banking users, based on 10 survey items that constitute the SUS measurement tool (see Appendix).

Related work

A number of studies have been conducted on factors affecting online banking adoption in individual economies across developed and developing economies. Most of these studies have been limited to identifying online banking benefits and inhibitors. Numerous factors such as service usefulness, ease of use, security, usability, lack of Internet infrastructure and Internet access costs have been brought to the fore. Apart from single-case studies, some scholars have performed a systematic analysis of studies on the topic of online banking adoption across cases to identify significant factors. The following paragraphs outline some of the recent studies in this area.

In India, Sikdar et al.[24] identified trust, usage constraint, ease of use, accessibility and intention to use as significant factors determining Internet banking adoption, while accessibility, usage constraints and intention to use displayed a strong and significant relationship with overall customer satisfaction. Szopiński[25] identified factors that determined the adoption of online banking in Poland as mainly the use of the Internet, taking advantage of other banking products and trust in commercial banks.

In the South African context, Brown and Molla[26] investigated the determinants of Internet and cellphone banking in 2005. Their findings identified a significant difference between factors affecting Internet and cellphone banking. It is important to note that the study was done during the initial stages of both Internet and cellphone banking in South Africa and that Internet and cellphone network coverage was not countrywide; in addition, the cost of data was exorbitant. Given the current overlap of online banking users accessing the service through multiple devices, such as mobile and desktop computers, no significant difference exists in adoption factors of the service between mobile or non-mobile devices. Since then, the technology landscape for both technologies has significantly changed.

A more recent study by Ramavhona and Mokwena[4] highlighted the lack of online banking adoption in rural South Africa, irrespective of efforts by financial institutions to promote the service. The study identified perceived security risk and lack of Internet infrastructure as the main inhibitors of adoption. Although security risk is still a concern countrywide, lack of Internet infrastructure is not necessarily a factor in urban areas. The authors cited the complexity of using the service in terms of lack of user-friendliness as a contributing factor in lack of adoption. In our previous study, we identified security as the main concern among online banking users, while service convenience was cited as the main attraction among current service users, mainly in Gauteng.[27]

Hanafizadeh et al.[28] conducted a systematic review of 165 Internet banking adoption studies between 1999 and 2012 and reported on a wide range of types of studies. The three types included studies that sought a description of the phenomenon, interplay between identified factors, and high-level comparison across populations, channels and methods. Another systematic review of 25 265 cases of online banking adoption by Montazemi and Qahri-Saremi[29] identified factors that had differing importance based on consumers' pre-adoption and post-adoption of the service. The findings reinforced the significance of original technology acceptance model constructs of perceived usefulness and perceived ease of use in online banking adoption. The authors also highlighted the significant influence of trust in online banking services as well as in the physical bank on the intention to use the service.

Besides determinants of online banking adoption, there is limited work on the relationship between usability and adoption of the service. Tassabehji and Kamala et al.[30] performed a SUS evaluation of an online banking biometric authentication system prototype. Their findings found the biometric system usable based on respondents' SUS scoring, with no major differences between gender and age groups. Although the findings were favourable in a prototype system, these cannot be generalised to a commercially developed system, as there are still major usability and privacy issues with biometric systems. Krol et al.[3] reported on the usability of a two-factor authentication mechanism for online banking that used hardware tokens. The respondents from the study highlighted a preference for biometric authentication because of the high degree of mental and physical workload and usability issues of hardware tokens.

Although users prefer biometric authentication systems for online banking and other online applications, there are still significant usability problems with biometric technology. For instance, Bhagavatula et al.[31] identified system reliability issues in popular smartphone devices' biometric technology. In general, biometric authentication methods are rarely used in online banking systems as a result of high cost and difficulty of use for people living with disabilities.[32] A study by Belanche et al.[21] showed that website usability affected satisfaction, which in turn had an impact on intention to use. Their findings concluded that usability did not necessarily affect intention to use, but had an indirect effect through consumer satisfaction.

This paper reports on a SUS assessment of the current online banking system in South Africa as provided by major retail banks to customers. The work provides a first insight into service usability, as no other such studies have been conducted in South Africa. Given the low adoption rate of online banking services in South Africa, we investigated the possible effect of the usability of the service on adoption decisions.

Research methodology

The study follows a positivist philosophy that is mainly associated with quantitative research that is focused on collecting quantitative data for theory testing to increase the predictive nature of understanding the phenomenon under investigation.[33] The study uses a deductive research approach that involves hypothesis testing evaluated through empirical observations.[34] Using a survey research method, an online questionnaire was distributed to online banking users, and 540 valid and usable responses were filled in by respondents from across all nine provinces of South Africa. The majority of responses came from clients of South

Africa's five major banks (ABSA, Capitec, FNB, Nedbank and Standard Bank). The respondents were current users of online banking services who evaluated the service based on the 10-item SUS measurement tool.

System usability scale

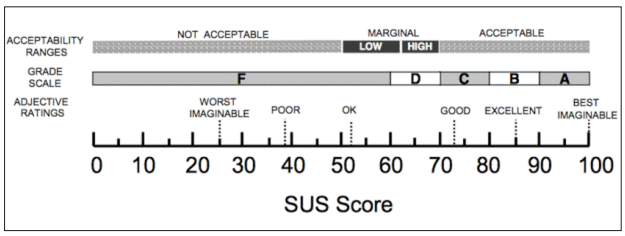
SUS is intended to provide an easy and quick subjective measure of usability. SUS is a standardised, inexpensive and reliable low-cost usability scale for system or product usability assessment.[35] SUS provides a quick and easy measurement of a user's subjective rating of a product or service usability. Brooke[23] developed SUS as a 10-statement measurement tool that uses a Likert scale for scoring SUS statements. An empirical systematic evaluation of SUS studies of over 10 years found the tool to be useful as a quick and easy method of measuring system usability.[36] We used a five-point Likert scale to measure the degree of agreement with SUS statements, from *strongly disagree* to *strongly agree,* scored from 1 to 5, respectively.

The SUS statements were modified to reflect online banking as the system under evaluation. A single SUS score indicates a composite measure of the system's overall usability as evaluated by the user, but individual item scores are meaningless on their own.[23] Hence, a variety of ratings was developed to interpret SUS scores and rate the usability of a system. The range of individual SUS item scores was from 0 to 4, covering the five-point Likert scale, as shown in Table 1.

Table 1: System usability scale scoring

Likert scale degree	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Scale position	1	2	3	4	5
Item score	0	1	2	3	4

SUS questionnaire items consisted of alternating five positive (numbered 1, 3, 5, 7 and 9) and five negative (numbered 2, 4, 6, 8 and 10) statements. The alternation between positive and negative statements was deliberate to avoid response biases.[37] Positive statements were scored by subtracting 1 from the Likert scale position, while negative statements were scored by subtracting the scale position from 5. Finally, the overall SUS score was obtained by multiplying the sum of all individual scores by 2.5 to obtain a score in the range 0–100.[37] The mean SUS score of all valid system evaluation responses gave an indication of the overall system usability. Bangor et al.[35] developed usability ratings based on mean SUS scores, as indicated in Figure 2.



Source: Bangor et al.[35]

Figure 2: Comparison of mean system usability scale (SUS) score ratings.

The figure shows three ratings, namely, acceptability ranges, grade scale and adjective ratings, all based on SUS score ranges. A SUS score below 50 indicates poor usability (not acceptable) of the artefact under investigation, while a score between 50 and 70 indicates marginal acceptability, and a score above 70 indicates an acceptable (good, excellent and best imaginable or better) level of usability.[35]

Ethical considerations

This study was reviewed and approved by the Research Ethics Committee of the College of Science, Engineering and Technology at the University of South Africa. The respondents were informed about the voluntary nature of participation, with the right to withdraw at any time, before submitting an anonymous online survey. The online survey provided a compulsory tick box for giving consent to participate. The results do not mention the bank names to avoid any undue prejudice, although these data were collected.

Empirical results

The collected data were prepared for analysis using IBM SPSS Statistics 24. The paper provides descriptive statistics of frequencies performed on the data, followed by more advanced techniques on SUS scores that included *t*-testing, analyses of variance (ANOVA) and correlation analyses.

Descriptive statistics

Table 2 shows the frequency statistics and mean SUS scores for different groups of respondents as extracted from the IBM SPSS statistical package. More than half of the 540 respondents (58.1%) are male, while 41.9% are female. Furthermore, the results indicated that a plurality of respondents (36.1%) are in the 30–39 year age group, followed by those aged 20–29 years (with 22.6%).

The majority of respondents (66.5%) had 5 years' or more experience using online banking, while 48.5% of them had more than 7 years' experience. A significant proportion of users used the service once a week (with 39.6%), followed by daily users at 34.3%. Further analysis revealed that the majority (83.8%) of 185 daily users mainly used a mobile device to access online banking, while almost half (49.7%) used all three devices (mobile devices, laptops and personal computers), and only 23.2% used a single device.

Mobile devices were the device most frequently used (73.7%) to access online banking, followed by personal computers (72.2%) and laptops (61.3%); 70% of respondents used more than one device to access online banking, while 37.4% used all three devices, and 30% used a single device.

These findings support the logical assumptions that mobile devices facilitate frequent access and that the more devices users have at their disposal, the more frequently they will use the service.

Mean SUS scores rating

With regard to grade scale and adjective rating, 11% of respondents scored their online banking as grade A, with SUS scores between 90 and 100, 25% rated it as *excellent* (grade B), and 34% considered the service to be *good* (grade C). With an overall mean SUS score of 73.8, the majority of respondents (70%) rated their online banking service as having a 'good' subjective rating and an 'acceptable' acceptability range. An analysis of the extreme low and high SUS scores showed that only 9 (2%) users considered the service to be 'best imaginable', with a perfect 100 SUS score, while 67 (12%) users gave SUS scores below 60 that resulted in a grade F, with 7% of users rating the service either as 'poor' or 'worst imaginable'. The extreme scores and the overall average rating show that there is significant room for improvement and a critical need to address usability issues of online banking in South Africa.

Inferential statistics

We performed a set of inferential statistical analyses to investigate the differences in mean SUS scores based on a set of study propositions. The overall null hypothesis for this study contended that there was no significant difference between mean SUS scores across individual moderating factors, including groups within these factors. To investigate this hypothesis, we used a series of analysis techniques, namely *t*-tests, ANOVAs and correlations, in order to make a decision on whether to reject or accept the null hypothesis.

Table 2: Frequencies and mean system usability scale (SUS) scores, $n = 540$

Factor	Category	Frequency	Percentage	Cumulative percentage	SUS score
Gender	Male	314	58.1	58.1	73
	Female	226	41.9	100.0	74
Age	Below 20 years	26	4.8	4.8	65
	20–29 years	122	22.6	27.4	73
	30–39 years	195	36.1	63.5	75
	40–49 years	88	16.3	79.8	74
	50 years or older	109	20.2	100.0	74
Experience	Below 1 year	39	7.2	7.2	76
	1–2 years	46	8.5	15.7	74
	3–4 years	96	17.8	33.5	74
	5–6 years	97	18.0	51.5	67
	7 years and above	262	48.5	100.0	77
Income (ZAR)	Less than 10 000	73	13.5	13.5	67
	10 000–19 999	68	12.6	26.1	71
	20 000–29 999	102	18.9	45.0	74
	30 000–39 999	130	24.1	69.1	77
	40 000–49 999	46	8.5	77.6	77
	50 000 or more	121	22.4	100.0	75
Use frequency	Every day	185	34.3	34.3	64
	Once a week	214	39.6	73.9	70
	Once in 2 weeks	71	13.2	87.1	72
	Once a month	65	12.0	99.1	74
	Other	5	0.9	100.0	77
Employment	Employed	431	79.8	79.8	75
	Self-employed	53	9.8	89.6	71
	Unemployed	28	5.2	94.8	63
	Retired	14	2.6	97.4	71
	Other	14	2.6	100.0	71
Device	Any one device	134	24.8	24.8	72
	Two or more devices	406	75.2	100.0	74

Independent samples *t*-test

We performed *t*-tests on mean SUS scores as opposed to the actual Likert scale scores, as these are meaningless before conversion. Independent samples *t*-tests work with variables that have, at most, two categories; in this study, only gender and device satisfied this criterion. This allowed the testing of hypotheses that compared two groups in the population for a decision on whether to accept or reject the null hypothesis.

Gender

The results of two independent samples *t*-tests, Levene's test for equality of variances and the *t*-test for equality of means, are illustrated in Table 3. The results showed that the *p*-value of Levene's test was 0.817, which was greater than the threshold value of 0.05, so we accepted the null hypothesis that there was no significant difference between the means of Levene's test and concluded that the variance in SUS scores for male and female respondents was equal; hence, the difference was statistically insignificant, which meant that we had to consider the output of equal variance assumed in order to test for equality of means.

Device

Table 4 shows independent samples *t*-tests performed to compare mean SUS scores between users who used 'any one device' and those who used 'two or more devices'. The mean SUS scores difference of -2.5117 shows that there was a significant difference, but further tests were necessary to investigate if the difference was statistically significant.

One-way ANOVA

We performed one-way ANOVAs for the categorical moderating factors, to compare the means of two or more independent groups in order to determine whether there was statistical evidence that the associated population means were significantly different. One-way ANOVA and *t*-testing are equivalent tests that test mean differences between groups; however one-way ANOVA goes a step further to provide the capability to analyse more than two groups, while *t*-testing compares only two groups.[38] The test statistic for a one-way ANOVA is denoted as *F*. Table 5 summarises the one-way ANOVA *F*- and *p*-values for moderating effects on mean SUS scores (between groups values), including post-hoc tests for group comparisons. The factors are ordered in order of statistical significance.

All five age groups exhibited statistically significant differences, with *p*-values ranging between 0.001 and 0.022 ($p<0.05$) with 'between groups' values of ($F = 3.925$, $p = 0.004$). This finding led to the conclusion that SUS scores significantly differed based on online banking users' age, with older users scoring relatively higher than younger users. This trend was also supported by the plot graph, although the 40–49 years age group scored a little lower than the preceding 30–39 years age group.

Statistically significant differences in mean SUS scores were shown for experience among all groups, except 'less than 12 months' versus '1–2 years'. This finding was supported by the plot chart that showed SUS scores increasing as experience increased. It follows that more experienced users were more likely than novice users to score a system highly in usability evaluation assessments. Therefore, our findings are in line with previous studies which concluded that as users became familiar with an interface, they considered it more usable because they could locate functions more easily.[39] We can conclude that website and interface usability improve with increased familiarity, as users become less and less frustrated with the system because they can locate and use functions more quickly and easily with time.

When contrasting income groups against the 'less than ZAR10 000' group, all comparisons were statistically significant, except the one against the 'ZAR20 000–ZAR29 999' group. These findings suggested that users with a higher income tended to have higher perceptions of the usability of online banking service compared to their lower-income counterparts.

Although one-way ANOVA mean SUS scores between groups for 'use frequency' showed a statistically significant difference with $p<0.05$, only one group comparison between the 'daily' and 'once a month' groups was found to be statistically significant, with $p=0.001$. The same applied to 'experience', with only one group comparison between 'unemployed' and 'employed' being statistically significant with $p=0.001$. Although data were collected for additional moderating factors such as education, ethnicity, language, province and bank, these were excluded from the analysis because the differences in means for SUS scores were statistically insignificant among the respective group comparisons ($p>0.05$ in one-way ANOVA). Therefore, for these factors we accept the null hypothesis and conclude that there is no significant difference in mean SUS scores between groups.

Table 3: Independent samples test: Gender

		Levene's test for equality of variances		*t*-test for equality of means						
		F	Significance	*t*	d.f.	Significance (two-tailed)	Mean difference	Standard error difference	95% confidence interval of the difference	
									Lower	Upper
System usability scale (SUS) scores	Equal variances assumed	0.054	0.817	-0.948	538	0.343	-1.063	1.121	-3.266	1.139
	Equal variances not assumed			-0.948	484.3	0.344	-1.063	1.122	-3.267	1.141

The test for equality of means provided the actual independent samples t*-test. The mean difference in SUS scores between genders of 1.063 at 0.05 level* (p*=0.343) meant that we could accept the null hypothesis that there was no significant difference in SUS scores based on gender.*

Table 4: Independent samples test: Device

		Levene's test for equality of variances		*t*-test for equality of means						
		F	Significance	*t*	d.f.	Significance (two-tailed)	Mean difference	Standard error difference	95% confidence interval of the difference	
									Lower	Upper
System usability scale (SUS) scores	Equal variances assumed	4.627	0.032	-1.967	538	0.050	-2.5117	1.277	-5.020	-0.003
	Equal variances not assumed			-1.817	200.8	0.071	-2.5117	1.383	5.238	0.2148

We test the null hypothesis that the mean difference in SUS scores is insignificant for users using one device and those using at least two devices. Levene's test for equality of variances has p*=0.032, which is lower than the threshold value of 0.05. Therefore, we reject the null hypothesis of Levene's test and conclude that the variance in SUS scores between the two groups is significant, and as a consequence we consider the row that assumes unequal variances. The output for equal variance not assumed indicates that the difference in means is statistically insignificant at* p*=0.071 (*t*= -1.817,* p*>0.05).*

Table 5: One-way ANOVA of system usability scale (SUS) scores versus moderating factors

Factor	d.f.	F	Signfiicance	Post-hoc tests (multiple comparisons)	
Age	4	3.925	0.004	All	$p < 0.05$
Experience	4	10.338	0.000	Less than 1 year versus 1–2 years	$p < 0.05$
Income	5	7.524	0.000	Less than ZAR10 000 versus ZAR10 000–ZAR19 999	$p < 0.05$
Use frequency	4	5.991	0.000	Daily versus once a month	$p > 0.05$
Employment	4	7.497	0.000	Unemployed versus employed	$p > 0.05$

Correlation

Correlation analysis was implemented to specify the strength and direction of relationships between the dependent variable SUS scores and independent categorical variables. A correlation coefficient (*r*) indicates the degree of relation among two or more variables, and it ranges between -1.00 and +1.00 (inclusive). A value of +1.00 means that there is a direct (positive) relation between the two variables, meaning that as one variable increases, the other increases.[40] Graphically, this is shown by a positive gradient. A value of -1.00 indicates an inverse (negative) relation. When there is no correlation between variables, the coefficient value is 0. Table 6 shows the results of the correlation analysis performed; interpreted findings are highlighted in bold font.

According to Pallant[41], a relationship that has a Pearson correlation value (magnitude) below 0.3 is weak, while a Pearson correlation value between 0.3 and 0.5 is termed moderate, and anything with a magnitude between 0.5 and 1 is termed strong. Table 6 indicates that the relationships between SUS scores and the categorical moderating factors income, experience, use frequency, and employment were significant at the 0.01 level (two-tailed), while age was statistically significant at the 0.05 level. Although there is a correlation between these moderating factors and SUS scores, the value of $r < \pm 0.3$ indicates weak relationships.[41] Income and experience had a positive weak relationship with SUS scores, while use frequency and employment exhibited a negative weak relationship with SUS scores. Additionally, the correlations between 'income versus age', 'experience versus age', and 'experience versus income' showed positive, moderately strong relationships at $p < 0.01$.

Table 6: Correlation analysis between system usability scale (SUS) scores and categorical variables

		SUS scores	Age	Experience	Income	Use frequency	Employment
SUS scores	Pearson correlation	1					
	Significance (two-tailed)						
Age	Pearson correlation	**0.090***	1				
	Significance (two-tailed)	**0.036**					
Experience	Pearson correlation	**0.259****	**0.464****	1			
	Significance (two-tailed)	**0.000**	**0.000**				
Income	Pearson correlation	**0.192****	**0.495****	**0.556****	1		
	Significance (two-tailed)	**0.000**	**0.000**	**0.000**			
Use frequency	Pearson correlation	**-0.176****	-0.017	**-0.196****	**-0.212****	1	
	Significance (two-tailed)	**0.000**	0.699	**0.000**	**0.000**		
Employment	Pearson correlation	**-0.167****	**-0.119****	**-0.365****	**-0.382****	**0.189****	1
	Significance (two-tailed)	**0.000**	**0.006**	**0.000**	**0.000**	**0.000**	

*Correlation is significant at the *0.05 level or the **0.01 level (two-tailed)*

Interpreted findings are highlighted in bold font.

Conclusion

We report on the first SUS assessment of online banking services in South Africa. With a sample size of 540, the results of the statistical analyses prove the findings are statistically significant to be generalised to the population at large. The moderating factors analysed in the study had varying effects on mean SUS scores. Using independent samples *t*-tests, gender did not have a statistically significant influence on mean SUS scores while device had statistical significance based on Levene's test but not with *t*-test equality of means. ANOVA results showed statistically significant mean differences for all age group comparisons. While the comparisons among income and experience groups were statistically significant for most groups, only one group comparison from each variable was found to be statistically insignificant. The comparison among use frequency and employment groups found that most group comparisons had mean differences that were statistically insignificant, with both variables having only one group comparison that was statistically significant.

The study highlights that, generally, younger users rated online banking service in South Africa poorly compared to older users. This can partly be explained by a lack of experience and familiarity with the user interface, as older and higher-income users tended to use the service more frequently. We also assume that with the prevalence of social media applications, younger users might be rating online banking against applications with better usability and user experience. This paper contributes to the development of the literature with regard to the usability of online banking services and adoption decisions in the context of the South African financial environment. Given the diverse group of online banking users and the one-size-fits-all service approach, insight into the usability evaluation of the service by users is critical in improving the service and making sure that it is used effectively to maintain a secure online banking environment.

The paper is limited to the usability assessment based on the SUS measurement tool, which comprises 10 items to measure system usability. Therefore, a more comprehensive usability assessment of online banking can be obtained by using an in-depth survey tool that evaluates individual usability principles, thereby yielding a more insightful analysis of the phenomenon. Another limitation is the fact that perceptions were solicited from current users of the service; it would be helpful to get the views of non-users of the service on their reasons for non-adoption. There is a need to conduct an in-depth investigation to identify factors that lead to younger users scoring usability poorly. This can be achieved by conducting an in-depth usability evaluation that specifically investigates a variety of usability principles identified in the literature. Financial institutions can then address these specific factors to improve overall service usability, which might improve overall service adoption.

Acknowledgements

This article is based on doctoral research by Mathias Mujinga. The research was supported in part by the National Research Foundation (NRF) of South Africa. The University of South Africa (UNISA) is acknowledged for granting a period of research and development leave to the first and third authors during which time the article was completed. Any opinion, finding, conclusion and recommendation expressed is that of the authors, and neither the NRF nor UNISA accepts any liability in this regard.

Authors' contributions

M.M. was responsible for the conceptualisation, methodology, data collection, data analysis, writing the initial draft, and writing revisions. M.M.E. was responsible for validation, critically reviewing the writing and revisions, student supervision and project leadership. J.H.K. was responsible for validation, critically reviewing the writing and revisions, and student supervision.

References

1. Gkoutzinis AA. Internet banking and the law in Europe: Regulation, financial integration and electronic commerce. London: Cambridge University Press; 2006. https://doi.org/10.1017/cbo9780511494703.003
2. Turban E, Rainer RK, Potter RE. Introduction to information technology. 3rd ed. Hoboken, NJ: John Wiley and Sons; 2005.
3. Krol K, Philippou E, De Cristofaro E, Sasse MA. "They brought in the horrible key ring thing!": Analysing the usability of two-factor authentication in UK online banking. In: Proceedings of USEC 2015: NDSS Workshop on Usable Security; 2015 February 08–11; San Diego, CA, USA. Reston, VA: Internet Society; 2015. p. 1–10.
4. Ramavhona TC, Mokwena S. Factors influencing Internet banking adoption in South African rural areas. S Afr J Inf Manage. 2016;18(2):1–8. https://doi.org/10.4102/sajim.v18i2.642
5. Stats SA. Mid-year estimates 2016 [document on the Internet]. c2016 [cited 2017 Feb 14]. Available from: http://www.statssa.gov.za/?p=8176
6. FinMark Trust. FinScope South Africa 2016 [document on the Internet]. c2016 [cited 2017 Nov 06]. Available from: https://www.finmark.org.za/results-from-finscope-south-africa-2016-survey-on-financial-inclusion/
7. Internet Live Stats. South Africa Internet users [document on the Internet]. c2017 [cited 2017 Nov 06]. Available from: http://www.internetlivestats.com/internet-users/south-africa/

8. Van Zyl G. FNB rated SA's 'top internet banking provider'. Fin24. 2015 May 07;Tech News [cited 2017 Nov 06]. Available from: http://www.fin24.com/Tech/News/FNB-rated-SAs-top-internet-banking-provider-20150507.
9. Statista. Online banking penetration in selected European markets in 2016 [homepage on the Internet]. c2017 [cited 2017 Nov 06]. Available from: https://www.statista.com/statistics/222286/online-banking-penetration-in-leading-european-countries/
10. Statista. Global online banking penetration in April 2012, by region [homepage on the Internet]. c2012 [cited 2017 Nov 06]. Available from: https://www.statista.com/statistics/233284/development-of-global-online-banking-penetration/
11. Karjaluoto H, Mattila M, Pento T. Electronic banking in Finland: Consumer beliefs and reactions to a new delivery channel. Journal of Financial Services Marketing. 2002;6(4):346–361. https://doi.org/10.1057/palgrave.fsm.4770064
12. Kaspersky Lab and B2B International. Consumer security risks survey 2016: Connected but not protected. USA: Kaspersky Lab and B2B International; 2016.
13. Barnum CM. Usability testing essentials: Ready, set... test! Burlington, MA: Morgan Kaufmann; 2011. https://doi.org/10.1109/tpc.2011.2159642
14. Preece J, Rogers Y, Sharp H. Interaction design: Beyond human-computer interaction. 4th ed. Chichester: John Wiley & Sons; 2015.
15. Mitnick KD, Simon WL. The art of deception: Controlling the human element of security. Indianapolis, IN: Wiley Publishing; 2002.
16. ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs): Guidance on usability. Geneva: International Organization for Standardization; 1998.
17. Bevan N, Carter J, Harker S. ISO 9241-11 revised: What have we learnt about usability since 1998? In: Kurosu M, editor. Human-computer interaction: Design and evaluation. Los Angeles, CA: Springer; 2015. p. 143–151. https://doi.org/10.1007/978-3-319-20901-2_13
18. Nielsen J. What is usability? In: Wilson C, editor. User experience re-mastered: Your guide to getting the right design. Burlington, MA: Morgan Kaufmann; 2010. p. 3–22. https://doi.org/10.1016/B978-0-12-375114-0.00004-9
19. Rubin J, Chisnell D. Handbook of usability testing: How to plan, design and conduct effective tests. 2nd ed. Indianapolis, IN: John Wiley & Sons; 2008.
20. Nielsen J. Designing web usability: The practice of simplicity. Berkeley, CA: New Riders Publishing; 2000.
21. Belanche D, Casaló LV, Guinalíu M. Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk. Journal of Retailing and Consumer Services. 2012;19(1):124–132. https://doi.org/10.1016/j.jretconser.2011.11.001
22. Flavián C, Guinalíu M, Gurrea R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Inf Manage. 2006;43(1):1–14. https://doi.org/10.1016/j.im.2005.01.002
23. Brooke J. SUS – a quick and dirty usability scale. Usability Eval Industry. 1996;189(1996):194–200.
24. Sikdar P, Kumar A, Makkad M. Online banking adoption: A factor validation and satisfaction causation study in the context of Indian banking customers. Int J Bank Marketing. 2015;33(6):760–785. https://doi.org/10.1108/IJBM-11-2014-0161
25. Szopiński TS. Factors affecting the adoption of online banking in Poland. J Bus Res. 2016;69(11):4763–4768. https://doi.org/10.1016/j.jbusres.2016.04.027
26. Brown I, Molla A. Determinants of internet and cell phone banking adoption in South Africa. The Journal of Internet Banking and Commerce. 2005;10(1):1–9.
27. Mujinga M, Eloff MM, Kroeze JH. Online banking users' perceptions in South Africa: An exploratory empirical study. In: Proceedings of the IST-Africa Conference 2016; 2016 May 11–13; Durban, South Africa. Durban: IIMC; 2016. p. 1–7. https://doi.org/10.1109/istafrica.2016.7530617
28. Hanafizadeh P, Keating BW, Khedmatgozar HR. A systematic review of internet banking adoption. Telematics Inf. 2014;31(3):492–510. https://doi.org/10.1016/j.tele.2013.04.003
29. Montazemi AR, Qahri-Saremi H. Factors affecting adoption of online banking: A meta-analytic structural equation modeling study. Inf Manage. 2015;52(2):210–226. https://doi.org/10.1016/j.im.2014.11.002
30. Tassabehji R, Kamala MA. Evaluating biometrics for online banking: The case for usability. Int J Inf Manage. 2012;32(5):489–494. https://doi.org/10.1016/j.ijinfomgt.2012.07.001
31. Bhagavatula C, Ur B, Iacovino K, Kywe SM, Cranor LF, Savvides M. Biometric authentication on iPhone and Android: Usability, perceptions, and influences on adoption. In: Proceedings of the USEC Conference; 2015 February 08; San Diego, CA, USA. Reston, VA: Internet Society; 2015. p. 1–10. https://doi.org/10.14722/usec.2015.23003
32. Althobaiti MM, Mayhew P. Security and usability of authenticating process of online banking: User experience study. In: Proceedings of the 2014 International Carnahan Conference on Security Technology (ICCST); 2014 October 13–16; Rome, Italy. IEEE; 2014. p. 1–6. https://doi.org/10.1109/CCST.2014.6986978
33. Turban E, King D, Lee JK, Liang T, Turban DC. Electronic commerce: A managerial and social networks perspective. London: Springer; 2015.
34. Gray DE. Doing research in the real world. 3rd ed. London: Sage Publications; 2014.
35. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud. 2009;4(3):114–123.
36. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact. 2008;24(6):574–594. https://doi.org/10.1080/10447310802205776
37. Brooke J. SUS: A retrospective. J Usability Stud. 2013;8(2):29–40.
38. Hair Jr JF, Hult GTM, Ringle C, Sarstedt M. A primer on partial least squares structural equation modeling (PLS-SEM). 2nd ed. London: Sage Publications; 2016.
39. Arfaa J, Wang YK. A usability study on elder adults utilizing social networking sites. In: Proceedings of the International Conference of Design, User Experience, and Usability; 2014 June 22–27; Crete, Greece. Crete: Springer International Publishing; 2014. p. 50–61. https://doi.org/10.1007/978-3-319-07626-3_5
40. Antonius R. Interpreting quantitative data with IBM SPSS statistics. 2nd ed. London: Sage Publications; 2013.
41. Pallant J. SPSS survival manual. 5th ed. Berkshire: McGraw-Hill Education; 2013.

Appendix: System usability scale (SUS) measurement tool

SUS items	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
I think that I would like to use online banking more frequently					
I find online banking unnecessarily complex					
I thought online banking was easy to use					
I need the support of a technical person to be able to use online banking					
I found the various functions in online banking to be well integrated					
I thought there was too much inconsistency in online banking					
I would imagine that most people would learn to use online banking websites very quickly					
I find online banking very difficult to use					
I feel confident using online banking					
I needed to learn a lot of things before I could get going with online banking					

Impact of a biomedical research project on the human capital development of emerging researchers

AUTHORS:
Christabelle S. Moyo[1]
Joseph Francis[2]
Pascal O. Bessong[1]

AFFILIATIONS:
[1]Department of Microbiology, University of Venda, Thohoyandou, South Africa
[2]Institute for Rural Development, University of Venda, Thohoyandou, South Africa

CORRESPONDENCE TO:
Christabelle Moyo

EMAIL:
christabellem6@gmail.com

DATES:
Received: 24 Mar. 2017
Revised: 18 July 2017
Accepted: 06 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
skills; knowledge; community-based research; rural-based university

HOW TO CITE:
Moyo CS, Francis J, Bessong PO. Impact of a biomedical research project on the human capital development of emerging researchers. S Afr J Sci. 2018;114(3/4), Art. #2017-0090, 6 pages. http://dx.doi.org/10.17159/sajs.2018/20170090

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Fogarty International Center of the National Institutes of Health (USA; award no. D43 TW009359)

Various evaluative studies have been carried out to obtain the views of multiple stakeholders involved in community-based biomedical research projects. However, rarely have the viewpoints of postgraduate students and junior faculty involved in such initiatives been explored. Thus, the aim of this study was to examine the views of postgraduate students and junior faculty at a rural-based university on the effect of a longitudinal biomedical research project on their acquisition of relevant skills. In-depth interviews and a focus group discussion were conducted. The thematic content analysis technique was used to analyse the qualitative data. Both postgraduate student and junior faculty groups indicated that they had acquired considerable research skills and knowledge; gained experience; were exposed to practical reality; and strengthened their interpersonal skills and general personal development. However, some respondents highlighted that they still believed that training in data analysis and exposure to new laboratory techniques would have strengthened their individual capabilities to conduct cutting-edge research. The results of this study highlight the need for community-based biomedical researchers to equip members of their teams with the skills and knowledge that will help them achieve their academic and career goals.

Significance:

- The importance of engaging study participants to get their views on the 'hidden' value of biomedical research projects has been identified.
- The need for community-based researchers to equip members of their teams with relevant research skills and knowledge has been emphasised.
- The results of this study can be useful in the planning and implementation of similar projects in the future.

Introduction

Various evaluative studies have been conducted in different parts of the world, which aim to elicit multiple stakeholder viewpoints and experiences. Among these have been studies focused on community-based biomedical projects or programmes.[1-4] Results obtained in some of the aforementioned studies have highlighted the importance of effective engagement of study participants in projects carried out in their localities. Engagement can be partially achieved through creating communication platforms in which project participants share their experiences and thoughts regarding research projects conducted in their communities. Through such interactive processes, researchers and development practitioners would gather information that might help in planning and implementing future initiatives. Although numerous studies on community-based biomedical research projects have been conducted, the viewpoints of postgraduate students and junior faculty involved in such initiatives have rarely been explored. Exploring the perspectives of postgraduate students and junior faculty might assist research project leaders to better plan similar initiatives and to take some corrective action when the need arises. Moreover, studies of this nature are likely to reveal the effects of community-based biomedical research projects beyond merely generating scientific knowledge. In addition, exploring the 'hidden' issues of biomedical research projects has the potential to bring about a holistic understanding of community-based projects and the benefits accruing to members of study teams. The preceding arguments necessitated carrying out the current study. In this study, the perspectives of postgraduate students and junior faculty on the effects of a biomedical community research project on their human capital development were explored.

Marimuthu et al.[5] define human capital as 'processes that relate to training, education and other professional initiatives in order to increase the levels of knowledge, skills, abilities, values and social assets of an employee which will lead to the employee's satisfaction and performance and eventually on a firm's performance'. The Organisation for Economic Co-operation and Development[6] concurs with these views and defines human capital as knowledge, skills, competencies and attributes embodied in individuals that assists in the creation of personal, social and economic well-being. Included in this definition are non-economic attributes such as motivation, communication, inter-personal skills, perseverance, self-discipline, behaviour, and physical, emotional and mental health of individuals.[6] Thus, investing in human capital is a strategy that helps increase economic returns.[5] Human capital can be acquired either through formal early childhood learning, primary and high schooling, tertiary and post-tertiary training, adult education or through informal education such as family, peers or on the job learning.[6] It is evident from these perspectives that education is a primer for development. In the current study, human capital refers to the acquisition of skills and knowledge, experience and exposure, interpersonal skills and personal development through informal post-tertiary training. Human capital attributes such as motivation, confidence building and working independently are also included in the definition.

Human capital theory, which is rooted in the works of Schultz and Becker[7], anchored the current study. Becker[8] argues that human capital is linked to economic growth and that education, training and health are the crucial elements of human capital. Becker[8] also points out that education, training and health are important investments that precipitate human capital, because they assist in raising income, improving well-being and enhancing the good habits of an individual over their lifetime. Education and training increase the chances of being employed.[9]

When individuals acquire relevant skills and knowledge, they often make better informed decisions about their lives, such as those relating to career path, financial investments, how to live amicably with others, eradicating poverty, nutrition, healthy lifestyles and various other ways of improving their well-being.[5,6,8] Nhamo and Nhamo[10] postulate that education enhances the potential earning capacity of individuals, given that enlightened individuals tend to be more productive and are able to perform more complex tasks. Mincer[11] lends weight to this view and states that employers set better rewards for employees with higher qualifications because they are more skilled and, in general, are more productive. It is also worth noting that globalisation is now heightening competition for skilled human resources.[12] Personnel with higher educational qualifications, relevant expertise and experience can be recruited from any part of the world. Thus, investments in human capital are crucial.

The current study was embedded in the 'Etiology, Risk Factors, and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development' (MAL-ED) South Africa project, which was initiated in 2009 and undertaken in the Dzimauli community of the Vhembe District in the Limpopo Province of South Africa. Bessong et al.[13] and the MAL-ED Network Investigators implemented the MAL-ED project. Postgraduate students and junior faculty from a rural-based university were recruited and trained to collect data on feeding habits, childhood diseases, vaccinations and biospecimens, among others. As highlighted earlier, the aim of the current study was to interrogate the perspectives of postgraduate students and junior academic staff who participated in the MAL-ED project with respect to the impact of the project on their human capital development.

Description of the study site

The study was conducted between 1 January and 28 February 2016 at a rural-based university in Thohoyandou in the Limpopo Province of South Africa. The study site lies approximately 180 km north of Polokwane, the capital city of the Limpopo Province.

Research methodology

Research design, population and sampling

A case study research design was adopted. This study was exploratory and qualitative. According to Kothari[14], qualitative research entails obtaining subjective assessments of attitudes, opinions and behaviour. Santha et al.[15] also explain that this type of research is concerned with the opinions, experiences and feelings of individuals, implying that subjective data are collected. It was deemed appropriate to adopt the qualitative research approach in the current study because it allowed the investigators to interrogate the perceptions of participants with a view of getting deeper insights on their acquisition of relevant skills.

Out of a target population of 34 people, 15 postgraduate students (pursuing honours, master's and doctoral studies), 7 junior faculty and 3 heads of departments (HoDs) voluntarily participated in the study. Of the 25 people who volunteered to participate in the study, 23 of them had been involved in the initial phase (24 months of follow-up) of the MAL-ED project. The remaining two participants were HoDs who had not been directly involved in the project but supported junior faculty from their departments to actively participate. It was deemed necessary to obtain their views on the academic performance of the members of staff from their departments who had been involved in the MAL-ED project. Out of the 15 postgraduate students who participated in the study, 11 took part in the individual face-to-face interviews. The remainder were interviewed telephonically because work-related commitments made it impossible for them to present themselves for scheduled interview sessions.

Open invitation letters were sent to the postgraduate students and junior faculty requesting them to participate in the current study. The purpose and benefits of participating in the study were clarified in the invitation letters. Furthermore, the letters revealed that participation in the study was not compulsory, meaning that they were free to withdraw at any point. The participants were 25–51 years old. The highest academic qualification was a doctorate, with the lowest being a diploma.

The University Research and Ethics Committee approved the study protocol (protocol no: SMNS/16/MBY/01/0701). This paved the way for the study to be conducted. Before commencement of the interviews, postgraduate students, junior faculty and their HoDs were made aware of the ethics to which the study would adhere. The right to anonymity, confidentiality and the freedom to withhold sensitive information if they so wished were specifically highlighted. Furthermore, permission to use an audio recorder during both the face-to-face in-depth interviews and the focus group discussion was sought from the participants. Signed informed consent was obtained from all the participants prior to their participation in the study.

Pilot testing research instruments

Interview guides were pilot tested with four respondents who were part of the target population. However, the latter were not included in the final study sample. Pilot testing was done in order to check the reliability of the data collection tools. This testing involved checking how long it took to conduct each interview, clarity of the questions and ease of understanding the interview questions.[16] The interview guides were given to other colleagues not involved in the study but with relevant knowledge and expertise for their comments. After the pilot testing exercise, the research instruments were re-worked to ensure that the questions were clearer, more precise and could be administered within the stipulated time frame.

Data collection and analysis

Data were collected through individual interviews and a focus group discussion. In-depth interview schedules guided the data collection. The interview schedule for postgraduate students and junior faculty sought to determine the relevant skills and knowledge that the participants had acquired as a result of their participation in the MAL-ED project. The principal researcher ensured that the individual interviews were held in a quiet environment and that they did not interfere with the academic commitments of the participants. On average, each interview was concluded within 50 minutes. The participants' responses regarding the benefits of the project were categorised into four identified categories derived from the themes of the study. However, the responses were so varied that data saturation was not reached.

After the individual interviews were complete, a focus group discussion was held. The group discussion involved 10 of the 22 postgraduate students and junior faculty who had participated in the individual interviews. The remaining 12 persons failed to participate in the focus group discussion because of other commitments. The focus group was designed to establish if group dynamics would elicit different perspectives from those obtained in the individual interviews. The participants chose their own chairperson and scribe from within the group. This approach enabled the group to select an impartial person who would allow them to express their views freely. The chairperson facilitated the group discussion and the scribe recorded the responses once there was consensus on an issue. The principal researcher ensured that there was order in the engagements. The focus group discussion took about 1 hour to complete.

As already reported above, HoDs were interviewed individually. The interviews took 50 minutes to complete. The focus of the interview questions for the HoDs was to find out whether there were any noticeable changes in the academic performance of staff members in their departments which could be attributed to their participation in the MAL-ED project.

Qualitative data obtained from the postgraduate students, junior faculty and HoD interviews were analysed using the thematic content analysis approach.[17] The principal researcher transcribed the audio-recorded data. Codes were generated and the emerging themes were recorded. The data were then fed into the Atlas.ti software version 7.5.10. The results of analysis were linked to the themes that had been identified earlier. The researchers deliberated on the outputs and reached consensus on the final themes and categories to be adopted.

Results of the study

The participants' perceptions on the various aspects of their human capital development as a result of active involvement in the MAL-ED project are described below and are presented one theme at a time. Research skills and knowledge, experience and exposure, interpersonal skills and personal development were the themes distilled from the results of the interviews and focus group discussion. Also addressed below are the expectations of postgraduate students and junior faculty regarding their participation in the biomedical community research project.

Research skills and knowledge

Of the 22 postgraduate students and junior faculty, 18 agreed that participating in the project resulted in them being empowered with research skills and knowledge. Cognitive testing, conducting anthropometric tests, and collection of bio-specimens such as stools and urine were specifically cited as the knowledge and skills they acquired. Some junior faculty were happy that they had been entrusted with leadership and supervision roles, which included overseeing data collection, checking completed forms, reviewing assessments and training field workers. The satisfaction is evident in the following views:

> *It was a privilege for me to be recruited for the project. The project brought a lot of things to my research. I learnt how to do cognitive testing, to get data from children, to adapt to new equipment and to collect quality data.* (Participant #14)

> *I am now overseeing data collection. I do quality checking of forms, reviewing assessments and training of fieldworkers. I collaborate with the Data Coordinating Centre. I was promoted to a lecturer position because of what I learnt from the MAL-ED project. I have been called to many interviews because of the skills I now possess.* (Participant #13)

Experience and exposure

A total of 14 of the 22 postgraduate students and junior faculty alluded to the fact that they had gained experience and valuable exposure. Some of the participants learnt how to relate with different people such as the mothers, children and community members involved in the studies. It was highlighted that the participants managed to put the theory they learnt at university into practice. Some participants attested to having gained experience in using laboratory equipment in addition to being exposed to the general 'world' of research. Some of these views are presented below:

> *The project has taught me to put theory into practice. At school, things were theoretical but the project gave me the practical side of things. I learnt to conduct anthropometric measurements. I gained a lot of experience in the field. I learnt about how people react to issues. I learnt how to relate to people.* (Participant #1)

> *The project is full of laboratory work, blood processing, fieldwork and ordering of laboratory supplies. I got a lot of experience in DNA extraction from stools and saliva. I also attended field meetings. I have gained a lot of experience and exposure to blood processing from the project. MAL-ED has widened my scope on many issues.* (Participant #12)

Interpersonal skills

Another benefit that accrued to participants in the MAL-ED project was the enhancement of interactive skills. It was explained that the acquisition of interpersonal skills resulted in continuous communication with community members and colleagues. In the process, the participants in the project worked well as a team, and respected and appreciated their diversity in terms of people, ideas and cultures – which is crucial for fieldworkers because it helps create conditions that enable collection of data of high quality. Some participants articulated their experiences as follows:

> *I have learnt how to relate to people and I have made friends in the process. I have learnt to communicate with different community members.* (Participant #2)

> *I have learnt how to communicate well with community members and to understand community issues. I have learnt to interact with everybody, poor or not. I have a good relationship with the participating mothers. I know how to maintain good relationships with children.* (Participant #5)

Personal development

It was revealed that the MAL-ED project helped expand the capabilities of those involved in implementing it, developed and strengthened their character and also motivated others to pursue further studies or advance seemingly stalled careers. These experiences highlight the fact that even though the participants were recruited to implement the project, they were empowered and learnt to enhance the capacities of others as well. Some respondents commented:

> *The project has improved my confidence. I have learnt to work with different people of different cultures. I now understand how people respond to pressure and commitment. I have learnt to understand my work colleagues.* (Participant #3)

> *The project helped me in my personal development. I also felt a sense of belonging. The project gave me enough ground to explore my capabilities. I was able to empower someone to take over from me. From MAL-ED, I learnt to write manuscripts, attend conferences and present papers and to interact with more experienced researchers. MAL-ED was like a family to me.* (Participant #22)

Other expectations of postgraduate students and junior faculty

Despite the varied human capital benefits that accrued to the postgraduate students and junior faculty through their involvement in the MAL-ED project, 9 out of the 22 postgraduate students and junior faculty indicated that they needed further training on data analysis and interpretation, exposure to a wider range of laboratory techniques as well as time and project management training. Active listening, how to handle children and research techniques were other areas that were identified for further training. Some participants wanted to be involved in other laboratory procedures and in data analysis that was conducted downstream. The following sentiments shed more light on these results:

> *An opportunity in analysing quantitative data and interpreting it would have helped me.* (Participant #8)

> *I wish I could be exposed to other new laboratory techniques and attend more training workshops.* (Participant #12)

Human capital development as perceived by heads of departments

In general, the HoDs interviewed in the current study corroborated the views expressed by junior faculty. They were of the view that research skills such as data collection, data analysis, writing of quality manuscripts and mentoring were enhanced. In addition, they believed that the project assisted the participants in their personal development. Some of the members of staff seemed to collaborate with other scientists much better than they had before. They demonstrated more

mature behaviour in the way they executed their work, with some gaining promotion mainly because of the quality of their research outputs as a result of their participation in the MAL-ED project. Below are some of the views of the HoDs:

> *The multi-disciplinary approach of the project (microbiology, psychology and nutrition) exposes one to different people and a different way of looking at a problem. The project exposed people to practical ways of understanding issues. Staff members have shown maturity in the way they manage focus group discussions, transcribing, collecting data and writing articles.* (Participant #25)

> *Research related skills have improved. There has been improvements in the skill of data collection, experience in research protocols like community entry and ethical issues. The skill of data collection has been enhanced. Staff members have been exposed to manuscript writing.* (Participant #23)

> *The quality of publications are much improved in the sense that works were accepted in high profile journals which in a way attests to the quality of the work. Some members of staff have become professors based on the output from the project. This has enhanced their careers.* (Participant #25)

Consolidated views from individual in-depth interviews, focus group and heads of departments

The most prominent responses drawn from individual in-depth interviews are presented in Table 1. It should be noted that the results obtained through individual interviews and focus group discussions were virtually the same. In general, the project contributed to considerable human capital development among postgraduate students and junior faculty.

Discussion

Participation of postgraduate students and junior faculty in the MAL-ED biomedical community research project was found to have particularly enhanced their skills and knowledge. The projects that postgraduate students and junior academic staff mount, the lectures prepared for delivery and manuscripts written for publication in scientific media depend on and demand that quality research be undertaken. A considerable number of the postgraduate students and junior faculty who participated in the MAL-ED project secured jobs in the public sector, highlighting the immense value and capacity enhancing power of participating in the project. Other participants explained that, as a result of their involvement in the project, they were promoted to senior positions. This is in agreement with Ibok and Ibanga's[18] contention that improvements in skills and knowledge increases individuals' employment prospects and related size of remuneration packages. The improved employability of the research participants is worth highlighting, especially considering the high levels of unemployment in the country. These results suggest that the quality of postgraduate programmes can be improved if students are integrated into ongoing research projects that senior academics run. Thus, it is prudent to take this into account in the course of designing innovative curricula for research-based postgraduate degrees.

The participants' views in the current study resonate with those expressed by the grassroots community members in Moyo et al.'s[16] study conducted in the community in which the MAL-ED project was implemented. Acquisition of knowledge and research skills enhanced personal development and enabled the grassroots community members in the forefront of project implementation to secure jobs. A study by Dongre et al.[19] in India which focused on the benefits of exposing medical undergraduate students to community-based surveys yielded similar results to those of the current study. The benefits that accrued to the medical students included the ability to conduct interviews, better communicate with local villagers, collect and enter or store data using computers, and apply learning to research work as well as enhanced awareness of the public health process.

Postgraduate students and junior faculty also reported that the project enabled them to gain experience and exposure to new research techniques and practical realities. For instance, they pointed out that the project had enabled them to apply the theory they had learnt at university to real-life situations. This enhanced their understanding and appreciation of key concepts that might have otherwise remained abstract. Many participants were exposed to the realities of community-based research for the first time through this initiative. They had the opportunity to work with and interact with children and their mothers together with other grassroots community members for the first time. This helped to build and strengthen their social relationships and networking and enhanced their communication skills. In the process, the research participants acquired valuable experience and exposure, resulting in them gaining confidence in what they do. These views corroborate the observations of Tingen et al.[20] in a study in Atlanta (Georgia, USA).

In the latter study, undergraduate nursing students gained professional experience and personal development through their participation in a research project conducted by their faculty. As was the case with the MAL-ED South Africa project, the nursing students in Tingen et al.'s[20] study had limited exposure to research processes prior to their involvement in the study.

Table 1: Consolidated views from individual in-depth interviews, the focus group discussion and individual heads of department interviews

Benefits	In-depth interviews	Postgraduate students (PS)/ junior faculty (JF)	Focus group	Heads of department
1. Experience in working with different people	The project helped me to deal with people of different cultures. (8)	PS	•	••
2. Other benefits	I benefitted financially. I used the money to further my studies. (8)	PS	•	••
3. Research knowledge and acquisition of research skills	The study brought a lot of insights about research, initial preparation for a study and the reason for carrying out a study. (22)	PS and JF	•	••
4. Personal development	The study motivated me to study further. (21)	PS and JF	•	••
5. Writing scientific research manuscripts	The project has helped me in writing manuscripts. At times we wrote as a team.(7)	JF	•	••

Perspectives from individual interviews regarding the benefit statements are shown in the first column. Responses to those statements by the respondents are indicated in numerals in the second column. The third column shows the group of participants who responded to each of the key statements. A bullet (•) in the fourth column is used to indicate that each response was supported by the focus group. The number of bullets in the heads of departments column represents the number of heads of departments who selected each of the statements.

The students revealed that their involvement in the project improved their research skills, in particular data review, analysis and writing manuscripts for scholarly publishing. This reinforced the widely held notion that 'experience is the best teacher'. Another study in Japan[21] highlighted the benefits of exposing health science students to off-campus education. Students were able to interact with local residents. Afterwards, they expressed positive views about their community-based experience. Thus, exposing students to real-life environments enhances their understanding of issues and is likely to result in better graduate attributes.

Research participants cited enhancement of interpersonal skills as another attribute of human capital development gained through the current study. They mentioned improved ability to communicate well with grassroots community members and peers and researching together with mothers and their children as they collected time series data. As they participated in the project and remained in constant contact with the participants and grassroots community members, they became less judgmental of others and learnt how best to interact with a diverse range of people. Having an accommodating personality helps improve social relationships and cohesion – all of which are important prerequisites for individual and collective development. Raman[22] concurs with this viewpoint and emphasises that effective interpersonal communication skills are a prerequisite of social interaction and building sustainable relationships.

Lastly, personal development was achieved through creating opportunities for the participants to apply the theory they had learnt in class to real-life situations, thereby boosting their confidence. When an individual's confidence is boosted, they often feel motivated to perform better in the tasks assigned to them. Ultimately, they improve the performance of the organisation for which they work. The fact that some participants became so motivated that they eventually pursued further studies was crucial because it helped build their professional careers. Their desire to improve their educational qualifications resonates with the ideals of the late Nelson Mandela who alluded to the fact that education is the most powerful weapon which one can use to transform oneself and the world around them. Others reported that they were better organised, more disciplined and were able to work more independently as a result of their involvement in the MAL-ED project. All these are positive attributes that demonstrate maturity, which their HoDs confirmed to be the case. Thus, it can be concluded that participation in the MAL-ED project enabled both postgraduate students and junior faculty to unlock and sharpen their capabilities.

Limitations of the study

Although only 25 and 10 people participated in the interviews and focus group discussions, respectively, their insights into the benefits accruing to them as a result of their involvement in the MAL-ED project were quite revealing. Ideally, having more than one focus group could have enabled deeper, independent interrogation of issues. However, the fact that the results of the individual interviews and focus group discussion were virtually the same, highlighted the fact that what is reported here was a true reflection of participants' experiences.

Conclusions

Skills and knowledge in research, experience and exposure, interpersonal skills and personal development were the major benefits that accrued to postgraduate students and junior faculty as a result of their participation in the biomedical community research project. Some participants highlighted the need for training workshops that focused on data analysis and interpretation, exposure to other laboratory techniques as well as time and project management training. These findings are useful as inputs into the planning and implementation of future similar projects.

Acknowledgements

We are grateful to the postgraduate students, junior faculty and heads of departments for their participation in the study. C.S.M. was supported by the Fogarty International Center of the National Institutes of Health, USA (award no. D43 TW009359).

Authors' contributions

P.O.B. conceptualised the study. C.S.M. and J.F. designed the study. C.S.M. collected and analysed the data, and drafted the manuscript with contributions from P.O.B. and J.F. All authors read and approved the final manuscript.

References

1. Mfutso-Bengo J, Masiye F, Moyneux M, Ndebele P, Chilungo A. Why people refuse to take part in biomedical research studies: Evidence from a resource-poor area. Malawi Med J. 2008;20(2):57–63.
2. Ojakaa DI, Ofware P, Machira WW, Yamo E, Collymore Y, Ba-Nguz A, et al. Community perceptions of malaria and vaccines in the South Coast and Busia regions of Kenya. Malar J. 2011;10, Art. #147, 11 pages. http://dx.doi.org/10.1186/1475-2875-10-147
3. Massawe IS, Lusingu JP, Manongi RN. Community perception on biomedical research: A case study of malariometric survey in Korogwe, Tanga Region, Tanzania. BMC Public Health. 2014;14, Art. #385, 9 pages. http://dx.doi.org/10.1186/1471-2458-14-385
4. Rennie S, Perry B, Corneli A, Chilungo A, Umar E. Perceptions of voluntary medical male circumcision among circumcising and non-circumcising communities in Malawi. Glob Public Health. 2015;10(5–6):679–691. http://dx.doi.org/10.1080/17441692.2015.1004737
5. Marimuthu M, Arokiasamy L, Ismail M. Human capital development and its impact on firm performance: Evidence from developmental economics. The Journal of International Social Research. 2009;2(8):265–272.
6. Organisation for Economic Co-operation and Development (OECD). The wellbeing of nations. The role of human and social capital, education and skills. Paris: OECD Centre for Educational Research and Innovation; 2001. Available from: http://www.oecd.org/site/worldforum/33703702.pdf
7. Gilead T. Human capital, education and the promotion of social cooperation: A philosophical critique. Stud Philos Educ. 2009;28:555–567. http://dx.doi.org/10.1007/s11217-009-9146-2
8. Becker GS. Human capital. 2nd edn. The concise encyclopaedia of economics [document on the Internet. c2008 [cited 2017 Jul 18]. Available from: http://www.econlib.org/library/Enc/HumanCapital.html
9. Gao N, Gill KJ, Schmidt LT, Pratt CW. The application of human capital theory in vocational rehabilitation for individuals with mental illness. J Vocat Rehabil. 2010;32:25–33. http://dx.doi.org/10.3233/JVR-2010-0492
10. Nhamo S, Nhamo G. Macroeconomics, (adult) education and poverty eradication in southern Africa. Int Rev Educ. 2006;52:305–322.
11. Mincer J. Human capital and economic growth. NBER Working Paper Series. Working Paper no. 803. New York: National Bureau of Economic Research; 1981.
12. Channar ZA, Talreja S, Bai M. Impact of human capital variables on the effectiveness of the organizations. Pak J Commer Soc Sci. 2015;9(1)228–240.
13. Bessong PO, Nyathi E, Mahopo TC, Netshandama V and MAL-ED network investigators. Development of the Dzimauli community in Vhembe District, Limpopo Province of South Africa, for the MAL-ED cohort study. Clin Infect Dis. 2014;59(suppl 4):S317–S324. http://dx.doi.org/10.1093/cid/ciu418
14. Kothari CR. Research methodology. Methods and techniques. 2nd edn. New Delhi: International (P) Limited Publishers; 2004.
15. Santha B, Sudheer H, Saxena V, Tiwari V. Qualitative research in dental public health care: An overview. The Pharma Innovation. 2015;4(2):83–86.
16. Moyo CS, Francis J, Bessong PO. Perceptions of community-based field workers on the effect of a longitudinal biomedical research project on their sustainable livelihoods. BMC Public Health. 2017;17, Art. #267, 11 pages. http://dx.doi.org/10.1186/s12889-017-4138-6
17. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analysing and presenting qualitative data. Br Dent J. 2008;204(8):429–432. http://dx.doi.org/10.1038/sj.bdj.2008.292.
18. Ibok EE, Ibanga SE. Impact of human capital development and economic empowerment on the socio-economic development of Akwa Ibom State, Nigeria. Glob J Hum Resour Manage. 2014;2(3):37–44.

19. Dongre AR, Kalaiselvan G, Mahalakshmy T. The benefits to medical undergraduates of exposure to community-based survey research. Educ Health. 2011;24(3):591.

20. Tingen MS, Burnett AH, Murchison RB, Zhu H. The Importance of nursing research. J Nurs Educ. 2009;48(3):167–170.

21. Nakagawa K, Asakawa Y, Yamada K, Ushikubo M, Yoshida T, Yamaguchi H. Benefits of off-campus education for students in the health sciences: A text-mining analysis. BMC Med Educ. 2012;12, Art. #84, 7 pages. http://dx.doi.org/10.1186/1472-6920-12-84.

22. Raman M. Interpersonal communication as an essential ingredient of organizational bonding. ELT Weekly. 2014;6(15).

Screening of the NIH Clinical Collection for inhibitors of HIV-1 integrase activity

AUTHORS:
Shaakira Abrahams[1,2]
Salerwe Mosebi[1]
Muhammed Q. Fish[1,2]
Maria A. Papathanasopoulos[2]
Raymond Hewer[3]

AFFILIATIONS:
[1]Centre for Metal-Based Drug Discovery, Advanced Materials Division, Johannesburg, South Africa
[2]Department of Molecular Medicine and Haematology, University of the Witwatersrand Medical School, Johannesburg, South Africa
[3]Discipline of Biochemistry, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Raymond Hewer

EMAIL:
hewerr@ukzn.ac.za

DATES:
Received: 21 Sep. 2017
Accepted: 12 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
drug repurposing; strand transfer; pan-assay interference compounds; PAINS

HOW TO CITE:
Abrahams S, Mosebi S, Fish MQ, Papathanasopoulos MA, Hewer R. Screening of the NIH Clinical Collection for inhibitors of HIV-1 integrase activity. S Afr J Sci. 2018;114(3/4), Art. #2017-0324, 5 pages. http://dx.doi.org/10.17159/sajs.2018/20170324

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
National Research Foundation (South Africa)

Drug repurposing offers a validated approach to reduce drug attrition within the drug discovery and development pipeline through the application of known drugs and drug candidates to treat new indications. Full exploitation of this strategy necessitates the screening of a vast number of molecules against an extensive number of diseases of high burden or unmet need and the subsequent dissemination of the findings. In order to contribute to endeavours within this field, we screened the 727 compounds comprising the US National Institutes of Health (NIH) Clinical Collection through an HIV-1 (human immunodeficiency virus type 1) integrase stand transfer inhibition assay on an automated scintillation proximity assay platform. Only two compounds were identified within the initial screen, with cefixime trihydrate and epigallocatechin gallate found to reduce integrase strand transfer activity at IC_{50} values of $6.03 \pm 1.29\ \mu M$ and $9.57 \pm 1.62\ \mu M$, respectively. However, both cefixime trihydrate and epigallocatechin gallate retained their low micromolar inhibitory activity when tested against a raltegravir-resistant integrase double mutant ($FCIC_{50}$ values of 0.83 and 0.06, respectively), were ineffective in an orthogonal strand transfer ELISA (<30% inhibition at 100 μM) and produced negligible selectivity index values (<1) in vitro. While no useful inhibitors of HIV-1 integrase strand transfer activity were found within the NIH Clinical Collection, the identification of two assay-disrupting molecules demonstrates the importance of consideration of non-specific inhibitors in drug repurposing screens.

Significance:
- This study is the first to screen the US NIH Clinical Collection for potential HIV-1 integrase inhibitors.
- The pervasive nature of promiscuous inhibitors is emphasised.

Introduction

Early-stage drug discovery fulfils a critical role within the broader drug discovery process and the entire drug discovery and development pipeline. Early-stage drug discovery is typically – but not always – undertaken following target identification and validation, and involves the screening of compounds with the intent purpose of identifying compounds with promising activity (HIT compounds) that can then be developed further (into LEAD compounds) within the drug discovery phase. Early-stage drug discovery activities can range from the evaluation of a limited set of compounds, typically selected through rational drug design methodologies, to the assessment of large compound libraries through high throughput screening (HTS; defined as the screening of >10 000 compounds per day) and even ultra-HTS (μHTS; defined as the screening of >100 000 compounds per day) operations. Owing to the sheer number of compounds screened, early-stage drug discovery ostensibly carries the highest failure rate and, accordingly, the highest risk of all activities within the pipeline. However, the true bottleneck to success in the broader drug discovery and development process lies less with the quantity of compounds identified as HITS during screening and more with the quality of these compounds and their suitability as drug candidates. Specifically, the highest cause for compound attrition in the pipeline, by far, is attributed to non-clinical toxicity which accounts for the termination of >40% of all compounds from the drug discovery and development pipeline.[1]

To mitigate the potential significant financial loss resulting from compound failures, in particular the high cost of late-stage failures, most pharmaceutical organisations adopt the 'fail early, fail cheap' paradigm. To support this approach, researchers aim to recognise ADMET-related issues through an ever-increasing number of tests undertaken at progressively earlier stages of the pipeline. Equally, findings from these tests have been retrospectively accumulated to delineate physiochemical properties (i.e. LogP, LogD, molecular weight, aromatic rings, rotatable bonds, polar surface area, etc.) that influence drug-likeness and then subsequently collated into 'rules of thumb' (such as the Lipinski rule of five, the rule of three and many other variations and extensions) and property prediction software programs. Application of these predictive models has allowed for the identification and judicious removal of non-favourable compounds either following screening or directly from the physical compound library prior to screening. While immeasurably useful, these tools have not proven infallible as evinced through a recent study of 812 failed compounds (oral development candidates from four different major pharmaceutical companies) that could draw no correlation between non-clinical toxicology failure and physiochemical properties.[1] Similarly, a subset of compounds eliciting growing interest because of their subversive effects in drug discovery efforts are promiscuous inhibitors[2,3] or pan-assay interference compounds (PAINS)[4]. These compounds yield convincing false-positive results in biological assays and significant efforts have been undertaken to identify them and ultimately remove them from screening libraries.[4] Broadly categorised and inclusive of several classes of compounds with varying mechanisms of action (i.e. aggregate-inducing compounds, redox-cyclers, covalent modifiers, metal complexes), these compounds do not readily lend themselves to predictive algorithms. Nonetheless, databases of existing PAINS highlight common structures (i.e. flavonoids, quinones, rhodamines) and some can also be searched for similarity.

Of other approaches aimed at minimising compound attrition, the concept of drug repurposing (or drug repositioning) has drawn significant interest. Herein, the underlying principle is the evaluation of clinically approved drugs or

late-stage clinical trial failures (all off-patent or generics) as disease-modifying agents in therapeutic areas other than the one for which they were designed or proved effective. The main appeal of this approach is the decreased risk of failure arising from safety issues while the extensive prior development allows for a quicker transitioning through the pipeline (up to 60%) with reduced costs (up to 40%). If successful, the drug can be granted patent protection on grounds of a new application or new formulation. The classic example of drug repositioning is Viagra® – the blockbuster erectile dysfunction drug from Pfizer which first served as an angina medication under the name Sildenafil. Numerous other examples exist, including azidothydimine (the cancer turned anti-HIV drug), ropinirole (a dual Parkinson's and restless legs syndrome treatment) and Rogaine® (a hair-loss drug repurposed from a blood pressure drug from a failed ulcer candidate), to mention but a few. The scientific merit of the concept has driven growth in its popularity as is clearly evident through interest from major pharmaceutical companies, the growth in focused start-up companies, the rise in related literature and a recently launched journal (*Drug Repurposing, Rescue and Repositioning*) with dedicated content.

Applying the drug repurposing approach to the field of HIV drug discovery has been previously described. While the treatment options for HIV-1 are formidable – both in the number of antiretroviral agents approved and the efficacy of combination therapy – the absence of an effective therapeutic vaccine or cure and the persistent threat of antiretroviral drug resistance has substantiated the continued exploration for novel inhibitors. In this study, we sought to identify an existing drug with activity against HIV-1 integrase (IN) – a virally encoded enzyme that catalyses the integration of viral DNA into the host chromosome. For this purpose we screened the US National Institutes of Health (NIH) Clinical Collection (NCC), which is a 727 small-molecule library of FDA-approved and late-stage candidates that has been previously explored for proteasome stimulators[5] and coronavirus inhibitors[6] but not, to the best of our knowledge, for HIV-1 IN inhibition. The NCC library was screened by means of an automated process through an HIV-1 IN strand-transfer (ST) inhibition scintillation proximity assay (SPA) in order to identify novel catalytic IN inhibitors.

Methods

Expression and purification of recombinant HIV-1 integrase

The reagent pINSD.His (Cat. #2957) was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr Robert Craigie.[7,8] Recombinant His-tagged HIV-1 IN was expressed and purified as previously described.[9] Briefly, wild-type HIV-1 IN was overexpressed in *E. coli* BL21 (DE3) bacterial cells using the NL4-3 histidine (HIS)-tagged HIV-1 IN coding sequence, pINSD, cloned into pET15B (Merck Millipore, Darmstadt, Germany). Cells were grown to logarithmic phase in Luria-Bertani medium and induced with 1 mM isopropyl-thio-galactoside. The recombinantly expressed protein was purified through affinity chromatography using a nickel (Ni)-affinity column and buffer exchanged into storage buffer (20 mM HEPES pH 7.2, 1 M NaCl, 4 mM EDTA, 2 mM dithiothreitol and 50% glycerol) using a PD-10 Sephadex column (GE Healthcare, Buckinghamshire, UK). The expression and purification of the HIV-1 IN was confirmed through SDS-PAGE and subsequent Western blot analysis. Similarly, recombinant IN which incorporated raltegravir-resistant mutations, $IN_{Q148H/G140S}$, was prepared. Briefly, the pINSD.His plasmid was used as a template for mutagenesis with the QuickChange Lightning Site Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA, USA). Raltegravir-resistant mutations were inserted into the pINSD.His backbone and included Q148H/G140S substitutions.

Radiolabelling of target DNA for scintillation proximity assays

Annealed oligonucleotides for target DNA (tDNA), T 56-S (AAAAGGAGGA-GAAGGAAAGGAGAGAGAGCGAATTAGCCCTTGGTC) and T 56-A (AAAAGGA GGAGAAGGAAAGGAGAGAGAGGACCAAGGGCTAATTCG) oligonucleotides (Inqaba Biotech, Pretoria, South Africa), were radiolabelled with ^{3}H-dCTP and ^{3}H-dTTP (AEC Amersham, Johannesburg, South Africa) by filling in the 5'-overhangs according to the Fermentas Klenow fragment DNA labelling kit instructions (ThermoFisher Scientific, Waltham, MA, USA). Unincorporated nucleotides were removed from the radiolabelled tDNA using the QIAquick nucleotide removal kit (Qiagen, Hilden, Germany).

HIV-1 integrase strand transfer scintillation proximity assay

The SPA was carried out as previously described[10] and adapted to an automated platform on a Hamilton Starlet robotic system (Hamilton, Bonaduz, Switzerland). Briefly, a 10x reaction buffer was prepared containing 200 mM HEPES (pH 7.5), 300 mM NaCl, 50 mM dithiothreitol and 0.5% Igepal (nonidet-P40). Polyvinyltoluene streptavidin-coated scintillation beads (GE Healthcare Sciences, Marlborough, MA, USA) were reconstituted in 1x reaction buffer at a final concentration of 10 mg/mL. Biotinylated donor DNA (dDNA) was added at a final concentration of 500 nM and rocked at room temperature for 1 h. The bead suspension was washed twice with 1x reaction buffer and centrifuged at 1000 *x g* for 5 min. The pellet was resuspended at 2 mg/mL in 2x reaction buffer to which recombinant IN (wild-type or Q148H/G140S mutant) was added at a final concentration of 1 μM and rocked at room temperature for 30 min. The final SPA reactions comprised, per well: 1 mg/mL SPA bead-dDNA-IN complex with 8 to 10 test compounds at 10 μM each for single-dose experiments or concentrations ranging from 100 to 0.78 μM for dose-response experiments (substituted with DMSO buffer solution for blank control). This reaction mixture was incubated at 22 °C for 30 min whilst shaking gently. The reactions were initiated by adding 500 nM ^{3}H-tDNA to each well at a final concentration of 50 nM and incubated at 37 °C shaking for 90 min before the enzymatic reaction was stopped using 62 mM EDTA. The reaction product formation was measured using the Top Count Scintillation Counter NXT (Perkin Elmer, Waltham, MA, USA). Percentage inhibition was determined for single-dose experiments while IC_{50} values were determined as the compound concentration required to reduce HIV-1 recombinant IN ST activity by 50% and calculated using OriginPro 8.0 software (Origin Lab Corporation, Northampton, MA, USA). All inhibition values are the average of at least triplicate experiments.

HIV-1 integrase strand transfer enzyme-linked immunosorbent assay

The HIV-1 IN strand transfer inhibition enzyme-linked immunosorbent assay (ELISA) was adapted from previously described methods.[9,11] Briefly, 0.15 μM double-stranded biotinylated dDNA (5'-biotin-ACCCTTTTAGTCAGTGT GGAAAATCTCTAGCA-3' and 5'-ACTGCTAGAGATTTTCCACACTGACTAA AAG-3') was added to the wells of streptavidin-coated 96-well microtitre plates (R&D Systems, Minneapolis, MN, USA). Following incubation at room temperature for 60 min and a stringent wash step, 1 μM purified recombinant HIV-1 subtype B IN (in the presence of $MgCl_2$) was assembled onto the pre-processed dDNA through incubation for 30 min at 22 °C. Following a wash step, the test compounds were titrated into individual wells at a final concentration of 100 μM for single-dose evaluation or concentrations ranging from 100 to 0.78 μM for dose-response experiments. The microtitre plates were incubated for 30 min at 37 °C, washed and the strand transfer reaction was initiated through the addition of 0.25 μM double-stranded FITC-labelled target DNA (5'-TGACCAAGGGCTAATTCACT-FITC-3' and 5'-AGTGAATTAGCCCTTGGTCA-FITC-3') in Hepes buffer containing $MgCl_2$ and $MnCl_2$. After an incubation period of 60 min at 37 °C, the plates were washed as before and an alkaline phosphatase conjugated anti-FITC secondary antibody (Sigma-Aldrich, St Louis, MO, USA) was added. Finally, the plates were washed and substrate (BluePhos, KPL, Gaithersburg, MD, USA) was added to allow for detection at 620 nm using a microplate reader (xMark™, Bio-Rad, Hercules, CA, USA). Percentage inhibition was determined for single-dose experiments while IC_{50} values were determined as the compound concentration required to reduce HIV-1 recombinant IN ST activity by 50% and calculated using OriginPro 8.0 software (Origin Lab Corporation). All inhibition values are the average of at least triplicate experiments.

Cytotoxicity assays

The reagent MT-4 (Cat. #120) was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr Douglas Richman.[12-14] The cytotoxicity assay was performed as per standard methods and as described previously.[9,15] Briefly, MT-4 cells were plated in 96-well microtitre plates at 3.0 x 10^5 cells/mL and allowed to stabilise

for 2 h at 37 °C and 5% CO_2. Thereafter, test compounds were added to the plate through twofold serial dilution to allow for eight final compound concentrations ranging from 200 to 1.56 μM in a total volume of 200 μL/well. The cells and compounds were then incubated for 96 h at 37 °C and 5% CO_2. To each well, 20 μL CellTiter 96 Aqueous One Solution (Promega, Madison, WI, USA) was added. The plates were incubated for 4 h and absorbance was read at 490 nm on a multiplate reader (xMark, Bio-Rad). CC_{50} values were determined as the concentration of the test compound required to reduce the cell viability by 50% and were calculated using OriginPro 8.0 software (Origin Lab Corporation). The values obtained are averages of at least three separate experiments.

Antiviral activity

To determine antiviral activity, 50 μL HIV-1_{NL4-3} virus was added to $3.0x10^5$ MT-4 cells/mL at a multiplicity of infection of 0.1 and the mixture was spinoculated at 3000 *x g* for 90 min. After washing off unbound virus, cells were plated in 96-well microtitre plates at 100 μL/well and allowed to stabilise for 1 h at 37 °C and 5% CO_2. Thereafter, test compounds were added to the plate through twofold serial dilution to allow for eight final compound concentrations ranging from 200 to 1.56 μM in a total volume of 200 μL/well. The cells and compounds were then incubated for 96 h at 37 °C and 5% CO_2. Cell-free supernatants were collected from each well, and p24 concentration was determined using the Vironostika HIV-1/2 p24 Antigen ELISA (bioMerieux, Marcy-l 'Etoile, France) as per manufacturer's instructions. EC_{50} values were determined as the concentration of the test compound required to reduce p24 concentration by 50% and were calculated using OriginPro 8.0 software (Origin Lab Corporation). The values obtained are averages of at least three separate experiments. Selectivity index (SI) values were calculated as the ratio of CC_{50}/EC_{50}.

Results and discussion

Inhibition of recombinant integrase strand transfer activity

To begin the screening process, all 727 small molecules within the NCC library were pooled in an orthogonal manner into 171 pools; each pool comprised a combination of 8 or 10 different compounds with each compound present in two pools. The pools were then screened through an automated SPA, in triplicate, with each compound evaluated at a final single-dose concentration of 10 μM. Initially, 24 pools were found to be affected by colour-quenching which was only resolved through the identification and removal of 12 coloured compounds. On completion of the screening process, four pools were found to reduce recombinant IN_{WT} ST activity by the pre-defined minimum cut-off of $\geq$50%, indicating the presence of two active compounds. Through cross-referencing, the two common compounds were identified as cefixime trihydrate (CEF) and epigallocatechin gallate (EGCG). CEF, a third-generation orally administered cephalosporin is a potent, broad-spectrum bactericidal. Like other cephalosporins, CEF possesses a dihydrothiazine ring fused to a beta-lactam ring and derives its therapeutic effect through inhibition of cell-wall synthesis by disruption of the transpeptidation process. Modulation of HIV-1 replication by CEF has not been previously described in the literature; however, cephalosporin oligonucleotides and monocyclic β-lactams have been reported as HIV-1 protease inhibitors. EGCG is a polyphenolic acid ester that has been proclaimed as an effective agent for an exceedingly broad range of diseases (including HIV-1 through several distinct mechanisms of action) despite its well-documented promiscuous nature.[4,16] A search of the US NIH registry and results database of clinical studies revealed 92 clinical trials in varying stages of completion (from recruiting to completed) examining EGCG as treatment for 276 different clinical conditions.[17]

In our experience, the HIV-1 integrase ST SPA is a robust assay that yields a low number of HIT compounds per screen. The inclusion of detergent within the protocol design serves, perhaps inadvertently, to reduce the identification of false positive responses caused by promiscuous molecules. We therefore opted to continue investigating EGCG to verify a non-specific mechanism, and as such, both CEF and EGCG were subjected to dose-response evaluation within the previously described SPA. Here we established IC_{50} values of 6.03$\pm$1.29 μM and 9.57$\pm$1.62 μM for CEF and EGCG, respectively (Table 1) with steep slopes observed for both compounds (Hill slope > 1). In comparison, raltegravir, the first-in-class HIV-1 integrase drug marketed as Isentress®, inhibited recombinant IN_{WT} activity by 92$\pm$5% at a single-dose concentration of 10 μM and produced an IC_{50} of 9.98$\pm$0.83 nM (Table 1) with a Hill slope of 1. Thereafter, the compounds were tested for inhibitory activity of the raltegravir-resistant $IN_{Q148H/G140S}$ double-mutant.[18] Replication capacity of this double mutant was reduced to 59% of IN_{WT} and, as anticipated,[19] raltegravir was significantly less effective against this mutant ($p<0.01$) as reflected by a fold change in IC_{50} ($FCIC_{50}$) value of 214 (Table 1). On the contrary, the two identified compounds retained their micromolar inhibitory activity against the raltegravir-resistant double mutant with $FCIC_{50}$ values of 0.83 and 0.06 calculated for CEF and EGCG, respectively (Table 1). As a further direct evaluation of activity, the compounds were tested for ST inhibition within an HIV-1 IN ST assay based on an ELISA platform. As a true, indisputable IN ST inhibitor, raltegravir efficiently inhibited IN_{WT} in this orthogonal assay to the same degree as that observed within the SPA-based assay (IC_{50} = 10.25$\pm$0.75 nM) while dose-response evaluations were not warranted for CEF and EGCG as both proved ineffective inhibitors at a high single-dose concentration (<30% inhibition at 100 μM; Table 1). Up to this point, the behaviour of CEF and EGCG strongly supported a non-specific role for both molecules through a related mechanism of action that disrupted the SPA-based platform. This observation subsequently led us to the work of another group that speculated a non-specific mechanism for CEF within an SPA – in this case as an inhibitor of both NADH dehydrogenase (at an $IC_{50} \sim 8$ μM) as well as MurG (at an $IC_{50} \sim 16$ μM).[20]

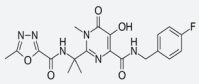
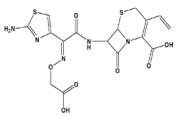
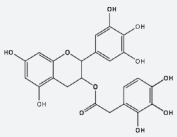
In-vitro evaluation

EGCG and CEF were both evaluated for toxicity within the MT-4 mammalian cell line. While EGCG yielded a CC_{50} of 23 μM, CEF was not found to be toxic within the limits of the assay ($\geq$200 μM). Thereafter, inhibition of HIV-1 replication by CEF and EGCG was evaluated in vitro in the MT-4 cell line following infection by HIV-1_{NL4-3}. For EGCG, an EC_{50} of 24 μM was determined through dose-response studies (Table 1). The observed viral inhibition by EGCG closely correlated the toxicity profile of the compound in the same cell line and led to a negligible SI value ($\sim$1; Table 1). An SI value for CEF could not be determined as no observable viral inhibition was found in the cell-based HIV-1 inhibition assay up to the maximum tested compound concentration of 200 μM (Table 1). As a control, raltegravir was found to inhibit HIV-1 replication within this assay with an EC_{50} of 16.82 nM with no observable toxicity up to the limit of the assay ($CC_{50} \geq 200$ μM; SI value $\geq$ 11 890.61).

Conclusion

Drug repurposing has proven successful in the past and offers a viable strategy for the discovery and development of therapeutic agents. In an endeavour to contribute to efforts in this field, we screened the NCC library to identify new inhibitors of HIV-1 integrase strand transfer activity. While no true inhibitors of HIV-1 IN ST activity were discovered, the identification of two non-specific inhibitors through our screen demonstrated that drug repurposing is not insusceptible to the presence of assay disruptors. In particular, the identification of EGCG demonstrates the invasiveness of even the most well-documented PAINS into chemical screening libraries. Furthermore, and perhaps more interestingly, the confirmation of the clinically relevant antibiotic CEF as an SPA disruptor demonstrates necessity to interrogate the action of well-characterised molecules within specific assay platforms and also supports the mandatory use of secondary or orthogonal assays to confirm inhibition. The findings from this study suggest that both EGCG and CEF disrupt the SPA through a similar non-aggregating mechanism that will be elucidated through future studies to facilitate further screening projects based on this assay platform.

Table 1: Comparison of cefixime trihydrate, epigallocatechin gallate and raltegravir as inhibitors of HIV-1 integrase strand transfer activity

	HIV-1 integrase strand-transfer inhibition					MT-4 cell based assays		
	Scintillation proximity assay (SPA)			ELISA				
Compound	**IN_{WT} IC_{50} ± SE (μM)**	**$IN_{Q148H/G140S}$ IC_{50} ± SE (μM)**	**$FCIC_{50}$**	**% Inhibition**	**IC_{50} ± SE (nM)**	**CC_{50} ± SE (μM)**	**EC_{50} ± SE (μM)**	**SI**
Raltegravir 	9.98 ± 0.83 (x 10^{-3})	1.50 ± 0.50	214	96% (at 10 μM)	10.25 ± 0.75	>200	16.82 (x 10^{-3})	≥ 11 890.61
Cefixime trihydrate 	6.03 ± 1.29	5.01 ± 1.29	0.83	21% (at 100 μM)	ND	>200	>200	–
Epigallocatechin gallate 	9.57 ± 1.62	0.62 ± 1.50	0.06	29% (at 100 μM)	ND	23 ± 1	24 ± 3	~ 1

IN_{WT}, wild-type HIV-1 integrase
IC_{50}, 50% inhibitory concentration; calculated as the concentration of compound required to reduce the HIV-1 integrase strand transfer activity by 50%
SE, Standard error; standard error of the mean for at least three separate experiments
$FCIC_{50}$, fold-change in IC_{50}: calculated as the ratio of $IN_{Q148H/G140S}/IN_{WT}$
CC_{50}, 50% cytotoxic concentration; calculated as the concentration of compound required to reduce cell viability by 50%
EC_{50}, 50% effective concentration; defined as the concentration of compound required to reduce HIV-1 replication by 50%
SI, selectivity index; calculated as the ratio of CC_{50}/ EC_{50}
ND, not done

Acknowledgements

We thank the National Research Foundation of South Africa, Mintek, the University of the Witwatersrand and the University of KwaZulu-Natal for funding and permission to publish these findings.

Authors' contributions

R.H. was responsible for the conceptualisation, methodology, data analysis, validation, critically reviewing the writing, writing revisions, student supervision, project leadership, project management and funding acquisition. S.M. was responsible for the methodology, data analysis, validation, data curation, writing revisions, student supervision and project leadership. S.A. was responsible for the methodology, data collection, data analysis, sample analysis, validation, data curation and writing the initial draft. M.Q.F. was responsible for the methodology, data collection and data analysis. M.A.P. was responsible for the methodology, data analysis, validation, writing revisions and student supervision.

References

1. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov. 2015;14:475–486. https://doi.org/10.1038/nrd4609
2. McGovern SL, Caselli E, Grigorieff N, Schoichet BK. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem. 2002;45:1712–1722. https://doi.org/10.1021/jm010533y
3. McGovern SL, Helfand BT, Feng B, Shoichet BK. A specific mechanism of nonspecific inhibition. J Med Chem. 2003;46:4265–4272. https://doi.org/10.1021/jm030266r
4. Baell J, Walters MA. Chemistry: Chemical con artists foil drug discovery. Nature. 2014;513(7519):481–483. https://doi.org/10.1038/513481a
5. Trader DJ, Simanski S, Dickson P, Kodadek T. Establishment of a suite of assays that support the discovery of proteasome stimulators. Biochim Biophys Acta. 2017;1861(4):892–899. https://doi.org/10.1016/j.bbagen.2017.01.003
6. Cao JJ, Forrest C, Zhang X. A screen of the NIH Clinical Collection small molecule library identifies potential anti-coronavirus drugs. Antiviral Res. 2015;114:1–10. https://doi.org/10.1016/j.antiviral.2014.11.010
7. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc Natl Acad Sci USA. 1993;90:3428–3432. https://doi.org/10.1073/pnas.90.8.3428
8. Craigie R, Hickman AB, Engelman A. Integrase. In: Karn J, editor. HIV. Volume 2: A practical approach. Oxford: Oxford University Press; 1995. p. 53–71.

9. Harrison AT, Kriel FH, Papathanasopoulos MA, Mosebi S, Abrahams S, Hewer R. The evaluation of statins as potential inhibitors of the LEDGF/p75 – HIV-1 integrase interaction. Chem Biol Drug Des. 2014;85(3):290–295. https://doi.org/10.1111/cbdd.12384

10. Grobler JA, Stillmock KA, Hazuda DJ. Scintillation proximity assays for mechanistic and pharmacological analyses of HIV-1 integration. Methods. 2009;49(4):249–253. https://doi.org/10.1016/j.ymeth.2009.03.002

11. Hazuda DJ, Hastings JC, Wolfe AL, Emini EA. A novel assay for the DNA strand-transfer reaction of HIV-1 integrase. Nucleic Acids Res. 1994;22:1121–1122. https://doi.org/10.1093/nar/22.6.1121

12. Harada S, Koyanagi Y, Yamamoto N. Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. Science. 1985;229:563–566. https://doi.org/10.1126/science.2992081

13. Larder BA, Darby G, Richman DD. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science. 1989;243:1731–1734. https://doi.org/10.1126/science.2467383

14. Pauwels R, De Clercq E, Desmyter J, Balzarini J, Goubau P, Herdewijn P, et al. Sensitive and rapid assay on MT-4 cells for detection of antiviral compounds against the AIDS virus. J Virol Meth. 1987;16:171–185. https://doi.org/10.1016/0166-0934(87)90002-4

15. Mphahlele MK, Papathanasopoulos MA, Cinellu MA, Coyanis EM, Mosebi S, Traut T, et al. Modification of HIV-1 reverse transcriptase and integrase activity by gold(III) complexes in direct biochemical assays. Bioorg Med Chem. 2012;20:401–407. https://doi.org/10.1016/j.bmc.2011.10.072

16. Ingolfsson HI, Thukur P, Herold KF, Hobart EA, Ramsey NB, Periole X, et al. Phytochemicals perturb membranes and promiscuously alter protein function. ACS Chem Biol. 2014;9(8):1788–1798. https://doi.org/10.1021/cb500086e

17. ClinicalTrials.gov. A service of the U.S. National Institutes of Health [homepage on the Internet]. No date [cited 2017 Mar 20]. Available from: https://clinicaltrials.gov

18. Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M, et al. Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. Antimicrob Agents Chemother. 2008;52(4):1351–1358. https://doi.org/10.1128/AAC.01228-07

19. Witmer M, Danovich R. Selection and analysis of HIV-1 integrase strand transfer inhibitor resistant mutant viruses. Methods. 2009;47:277–282. https://doi.org/10.1016/j.ymeth.2009.02.025

20. Ravishankar S, Kumar VP, Chandrakala B, Jha RK, Solapure SM, De Sousa SM. Scintillation proximity assay for inhibitors of *Escherichia coli* MurG and, optionally, MraY. Antimicrob Agents Chemother. 2005;49:1410–1418. https://doi.org/10.1128/AAC.49.4.1410-1418.2005

An approach for the determination and correlation of diversity and efficiency of software development teams

AUTHORS:
Seher Altiner[1]
Mustafa B. Ayhan[1]

AFFILIATION:
[1]Industrial Engineering Department, Marmara University, Istanbul, Turkey

CORRESPONDENCE TO:
Mustafa Ayhan

EMAIL:
batuhan.ayhan@marmara.edu.tr

DATES:
Received: 16 Sep. 2017
Revised: 01 Nov. 2017
Accepted: 13 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
diversity; efficiency; data envelopment analysis; correlation; performance evaluation

HOW TO CITE:
Altiner S, Ayhan MB. An approach for the determination and correlation of diversity and efficiency of software development teams. S Afr J Sci. 2018;114(3/4), Art. #2017-0331, 9 pages. http://dx.doi.org/10.17159/sajs.2018/20170331

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
None

We examined the impact of diversity on team efficiency. To do so, a model was developed to measure both the efficiency and diversity of the teams. Based on these measures, the correlation between efficiency and diversity was also analysed. In addition, to demonstrate the applicability of the model, it was applied to a real-life problem involving five teams dealing with different software development projects. Firstly, diversity indices were calculated based on age, experience, education and gender information on each member for each team by using Simpson's Diversity Index. Then, four key performance indicators (KPIs) were defined to measure the success rate of the teams. Depending on these KPIs, efficiencies of the teams were measured through data envelopment analysis (DEA). The correlation between team efficiency and each diversity factor was analysed and all four factors had positive correlation with efficiency. That is, in order to increase efficiency, teams should be composed of members with diverse characteristics. Education was the diversity factor that had the most positive correlation with team efficiency. This result highlights the importance of different educational backgrounds on team efficiency.

Significance:

- This study represents the first attempt to measure team diversity using Simpson's Diversity Index.
- A new technique is proposed to measure team efficiency through DEA.
- Team efficiency is positively correlated with diversity, specifically educational level, which is important for many software development teams to consider.

Introduction

The efficiency of software development teams is an important topic. In order to optimise efficiency, team success rates should be monitored and evaluated. Currently, the success rate of software project development teams is much lower than desired. These projects are often completed years behind schedule and exceed their budgets by millions of dollars, and, if completed, sometimes fail to meet their users' needs. A long-standing key question that has intrigued the minds of researchers concerns the problem in managing software projects efficiently while promoting team performance.

As software development is a labour- and knowledge-intensive task, teamwork in software projects has been long acknowledged as a crucial criterion for the successful design and deployment of software projects. In search of factors for successful team performance, researchers have examined the various personality characteristics of team members. Every software project will inevitably face the issue of team composition. Whether increasing diversity in team composition promotes successful teamwork and further leads a project towards fulfilment of its mission, vision and values is therefore of interest to researchers.[1]

Organisational scholars considering the link between team diversity in a workgroup and the group's performance have generally concluded that the relationship is neither simple nor direct. In some studies, diverse groups outperformed homogeneous groups, while in others, homogeneous groups avoided the conflicts and communication problems that often beset diverse groups.[2]

Because team diversity can affect project performance, our aim in this study was to develop an integrated model to determine a relation between team diversity and the efficiency of software development project teams. To do so, firstly, a tool is presented to measure the diversity in software development teams by Simpson's Diversity Index according to four different diversity factors. Secondly, four key performance indicators are defined to be used to measure the team efficiency through data envelopment analysis (DEA). Then correlation analysis is performed between the team efficiency and each of the diversity factors. As a case study, this integrated model is also applied to check the correlation between diversity and efficiency of teams for software development.

The main contributions of this study are threefold. Firstly, although Simpson's Diversity Index is used in many areas, as far as we know, this is the first attempt to measure team diversity using Simpson's Index. Techniques to measure team diversity are very limited, inadequate and ambiguous in the literature. Because Simpson's Index is easy to clarify, it can be applied to measure team diversity satisfactorily. Secondly, a new technique is proposed to measure team efficiency through DEA. As there are many inputs and outputs affecting the performance of teams, it is not easy to assess their efficiency. Hence DEA can measure the relative performance of organisational units for which the presence of multiple inputs and outputs makes comparisons difficult.[3] Thirdly, by performing a correlation analysis, the relation between team diversity and efficiency can be revealed. This relation is important for many software development teams to understand which diversity factor has the greatest effect on efficiency.

Literature review

The literature is reviewed according to diversity measurement and key performance indicators in two sub-sections.

Diversity measurement

Diversity can be defined as acknowledging, understanding, accepting and valuing differences among people with respect to age, class, race, ethnicity, gender, disabilities, etc.[4] Companies need to embrace diversity and look for ways to become inclusive organisations because diversity has the potential to yield greater work productivity and competitive advantages.[5] The world's increasing globalisation requires more interaction among people from diverse backgrounds. Everyone is part of a worldwide economy competing within a global framework. For this reason, profit and non-profit organisations need to become more diversified to remain competitive.[6] Maximising and capitalising on workplace diversity is an important issue for management.

Many different types of workplace diversity have been studied[7], including age, gender, race and ethnicity, education, physical ability, culture, problem-solving ability, communication ability, motivation ability, listening ability and conflict resolution ability.

Having a diverse workforce provides many benefits such as increased adaptability, broader service range, a variety of viewpoints and more effective execution.[8,9] On the other hand, more diversity also creates many challenges, such as lack of effective communication, lack of freedom of speech, increased cost of training, integration problems, increased competition and disrespect.[10]

It is important to manage culturally diverse workforces with different perspectives.[11] In order to manage diversity, it should first be measured. Simpson's Diversity Index is one of the most commonly used indices to measure diversity, especially in ecology.[12] The Shannon Diversity Index has also been a popular diversity index in ecology, where it is also known as the Shannon–Weaver Index.[13] Moreover, McIntosh[14] developed a diversity index, which is independent of sample size and yields values which are a percentage of the maximum possible diversity for a sample of the same size. DeJong[15] analysed these three indices and compared them based on their richness and evenness components. Although there are various studies on diversity indices, the most commonly used one is Simpson's Index.[15] However, as stated by Pitts and Wise[16], workforce diversity research suffers from inadequate data and insufficient attention to empirical connections between diversity and organisational performance. So in the next section, the literature is reviewed regarding key performance indicators with respect to diversity.

Key performance indicators

In order to evaluate the efficiencies of teams, their performance should be measured. Performance measurement is a fundamental principle of management. It is important because it identifies performance gaps between current and desired performance and provides an indication of progress towards closing the gaps.[17] Performance measurement focuses on results. It is a process of assessing the results of a company, project, or an individual to determine how effective the operations are and it is also used to detect performance gaps, shortfalls, and other problems.[18] Many leading organisations employ an enterprise-wide formal performance measurement system such as Goal-Driven Measurement, Balanced Scorecard, Six Sigma, Practical Software and Systems Measurement, and variations of Shewhart's Plan-Do Check-Act Paradigm.[19] Each of these approaches emphasises the need to take the following steps:

- Set clear and achievable performance goals or objectives.
- Define key performance indicators (KPIs) to characterise performance relative to the goals or objectives.
- Establish targets for each KPI.
- Collect the measurement data (i.e. results).
- Evaluate the data and use the results to make adjustments in order to achieve the targeted levels for each KPI.[18]

Defining the KPIs is an important step. They help an organisation define and measure progress toward organisational goals. Once an organisation has analysed its mission, identified all its stakeholders, and defined its goals, it needs a way to measure progress toward those goals. Hence, KPIs can be used as tools to measure the success rate of this progress.[20]

KPIs are goals or targets that measure how well an organisation is achieving its overall operational objectives or critical success factors for a particular project.[21] KPIs must be objectively defined in order to provide a quantifiable and measurable indication of the organisation's progress towards achieving its goals.[22]

KPIs can change according to the organisation, mission and project. Despite hundreds of KPIs defined for software projects, there are some key measures that every organisation should collect and use as a basis to compare performance between projects. For software projects, the following KPIs are recommended: project effort, productivity, project duration, schedule predictability, requirements completion ratio and post-release defect density.[18]

Proposed model

An integrated model is proposed to measure the relation between team diversity and efficiency. This model includes four phases: diversity phase, key performance indicators phase, team efficiency phase, and correlation analysis phase, as shown in Figure 1. The first phase is devoted to evaluate team diversity according to four factors. The second phase is determination of the four KPIs to measure the performance of teams. The third phase is measuring the team efficiency based on the four KPIs through DEA. The last phase is a correlation analysis to detect whether there is a meaningful relation between diversity and efficiency of teams.

Measuring team diversity

Although there are numerous diversity factors in a work place, the most commonly used diversity factors were selected: age, gender, experience and education.

Age diversity

A workforce composed of different age demographics creates an environment in which each generation brings different skills and talents. For example, young professionals grew up in a high-tech world and have greater familiarity with business technology tools compared with their more mature counterparts, whereas more mature professionals often have exceptional interpersonal skills and perform well in environments in which traditional in-person communication is used.[23] Hence different age groups have different skills in terms of computer use and communication. Age diversity is an important factor to be studied in team efficiency.

Gender diversity

In the past, women in the workplace were automatically assigned to temporary, part-time or low responsibility jobs because it was thought that their first priority was taking care of their families. In addition, there was a widespread belief that women were not as capable as men, either physically or mentally or emotionally. Today, organisations are slowly adjusting to treating women and men equally. Discriminating against female employees (in terms of hiring and advancement) as well as treating them in a sexual manner (sexual harassment) are now against the law.[24]

It has been suggested that gender-diverse teams perform better than single-gender teams for several reasons:

- Men and women have different viewpoints, and different skills for problem solving.
- A gender-diverse workforce provides easier access to resources, and wider industry knowledge.
- A gender-diverse workforce allows the company to serve a diverse customer base.
- Gender diversity helps companies attract and retain talented women.[20]

Gender diversity is an important factor which affects team efficiency and hence it should be measured.

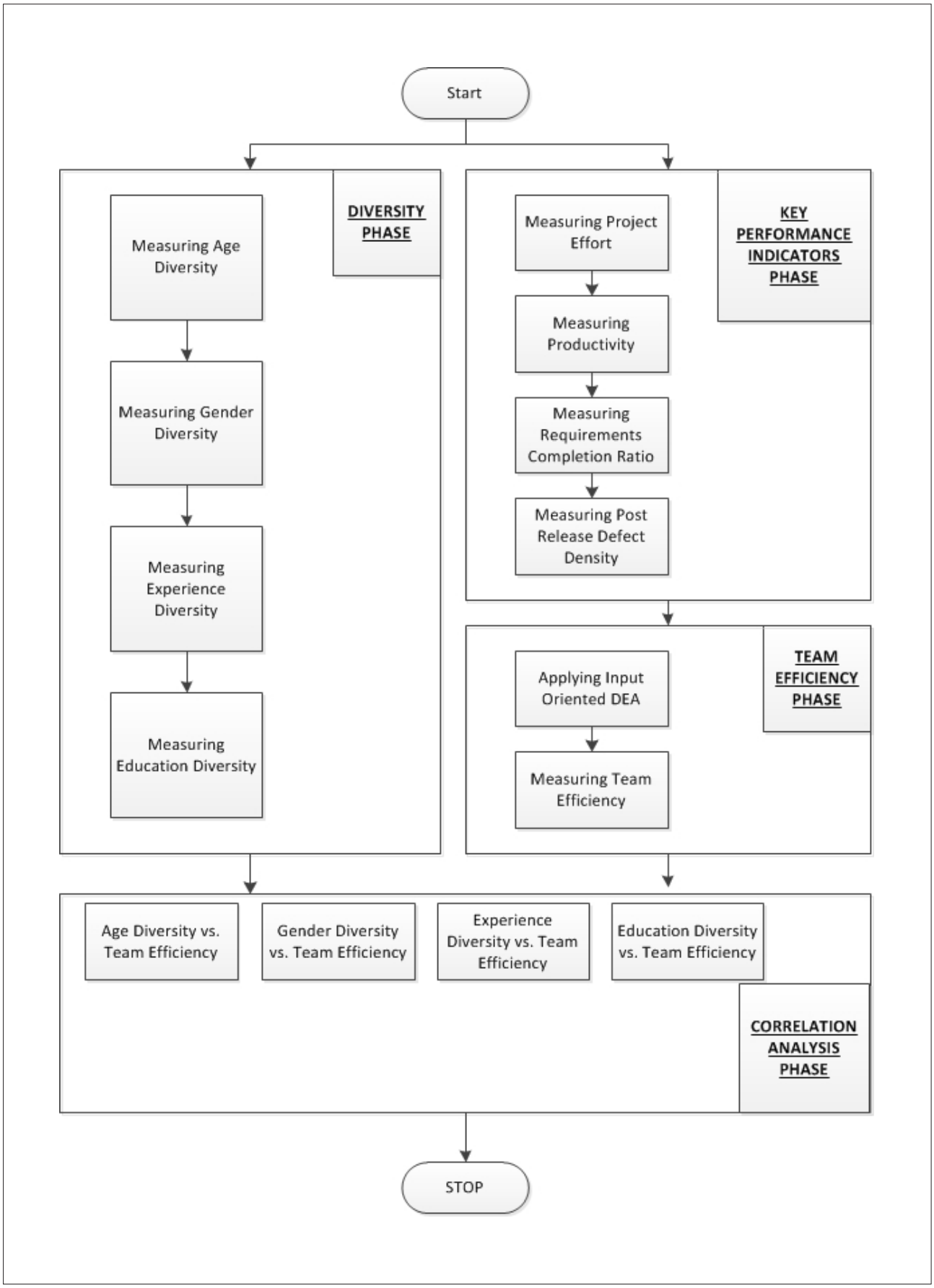


Figure 1: A summary of the research methodology employed.

Experience diversity

The concept of experience diversity is similar to that of informational diversity proposed by Jehn and Bezrukova[25]. Informational diversity has been measured by differences in education and functional area in the firm, such as position in the firm, major and level of education, or tenure.[25] Hence, in this study, experience diversity is referred to as tenure in the software project teams.

Education diversity

Education diversity refers to differences among members on the basis of technical knowledge and perspectives brought to the group. For example, team members who have a bachelor's degree, master's degree or doctorate have different skills and different perspectives. Because these skills and perspectives can affect the team efficiency in different ways, education diversity should be evaluated to detect if it has an effect on team efficiency.

Simpson's Diversity Index

The Simpson Index was introduced in 1949 by Edward H. Simpson to measure the degree of concentration when individuals are classified into types. The same index was described again by Orris C. Herfindahl in 1950. The square root of the index had already been introduced in 1945 by the economist Albert O. Hirschman. As a result, the same measure is usually known as the Simpson Index in ecology, and as the Herfindahl Index or the Herfindahl–Hirschman Index in economics. As far as we know, it has not yet been used to measure diversity in teams, but is well suited to the concept of team diversity. Moreover, it is a reasonable and simple measure of diversity in teams, as shown in Equations 1 and 2[12]:

$$\text{Simpson's Index} = \frac{\sum n(n-1)}{N(N-1)} \qquad \text{Equation 1}$$

$$\text{Diversity Index} = 1 - \frac{\sum n(n-1)}{N(N-1)}, \qquad \text{Equation 2}$$

where n is the total number of members of a particular species and N is the the total number of members of all species. The result is between 0 and 1, where 0 represents no diversity and 1 indicates maximum diversity.

Measuring team key performance indicators

As there are many criteria to measure the project performance of teams, it is important to determine which factors to use. Firstly, total time spent on the project by all team members ('project effort') is an important indicator to measure team performance. Secondly, 'productivity', which can be calculated by dividing the total time spent on the project by the number of team members, is also crucial to detect average performance. The third factor is the 'requirements completion ratio'. This ratio mainly calculates what percentage of the requirements can be satisfied at the end of the project. The last factor is the 'post-release defect density'. This indicator mainly focuses on the number of defects detected after the launch of the project. The total number of defects is divided by the number of team members to calculate the defect density. Each of these indicators and their calculations are explained in the following sections.

Project effort

Project effort is the total time spent on the software project by all of the team members. This total time should be that spent on only project-related activities during the life cycle of the project. Project-related activities are events in the software process life cycle for which effort data are collected and reported.

Activities that do not specifically contribute to the development and delivery of the software products are excluded from the calculation of project effort. For example, time spent on company-wide meetings, conferences, information seminars, and professional development training is excluded.[18] The formulation of project effort is given in Equation 3:

$$\text{Project Effort} = \sum_{i=1}^{N} \text{Working Hour}_i \qquad \text{Equation 3}$$

where Working Hour_i is the working hours spent on project-related activities for team member i and N is the total number of team members.

Productivity

Productivity can be simply expressed as output over input. For software project teams, outputs can vary (e.g. lines of code, function points, feature points, use cases, objects).[18] Also, the inputs can include many different factors such as labour hours, funds invested, etc. In this study, productivity can be calculated by Equation 4:

$$\text{Productivity} = \frac{\text{Project Effort}}{\text{Team Size}} \qquad \text{Equation 4}$$

Requirements completion ratio

The requirements completion ratio measures the extent to which planned functional requirements were satisfied in the final product implementation.[18] It is calculated as in Equation 5:

$$\text{Requirements Completion Ratio} = \frac{\text{Satisfied Requirements}}{\text{Planned Requirements}} * 100\% \qquad \text{Equation 5}$$

where 'planned requirements' is the number of requirements that were originally planned at the beginning of the project and those that were added or modified through negotiation with the user, and 'satisfied requirements' is the number of functional requirements that were satisfied in the delivered software product.

Post-release defect density

Post-release defect density is the number of unique defects per unit size discovered during the first 6 months after initial deployment of the software.[18] It is calculated as in Equation 6:

$$\text{Post-release Defect Density} = \frac{\text{Total Defects}}{\text{Team Size}} \qquad \text{Equation 6}$$

Measuring team efficiency

As explained previously, there are four KPIs to be measured. These KPIs have different units. Also, efficiency should consider inputs as well as outputs because it signifies a level of performance that describes a process that uses the lowest amount of inputs to create the greatest amount of outputs. Efficiency relates to the use of all inputs in producing any given output, including personal time and energy. It is a measurable concept that can be determined by determining the ratio of useful output to total input. It minimises the waste of resources such as physical materials, energy and time, while successfully achieving the desired output. The usual measure of efficiency is shown in Equation 7:

$$\text{Efficiency} = \frac{\text{Output}}{\text{Input}} \qquad \text{Equation 7}$$

However, this measure is often inadequate because of the existence of multiple inputs and outputs related to different resources, activities and environmental factors. Moreover, the units of these multiple outputs and inputs are generally different from each other. Therefore, it is not easy to calculate efficiency.

The DEA model allows relative efficiency measures as a result of multiple inputs and outputs. It is a linear programming based technique for measuring the relative performance of organisational units in which the presence of multiple inputs and outputs makes comparisons difficult.[3] It was originally developed by Charnes et al.[26] and Banker et al.[27] It has been widely used to measure performance in many areas. A key advantage of DEA is that it easily accommodates both multiple inputs and multiple outputs.

A range of DEA models has been developed to measure efficiency and capacity in different ways. These models largely fall into the categories of input-oriented or output-oriented models. With input-oriented DEA, the linear programming model is configured so as to determine how much a firm could contract if inputs are used efficiently in order to achieve the same output level. Adenso-Diaz et al.[28] used an input-oriented DEA model to assess the performance of the teams in a company and to estimate the relative efficiency scores of the projects. In their study, the project efficiencies were calculated based on three inputs – namely cost, duration and difficulty – and one output, namely project revenue.

In contrast, with output-oriented DEA, the linear program is configured to determine a firm's potential output given its inputs if it operated as efficiently as firms along the best practice frontier. Nazari-Shirkouhi and Keramati[29] used an output-oriented DEA model for selecting the best fuzzy regression models among 16. In their study, these fuzzy regression models were considered to understand the relationship between customer satisfaction and new product design.

The following steps were followed in the DEA method:

- Determine the number of service units being compared in the DEA analysis.
- Determine the inputs and outputs.
- Evaluate the efficiency rating of the service units by DEA.

Based on this DEA analysis, team efficiency can be evaluated through input and output parameters. However, because of the structure of the problem, it is more appropriate to use an input-oriented approach as in Adenso-Diaz et al.[28]

Correlation between diversity and efficiency

The correlation coefficient is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.[30]

The correlation coefficient can vary from +1 to -1. Values close to +1 indicate a high degree of positive correlation, and values close to -1 indicate a high degree of negative correlation. It can be calculated as in Equation 8:

$$\text{Corr}(x,y) = \frac{n\sum(xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2]*[n\sum y^2-(\sum y)^2]}} \qquad \text{Equation 8}$$

where x and y are the variables for which the correlation is to be determined.

In this study, the correlations between team efficiency and four diversity factors were calculated separately to determine which had the greater relation with performance.

Case study

The proposed model was applied to measure the correlation between different diversity factors and team efficiency at a bank's credit application development department. This department is developing a new banking package for the banking credit system within the .Net environment. In this department there are numerous teams dealing with different software development projects. For reasons of confidentiality, the name of the bank cannot be given, but the data can be revealed.

Five project teams were selected within the Credit Application Development Department as indicated below:

- Team 1: Retail Loans Team
- Team 2: Non-performing Loans Team
- Team 3: Revolver Loans Team
- Team 4: Instalment Loans Team
- Team 5: Loan Allocation Team

In addition, the age, experience, education and gender information of each member of each team was gathered from the Human Resources Department. The information required to calculate the performance of each team was also gathered.

Firstly, four diversity indices for each factor were calculated for all teams. Secondly, the performance of each team was measured according to four KPIs. Thirdly, based on these four KPIs, team efficiencies were measured using DEA. Lastly, the correlations of each factor with efficiency were analysed.

Measuring diversity index

In order to measure the diversity indices for age, gender, experience and education, the following criteria were applied.

- Team members were classified by age into three groups: young (those less than or equal to 30 years old); middle (those between 31 and 40 years old); and old (those older than or equal to 41 years old).
- Gender classification was female or male.
- Experience was categorised into three groups: working for less than 5 years; working for between 5 and 8 years; and working for more than 9 years. These years reflect the durations determined by the bank's Human Resources Department for promotions.
- Education was classified by bachelor's degree; master's degree; and doctorate.

According to the diversity index formula given in Equation 2, diversity indices of each factor for each team were calculated and are given in Table 1.

For example, the age diversity index for Team 1 can be calculated as in Equation 9. Team 1 is composed of nine members of whom two are young, six are middle-aged and one is old.

$$\text{Diversity Index} = 1 - \frac{[2*(2-1)]+[6*(6-1)]+[1*(1-1)]}{9*(9-1)} = 1-\frac{32}{72} = 0.56$$

Equation 9

Measuring key performance indicators

After calculating the diversity indices of each factor for each team, team performances were measured using the four KPIs: project effort, productivity, requirements completion ratio, and post-release defect density. For the sake of comparisons between teams, some assumptions were needed to sustain the uniformity and standardisation. The project period was assumed to be 180 days for each team. Starting and finishing dates of each project were assumed to be the same.

Project effort is the total time spent on the project by all team members. So, the average daily working hours of each team member on project-related activities is multiplied by 180 days to determine the total working hours spent on the project. Lastly, the summation of project efforts of all team members gives the total project effort. For example, total project effort for Team 1 is calculated as 10 006.40 hours and shown in Table 2.

By similar calculations, the total project efforts of all teams are given in Table 3.

Productivity is the ratio of total project effort over the number of team members. This ratio gives the average working hours for each worker for project-specific activities. For example, Team 1 is composed of nine members and has a project effort of 10 006.40 working hours. So the productivity of Team 1 is calculated as 1111.8222. When this figure is divided by the number of days, average daily productivity can be determined; 6.18 in this example, as shown in Table 4.

The requirements completion ratio can be calculated by dividing the satisfied requirements over planned requirements. A high ratio reflects good customer satisfaction. For each of the project teams, the number of planned and satisfied requirements are recorded and given in Table 5.

Table 1: Diversity indices of five teams according to four factors

	Age		Gender		Experience		Education	
	Group	***n***	**Group**	***n***	**Group**	***n***	**Group**	***n***
Team 1	Young	2	Male	4	≤5 years	3	Bachelor's degree	8
	Middle	6	Female	5	5–8 years	2	Master's degree	1
	Old	1			≥9 years	4	Doctoral degree	0
Diversity index	**0.56**		**0.56**		**0.72**		**0.22**	
Team 2	Young	3	Male	2	≤5 years	2	Bachelor's degree	5
	Middle	2	Female	4	5–8 years	3	Master's degree	0
	Old	1			≥9 years	1	Doctoral degree	1
Diversity index	**0.73**		**0.53**		**0.73**		**0.33**	
Team 3	Young	2	Male	5	≤5 years	1	Bachelor's degree	5
	Middle	3	Female	1	5–8 years	2	Master's degree	1
	Old	1			≥9 years	3	Doctoral degree	0
Diversity index	**0.73**		**0.33**		**0.73**		**0.33**	
Team 4	Young	2	Male	8	≤5 years	2	Bachelor's degree	8
	Middle	4	Female	0	5–8 years	2	Master's degree	0
	Old	2			≥9 years	4	Doctoral degree	0
Diversity index	**0.71**		**0.00**		**0.71**		**0.00**	
Team 5	Young	2	Male	5	≤5 years	1	Bachelor's degree	7
	Middle	4	Female	2	5–8 years	2	Master's degree	0
	Old	1			≥9 years	4	Doctoral degree	0
Diversity index	**0.67**		**0.48**		**0.67**		**0.00**	

Table 2: Total project effort for Team 1

Team member	Average daily working hours	Project duration (days)	Project effort
Member 1	6.95	180	1251.00
Member 2	7.80	180	1404.00
Member 3	6.00	180	1080.00
Member 4	5.75	180	1035.00
Member 5	7.20	180	1296.00
Member 6	7.20	180	1296.00
Member 7	7.00	180	1260.00
Member 8	7.08	180	1274.40
Member 9	2.00	55	110.00
Total project effort (hours)			**10 006.40**

Table 3: Total project effort for all teams

Team	Total project effort (hours)
Team 1	10 006.40
Team 2	7838.20
Team 3	8544.60
Team 4	11 244.60
Team 5	9108.00

Table 4: Average daily productivity for all teams

Team	Team size	Project effort (hours)	Productivity	Daily productivity
Team 1	9	10 006.40	1111.82	6.18
Team 2	6	7838.20	1306.37	7.26
Team 3	6	8544.60	1424.10	7.91
Team 4	8	11 244.60	1405.58	7.81
Team 5	7	9108.00	1301.14	7.23

For example, Team 1 had 252 planned requirements to complete at the beginning of the project. However, at the end, 251 of them were satisfied. Therefore, the requirements completion ratio for Team 1 is calculated as 99.60%.

Table 5: Requirements completion ratio for all teams

Team	Planned requirements	Satisfied requirements	Requirements completion ratio (%)
Team 1	252	251	99.60
Team 2	180	142	78.89
Team 3	178	171	96.07
Team 4	357	286	80.11
Team 5	482	352	73.03

Post-release defect density can be calculated by dividing the total number of defects detected within 6 months after launching the project by the number of team members. The defect density of each team member is shown in Table 6. For example, for Team 1, the total number of defects after release was 5041. Team 1 has nine members, therefore the post-release defect density is 560.11. Divided by 180 days, the average daily defect for each member of Team 1 is 3.11.

A comparison of the four KPIs of each of the five teams is given in Table 7.

Table 6: Post-release defect density for all teams

Team	Total number of defects	Team size	Average number of defects per team member	Average number of daily defects per team member
Team 1	5041	9	560.11	3.11
Team 2	2548	6	424.67	2.36
Team 3	5001	6	833.50	4.63
Team 4	2223	8	277.88	1.54
Team 5	3827	7	546.71	3.04

Table 7: Comparison of the four key performance indicators for each team

Team	Total project effort (hours)	Daily productivity	Requirements completion ratio (%)	Post-release defect density for each member
Team 1	10 006.40	6.18	99.60	3.11
Team 2	7838.20	7.26	78.89	2.36
Team 3	8544.60	7.91	96.07	4.63
Team 4	11 244.60	7.81	80.11	1.54
Team 5	9108.00	7.23	73.03	3.04

Measuring efficiency of teams

In order to calculate the efficiency of each team using DEA, the daily available time of each team should be found. Although each team member can work 8 hours a day, they also have to deal with post-release defects. As 15 minutes is the accepted time spent for each defect in the bank in general, the time spent on post-release defects was assumed to be 15 minutes for each defect.

The calculated time spent for defects (in minutes) was converted into hours to compare daily work time (8 hours) and the time spent on defects. For example, for Team 1, the average daily defect for each member was 3.11 (Table 6). When multiplied by 15 minutes, the average daily time spent for each member of Team 1 is 46.65 minutes or 0.78 hours. Hence, the available daily maximum working hours per member of Team 1 is 7.22 hours as shown in Table 8.

Table 8: Time spent for each post-release defect (hours) and available working hours of team members for all teams

Team	Daily number of post-release defects	Daily time spent for post-release defects (minutes)	Daily time spent for post-release defects (hours)	Daily maximum work hours per team member
Team 1	3.11	46.65	0.78	7.22
Team 2	2.36	35.40	0.59	7.41
Team 3	4.63	69.45	1.16	6.84
Team 4	1.54	23.10	0.39	7.61
Team 5	3.04	45.60	0.76	7.24

In order to calculate the efficiency of each team by DEA, the service units being compared in the DEA analysis were determined as well as the input and output parameters (Table 9). MaxDEA 7 Basic x64 software was used to perform the DEA to evaluate the efficiency rating of all the teams by input-oriented choice; the results are shown in Table 10. These efficiency ratings were used to calculate correlations between efficiency and the various diversity indices.

Table 9: The number of service units compared for each team in the data envelopment analysis

Team	Team size (input)	Project effort per team (input)	Daily productivity (output)	Requirements completion ratio per team (%) (output)	Daily maximum work hours per team member (output)
Team 1	9	10 006.40	6.18	99.60	7.22
Team 2	6	7838.20	7.26	78.89	7.41
Team 3	6	8544.60	7.91	96.07	6.84
Team 4	8	11 244.60	7.81	80.11	7.61
Team 5	7	9108.00	7.23	73.03	7.24

Table 10: Efficiency of all teams

Team	Efficiency
Team 1	0.89
Team 2	1.00
Team 3	1.00
Team 4	0.79
Team 5	0.86

Correlation between team diversity and team efficiency

The efficiency ratings for each team and the diversity indices of each factor for each team are summarised in Table 11. Correlation analyses were performed for each diversity factor against team efficiency to determine the correlation coefficients.

All diversity factors show a positive correlation with team efficiency. However, education diversity shows the highest correlation with efficiency. Therefore, in forming software development project teams, it is better to select employees with different educational backgrounds. In addition, gender diversity and experience diversity show a medium effect on team efficiency. Age diversity had the lowest effect on team efficiency.

Table 11: Correlation between team efficiency and diversity factors

Teams	Age diversity index	Gender diversity index	Experience diversity index	Education diversity index	Team efficiency
Team 1	0.56	0.56	0.72	0.22	0.89
Team 2	0.73	0.53	0.73	0.33	1.00
Team 3	0.73	0.33	0.73	0.33	1.00
Team 4	0.71	0.00	0.71	0.00	0.79
Team 5	0.67	0.48	0.67	0.00	0.86
Correlation coefficient	**0.30**	**0.56**	**0.57**	**0.93**	

Notably, Team 4 has the lowest diversity indices for all factors and their team efficiency is also low. This result highlights the importance of diversity among team members. Figure 2 summarises these results. We can conclude that software project teams should have a high degree of diversity. All team members differ in knowledge, social category and values. The importance of this diversity may be especially relevant for software development teams because software projects are complex and need different skills at different stages of the project.

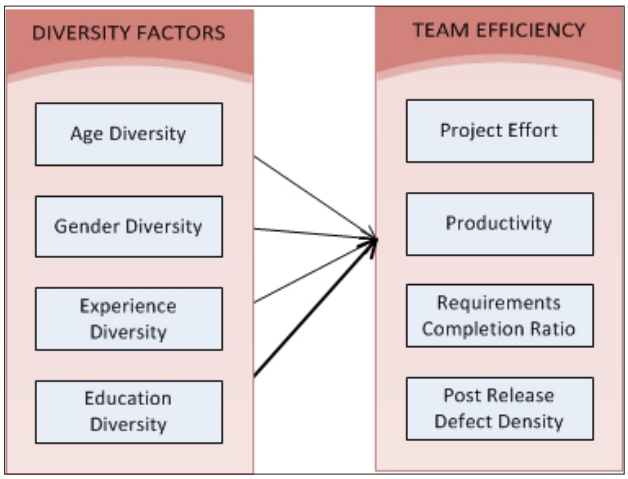


Figure 2: Diversity factors and their effects on software development team efficiency.

Conclusion

This study focused on the relation between efficiency and diversity of software development project teams. Based on four diversity factors (age, gender, experience and education), a methodology was proposed to measure the diversity index using Simpson's formula. Four KPIs (project effort, productivity, requirements completion ratio and post-release defect density) were used to measure efficiency by DEA. Lastly, correlation analyses were performed to detect the relations between each diversity factor and efficiency, using five software development project teams at a bank's Credits Application Development Department as a case study. Based on the result of this real-life application, we conclude that the more diverse teams – with respect to the four selected factors – were more efficient, and thus more successful, in their software development projects.

Some inferences can be made with respect to the managerial implications of our findings. Project team managers should consider different kinds of diversity when forming their teams or hiring new employees in order to increase the team's efficiency. Moreover, managers should also support different training alternatives for their employees. The staff with diverse educational backgrounds can open new horizons for the company. Although homogeneity in the workforce may provide for a smoother working environment, greater diversity in teams enhances their efficiency. Managing diverse teams may include ensuring harmony between different team members.

Although there are other methods, using Simpson's Index to calculate the diversity index of teams is a novel approach in this research area. In addition, using DEA to measure the efficiency of teams was a new application.

This study could be improved by adding other kinds of diversity to the methodology. For example, the ability to motivate other people and the ability to communicate are also important characteristics in a team and their diversities can affect team efficiency. This methodology should also be applied to other types of teams to strengthen the results applicable to software development teams. If the findings in this study are also valid for other teams, then the results can be generalised.

Authors' contributions

S.A. was responsible for the methodology, validation, data collection, data curation, data analysis, sample analysis and writing the initial draft. M.B.A. was responsible for the conceptualisation, methodology, validation, data curation, critically reviewing the first draft and revisions, student supervision, project leadership and project management.

References

1. Liang T, Liu C, Lin T, Lin B. Effect of team diversity on software project performance. Ind Manage Data Syst. 2007;107(5):636–653. https://doi.org/10.1108/02635570710750408
2. Ely RJ. A field study of group diversity, participation in diversity education programs, and performance. J Organ Behav. 2004;25:755–780. https://doi.org/10.1002/job.268
3. Emrouznejad A, Parker RP, Tavares G. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Socio Econ Plan Sci. 2008;42(3):151–157. https://doi.org/10.1016/j.seps.2007.07.002
4. Esty K, Griffin R, Schorr-Hirsh H. Workplace diversity. A manager's guide to solving problems and turning diversity into a competitive advantage. Holbrook, MA: Adams Media Corporation; 1995.
5. Green K, Lopez M, Wysocki A, Kepner K, Farnsworth D, Clark JL. Diversity in the workplace: Benefits, challenges, and the required managerial tools [document on the Internet]. c2002 [updated 2015 Oct; cited 2017 Nov 01]. Available from: https://edis.ifas.ufl.edu/pdffiles/HR/HR02200.pdf
6. Komaki JL, Minnich MLR. A behavioral approach to organizational change: Reinforcing those responsible for facilitating the climate and hence promoting diversity. J Organ Behav Manage. 2016;36(2–3):154–184. https://doi.org/10.1080/01608061.2016.1200514
7. Konrad AM, Prasad P, Pringle JK. Handbook of workplace diversity. Wiltshire, UK: SAGE; 2006.
8. Greenberg J. Diversity in the workplace: Benefits, challenges and solutions [document on the Internet]. c2004 [cited 2017 Nov 01]. Available from: http://www.multiculturaladvantage.com/recruit/diversity/diversity-in-the-workplace-benefits-challenges-solutions.asp

9. Marshall e-Learning Consultancy. Understanding equality and diversity in the workplace [document on the Internet]. c2015 [cited 2017 Nov 01]. Available from: http://marshallelearning.com/blog/understanding-equality-and-diversity-in-the-workplace/#sthash.xtl1PUiU.dpuf
10. Fleming RK. Diverse problems call for diverse analyses and interventions in behavioral safety. J Organ Behav Manage. 2001;21(1):65–69. https://doi.org/10.1300/J075v21n01_05
11. Podsiadlowski A, Gröschke D, Kogler M, Springer C, Van der Zee K. Managing a culturally diverse workforce: Diversity perspectives in organizations. Int J Intercult Rel. 2013;37:159–175. https://doi.org/10.1016/j.ijintrel.2012.09.001
12. Simpson EH. Measurement of diversity. Nature. 1949;163:688. https://doi.org/10.1038/163688a0
13. Shannon CE, Weaver W. The mathematical theory of communication. Urbana, IL: University of Illinois Press; 1949.
14. McIntosh RP. An index of diversity and the relation of certain concepts to diversity Ecology. 1967;48:392–404. https://doi.org/10.2307/1932674
15. De Jong TM. A comparison of three diversity indices based on their components of richness and evenness. OIKOS. 1975;26:222–227. https://doi.org/10.2307/3543712
16. Pitts DW, Wise LR. Workforce diversity in the new millennium: Prospects for research. Rev Public Pers Adm. 2009;30(1):44–69. https://doi.org/10.1177/0734371X09351823
17. Lavy S, Garcia JA, Dixit MK. Establishment of KPIs for facility performance measurement: review of literature. Facilities. 2010;28(9–10):440–464. https://doi.org/10.1108/02632771011057189
18. Kasunic M. A Data Specification for Software Project Performance Measures: Results of a Collaboration on Performance Measurement. Technical Report, 2008. CMU/SEI-2008-TR-012, ESC-TR-2008-012, Software Engineering Institute, Carnegie Mellon University, USA
19. Abernathy WB. Beyond the Skinner Box: The Design and Management of Organization-Wide Performance Systems. J Organ Behav Manage. 2014;34(4):235–254. https://doi.org/10.1080/01608061.2014.973631
20. Badal S, Harter JK. Gender Diversity, Business-Unit Engagement, and Performance. J Leadersh Org Stud. 2013;21(4):354–365. https://doi.org/10.1177/1548051813504460
21. Binder C. Integrating Organizational-Cultural Values With Performance Management. J Organ Behav Manage. 2016;36(2-3):185–201. https://doi.org/10.1080/01608061.2016.1200512
22. Parmenter D. Key Performance Indicators: Developing, Implementing, and Using Winning KPIs. New Jersey, USA. John Wiley and Sons Inc., 2015. https://doi.org/10.1002/9781119019855
23. Hopkins WE, Hopkins SA. Diversity Leadership: A Mandate for the 21st Century Workforce. J Leadersh Stud. 1999;5(3):129–140. https://doi.org/10.1177/107179199900500311
24. Cundiff NL, Komarraju M. Gender Differences in Ethnocultural Empathy and Attitudes Toward Men and Women in Authority. J Leadersh Org Stud. 2008;15(1):5–15. https://doi.org/10.1177/1548051808318000
25. Jehn K, Bezrukova K. A field study of group diversity, workgroup context, and performance. J Organ Behav. 2004;25(6):703–729. https://doi.org/10.1002/job.257
26. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. Eur J Oper Res. 1978;2(6):429–444. https://doi.org/10.1016/0377-2217(78)90138-8
27. Banker R, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci. 1984;30(9):1078–1092.
28. Adenso-Diaz B, Lozano S, Gutierrez E, Calzada L, Garcia S. Assessing individual performance based on the efficiency of projects. Comput Ind Eng. 2017;107:280–288.
29. Nazari-Shirkouhi S, Keramati A. Modeling customer satisfaction with new product design using a flexible fuzzy regression data envelopment analysis algorithm. Appl Math Model. 2017;50:755–771.
30. Devore JL. Probability and statistics for engineering and the sciences. 8th ed. Boston, MA: Richard Stratton; 2012.

The role of the Square Kilometre Array in South Africa's economic development strategy

AUTHOR:
Nishana Bhogal[1]

AFFILIATION:
[1]Graduate School of Business, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Nishana Bhogal

EMAIL:
nishanabhogal@gmail.com

DATES:
Received: 28 Aug. 2017
Accepted: 21 Nov. 2017
Published: 27 Mar. 2018

KEYWORDS:
knowledge-based economy; 4I model; institutions; innovation; astronomy

HOW TO CITE:
Bhogal N. The role of the Square Kilometre Array in South Africa's economic development strategy. S Afr J Sci. 2018;114(3/4), Art. #2017-0297, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170297

ARTICLE INCLUDES:
✓ Supplementary material
× Data set

FUNDING:
None

This study was undertaken to understand factors inhibiting and enabling the impact of the Square Kilometre Array South Africa (SKA SA) on the South African knowledge economy. A critical review of relevant literature revealed four themes, which are considered to be the main pillars crucial for engendering a knowledge economy: institutions, interrelationships, innovation and individuals. These pillars form the basis for the 4I model developed in this paper, the relevance of which in stimulating a knowledge economy was investigated. This study revealed no additional pillars, thus validating the 4I model in relation to SKA SA's contribution to the knowledge economy. SKA SA's success is underpinned by open and inclusive institutions, fostering and leveraging interrelationships, promoting innovation that may be commercialised, and attracting, retaining and training suitable individuals. Furthermore, this study provides a deeper insight into the 4I model by revealing new sub-themes that apply in a broader context, including the role of a nation's inherent competitive advantage in informing its competitive and innovation strategy, the nature of interrelationships that may be multidimensional, and politically astute leadership that is crucial for the ongoing support of a publicly funded project. This deeper understanding of the 4I model forms a basis for strengthening each pillar and its impact on the knowledge economy.

Significance:

- The 4I model, which is necessary for engendering a knowledge-based economy, is introduced.
- The role of a nation's inherent competitive advantage may inform its competitive and innovation strategy.
- A nation's institutions must be both inclusive and open for a knowledge-based economy to thrive.
- Interrelationships may be multidimensional in nature, including multidisciplinary, international and cross-sector collaboration.
- Politically astute leadership is crucial for the ongoing support of publicly funded projects.

Introduction

The main instrument and resource for economic development has evolved over time. Land was a key resource during a predominantly agrarian era.[1] Next, technological advances made way for the industrial revolution, which saw machinery as the key resource. More recently, further technological advances, mainly in information and communication technology (ICT), fostered the rise of an information age in which data and knowledge are the key resources.[1] Developed economies have turned to knowledge creation as an approach to realise competitive advantage and for sustained economic growth. Although South Africa has embarked on several initiatives to stimulate the economy, economic growth has steadily declined (refer to Supplementary figure 1 for South Africa's growth trend) and the economic outlook remains bleak.[2-4] In its search for sustained economic growth, South Africa is pursuing several knowledge-based initiatives.[5] The Square Kilometre Array South Africa (hereafter SKA SA or the project) is one such initiative.[6] The role and contribution of SKA SA to South Africa's knowledge economy is explored in this paper. The terms knowledge economy and knowledge-based economy (KBE) are used interchangeably hereafter.

Literature review

A KBE is an economy in which knowledge creation and its communication are the major engine for wealth generation and employment creation.[7] It results from the full recognition and exploitation of knowledge and technology as drivers for economic growth.[8] The creation of knowledge enables increased competitive advantage, not only for businesses but also for a national economy, which is difficult to imitate.[9-13] The concept of knowledge being linked to economic growth is not new, and the relationship has been acknowledged by Adam Smith, a pioneer of political economy.[8] However, knowledge remains directly linked to new economic growth theory[9], and must be converted, using innovation, and commercialised for economic growth[8]. For effective commercialisation, there must be close collaboration between industry, universities[1,8] and government: the role players in the triple helix model[14,15]. Although government may not be responsible for conducting research, innovation and commercialisation, it is responsible for establishing institutions with incentives such that a KBE may be fostered, such as fora for triple helix collaboration.

Government is the lynchpin in establishing and promoting a KBE. The determination of whether a nation has inclusive or extractive institutions, as described by Acemoglu and Robinson[16], refers to a nation's internal dynamics and is a direct consequence of policies adopted and implemented by its government[15,16]. In addition, nations need political and economic openness, which applies within a nation's borders as well as to its relationships with other nations.[1] Government is responsible for guiding the market system and thus economic growth.[17] A synchronised strategy is required, which accounts for institutions as well as policies required to enable the creation of a KBE. Importantly, the success of an institution hinges on the extent to which it is embedded in the economy. In addition, governments may adopt a structured ICT approach to support their synchronised strategy, thus closing

the gap between developed and developing economies by leveraging the learning of other nations, without following the same lengthy learning process.[1,18-20]

The role of individuals is another key factor influencing the creation of a KBE, as knowledge workers propel a KBE through their discoveries. Therefore, individuals must be supported and incentivised to innovate and discover new knowledge.[8] However, to support greater participation in the KBE, a broader set of incentives, in addition to the higher income earned, may be required. A diverse workforce or team promotes the convergence of different ideas, thus promoting creativity, and often yields commercially innovative solutions.[21,22] Diverse participation extends to the inclusion of women and the youth. A full and diverse workforce is required to optimise economic growth through innovation.[1] Although South Africa has some knowledge-based initiatives, such as SKA SA, transformation remains elusive. Income and wealth distribution continues to be inequitable and business structures remain non-representative.[23] Furthermore, the South African economy does not have full participation of its citizenry.[24] South Africa needs to identify sustainable initiatives, including those driven by innovation, to uplift those living below the poverty line and determine how to effect transformation such that broader citizenry contribution and participation takes place.

Science and technology (S&T) endeavours often result in innovation. The Organisation for Economic Co-operation and Development[8] considered that innovation might be either mission-oriented or diffusion-oriented. Typically, the benefits of fundamental research are realised in the long term.[8] Low-cost manufacturing and 'copy-cat innovation'[1(p.217)] may yield short-term benefits. However, fundamental research is essential for innovation, which yields competitive advantage, and thus sustained economic growth.[1,25-27] Innovation and development go hand in hand.[12] Global trends and technological advances have far-reaching consequences,[9,28,29] and their significance should influence national strategies and policies. Technological advances are disrupting existing markets and industries, thus fundamentally influencing how firms compete[10] and how national economies remain relevant. Mainly as a result of pressure from investors for short-term gains, the responsibility to drive fundamental research has shifted from corporations to government.[20] Fundamental research is extremely risky and by government funding fundamental research, market risk is reduced.[26,27] Skilled individuals and corporations will be attracted through government fostering of a collaborative climate in which innovation is pursued.[1,30] Moreover, S&T and innovation create the potential for developing nations to leapfrog or close the economic divide that exists between developed and developing economies, because developing economies do not have to follow the same innovation path previously followed by developed economies.[1,18,19] Seemingly, a paradox exists: sustained economic growth often depends on cutting-edge innovation, and the scale of cutting-edge innovation demands collaboration or sharing, which compromises competitive advantage. Indeed, sustained competitive advantage demands sustained and novel innovation. One leading scientific research area, which arguably rests at the pinnacle of innovation, is astronomy.

Astronomy – which refers collectively to astronomy, astrophysics and cosmology hereafter – stirs the imagination and spirit, and 'enjoys broad public appeal', from people of all ages[31(p.103)], by inducing awe and wonder[31]. It is the oldest science[31,32], and arguably has had the greatest enduring impact on civilization[31]. The human intrigue with the heavens has existed from time immemorial and inspired great feats.[32] Among its earliest uses, astronomy shaped the progress of agriculture and navigation.[32,33] Lunar and solar cycles underpinned the recording of time, and continue to do so even in modern time.[32] Arguably, the greatest technology-led breakthrough for astronomy was in 1609, when Galileo Galilei first used a telescope to view the night sky.[31] Since then telescopes have steadily improved.[31] Spurred by technological advances, astronomy has made unprecedented progress in recent decades.[31,34] Innovation resulting from astronomy has had far-reaching commercial spin-offs[35-37] (refer to Supplementary table 1 for a list of commercial spin-offs from astronomy-led initiatives). Astronomy embarks on fundamental research endeavours.[35,37] Although the fruit of these endeavours is often only realised in the long term[31], investment in astronomy fosters innovation, enhances educational opportunities in science, technology, engineering and mathematics (STEM) and contributes to developing the workforce[31,35,37]. Importantly, the high-quality feeder from both primary and secondary education significantly influences the success of universities to attract and develop good students.[20,31] Training must be supported by 'significant investments in world-class and innovative research infrastructure, [which enables a nation] to attract and retain excellent researchers'[37(p.7)].

The scale of modern day astronomical projects demands international collaboration[31], because partner countries may leverage cutting-edge data and technologies, as well as scientific funding[37]. Notably, these partnerships are not limited to partnerships between different nations, but may include triple helix collaborations[31,37], which are promoted by nations' institutions[36]. In order for commercialisation to be fast-tracked, researchers require incentives to promptly share findings with the public and industry.[37] Astronomy may underpin a nation's innovation plan.[31,37] A coherent strategy for astronomy is necessary to optimise the impact of available resources and may be an asset in negotiating international partnerships.[31] Astronomy is heavily reliant on government for funding. During periods of sluggish economic growth, governments may opt for short-term prioritisation, thus limiting or foregoing initiatives with benefits that are realised in the long term, such as astronomy-related fundamental research endeavours. Astronomy in South Africa lacks a comprehensive strategy and has faced budget reductions.[38] South African astronomy requires sustained funding and a comprehensive and coherent strategy, commencing with primary and secondary education, which is the feeder into universities. The quality of education, especially that of mathematics and science, is of great concern in South Africa.[39] Furthermore, the plan should ensure that institutions are established in which triple helix collaborations thrive and knowledge is diffused. The plan should identify astronomy projects, such as SKA SA, that will be pursued so that resources appropriated for astronomy in South Africa may be focused.

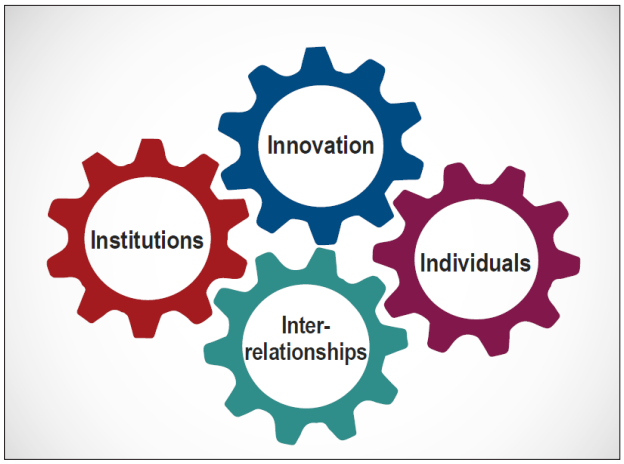
Arguably, as South Africa's largest science project and one of 18 Special Infrastructure Projects,[38,40] SKA SA plays a major role in the endeavour to build the world's largest radio telescope. Co-located in South Africa and Australia, the telescope is expected to have a total collecting area of approximately one square kilometre.[41] Significant technological advances are required to complete the telescope's construction, which will result in image resolution superior to that of the Hubble Space Telescope by a factor of 50.[42] The SKA telescope will observe stars and galaxies up to billions of light years away to better understand the formation and evolution of celestial bodies.[41] It will be constructed in phases to manage risk and to allow technological advancement.[43] Notably, as more dishes are added to the telescope, more data will be generated (refer to Supplementary figure 2 for the anticipated data volume of SKA SA). This iconic project has been attracting the best scientific talent from around the world. SKA will pioneer advances in high-performance and big data computing, provide a platform to create skills for KBEs, and stimulate interest in mathematics-related careers. In addition, the new infrastructure development will create opportunity for innovation, which will yield commercial spin-offs.[41]

The preceding literature suggests that South Africa can improve its competitive position as well as the World Economic Forum's ranking regarding the quality of its scientific research institutions, and its capacity for innovation,[39] by adopting a synchronised plan, which includes open and inclusive institutions. The plan should promote collaboration among individuals, and incentivise innovation so that a KBE may flourish, and competitive advantage is achieved. The Department of Science and Technology – the organ of state responsible for promoting scientific research in South Africa – has identified SKA SA as its flagship project.[38,44] This paper considers the requirements for creating a KBE, and focuses on the relevance of SKA SA as a knowledge engine and its contribution to economic growth.

Theoretical model

Four main themes crucial for creating and sustaining a KBE emerged from the literature review: institutions, interrelationships, innovation

and individuals. These themes may be considered as the pillars crucial for a knowledge economy. The interplay of these pillars is referred to here as the 4I model. Notably, these pillars do not exist in isolation. To successfully promote any one, depends on the strength of the remaining pillars. Each of these pillars must be strengthened to strengthen a knowledge economy. The relationships among the pillars may be likened to that of gears. The speed, size, and strength of any of the pillars will influence the remaining gears, as illustrated in Figure 1. There is a complex interaction and inter-dependence among the four pillars. *Individuals* exist, operate and influence each of the other three themes. *Institutions* must be open and inclusive to attract and retain individuals who innovate. Open and inclusive institutions foster an environment in which ideas may be transformed, by individuals, into *innovations* for commercialisation. Institutions establish the platforms on which triple helix *interrelationships* may flourish. Arguably, the greater the number of symbiotic interrelationships, the greater the prospect of sharing, and the greater the prospect of success. The 4I model is the basis of this research study.



© Sandhya Singh

Figure 1: The gear-like nature of the 4I model.

Methodology

A qualitative research approach was adopted because it supports constructing a detailed description of a participant's social construct or perceptions. This study is inductive in nature and employed interpretivist epistemology. A cross-sectional research design, rather than a longitudinal design, was employed because the SKA project is in its infancy and it will take many years before its impact as a knowledge engine in the South African economy can be measured more meaningfully. Although this study is cross-sectional in nature, it is replicable. A future study might provide more clarity on SKA SA's impact on the South African economy, as more information may then be available. However, the same themes are expected to emerge. This study was segmented into five steps, which are described briefly below.

Firstly, the literature was reviewed to broadly determine the factors that contribute to economic growth and development. The literature review was then focused on S&T, particularly on astronomy, to determine the role of S&T and astronomy on economic growth. The literature review was interrogated to identify themes, which formed the basis for the research questions devised. Secondly, data were gathered through a process of semi-structured interviews. The interview questions were formulated to investigate the themes that emerged from the literature review, as encapsulated in the 4I model – that is, the role of institutions, interrelationships, innovation and individuals in creating a KBE. A question was included to investigate impediments and enablers for SKA SA's success that were not already proposed by the respondent. This additional question was important because it investigated gaps in the literature review, and could have highlighted additional findings. Thirdly, the findings were formulated, drawing on themes that emerged from the interviews. Fourthly, recommendations are made on how the 4Is should be strengthened to augment SKA SA's impact in promoting a knowledge economy in South Africa, and, lastly, the conclusions drawn from this study are discussed.

Primary data were collected over a 7-week period through interviews. Interviews were conducted via Skype™ as the most convenient method to access participants in the available time frame. One-on-one interviews were chosen over focus groups, for greater control, and allowed respondents greater freedom to express themselves.[45] In addition, one-on-one interviews were more practical to schedule and allowed for a more in-depth investigation. The interviews were recorded to preserve the data for subsequent review.[45] Homogenous and purposeful sampling was adopted, and senior officials associated with SKA SA were selected as respondents for this study, as they have the specialist expertise and knowledge regarding SKA SA. In collaboration with the business manager of SKA SA, 11 suitable candidates to participate in this study were identified. Interviews were audio recorded, and supported by field notes[46,47], to determine themes that emerged.

Ethical clearance for the study was obtained from the Graduate School of Business, University of Cape Town (ref. GSB/MBA/2016/101). The respondents were informed that their participation was voluntary and that they could withdraw at any time. The respondents were also informed that their identities would be protected; all references to the respondents ensure anonymity.

The results of the study may be limited by participant bias and the small sample size. Participants may have shown a bias to support SKA SA, because the project's success is key to their continued employment. In addition, the sample size was by necessity small, as the number of senior officials associated with SKA SA is small.

Results and discussion

The results confirm that the SKA SA project contributes to promoting South Africa's knowledge economy. The study verified that the four main themes identified in the literature review (institutions, individuals, interrelationships and innovations), as encapsulated in the 4I model, are crucial for creating a knowledge economy, as demonstrated in the context of SKA SA. No additional themes were identified. In addition, four new sub-themes emerged that were not identified in the initial literature review. However, each of these new sub-themes is directly linked to one of the main themes. The sub-themes relate to the role of inherent competitive advantage in formulating a nation's strategic policies, the introduction of the concept of multidimensional interrelationships, the significance of leadership being politically astute, and the significance of developing and commercialising value chain products. The main findings relevant to each pillar, including the new sub-themes, are discussed in further detail below.

Institutions

This study has revealed that stable and consistent policies and funding strengthen the institutions pillar, especially for long-term projects.[20,48,49] Astronomy and SKA SA, a long-term project, have both benefitted from stable and consistent policies and funding.[50-53] The respondents noted that while there is significant engagement with other government departments to maximise support for the telescope, and to maximise the economic benefit that the telescope may yield, there remains scope for improvement. Six respondents noted that SKA SA is adequately funded. State investment in fundamental research, which is extremely risky, reduces the overall market risk.[26,27] As a business unit of the National Research Foundation, SKA SA is subject to standardised processes, some of which are ill-suited for a development project such as SKA SA. For example, the onerous procurement requirements of the *Public Finance and Management Act* promote governance, but occasionally slow the rapid progress of the SKA SA. The respondents expressed the view that there is a national focus on quantity rather than quality: for example, there is a drive to produce an increasing number of students and publish an increasing number of papers. Respondent 6 conjectured that South African scientists will be ill-equipped to lead the big science endeavours in years to come, if quality continued to be ignored.

Although SKA SA enjoys widespread support from government, certain regulations, structures and policies slow SKA SA's progress.

A new sub-theme was revealed – namely that a nation's competitive and innovation strategy may be informed by its inherent competitive advantage. Inherent competitive advantage may be leveraged to strengthen the key pillar of institutions, thereby fostering a knowledge economy. South Africa's geographical advantage was considered when determining fields in which the country's fundamental research endeavours should be promoted.[54] Importantly, the geo-location advantage is latent unless it is coupled with sound institutions, which is an argument overlooked by Landes[55], but appreciated by Acemoglu and Robinson[16]. Moreover, government efforts must focus on a subset of industries.[1,17,56] Arguably, the same is being done in South Africa, where focus is directed at areas with inherent competitive advantage, such as radio astronomy.

Interrelationships

All respondents noted SKA SA's interrelationships with universities, government and local businesses. The interview data revealed that SKA SA has impacted and maintained interrelationships with universities. Seven respondents noted that SKA SA enjoys widespread support from government, and that SKA SA engages with external companies to pursue research and development opportunities. The respondents further noted that developing these local industry partners during the design phase would strengthen the likelihood of these industry partners being awarded tenders for international contracts during the SKA construction phase. These data demonstrate that SKA SA's interrelationships are consistent with the triple helix model as described by Etzkowitz and Leydesdorff[15] and Etzkowitz[14].

However, the astronomy landscape is evolving, which heightens the need for collaboration and knowledge sharing. Respondent 10 noted that the scale and complexity of modern projects transforms the sociology of astronomy. This respondent further noted that astronomy today calls for multidisciplinary, as well as cross-sector collaboration, which is founded on knowledge sharing. Evidently, knowledge sharing underpins the success of astronomical projects, such as SKA SA. Moreover, knowledge sharing is underpinned by interrelationships, that is, both SKA SA's relationships with other partner countries as well as cross-sector collaboration. Six respondents mentioned that the astronomy domain has had a culture of knowledge sharing. Respondent 11 contrasted astronomy with the military, noting that confidentiality is imperative in the military. However, the respondent highlighted that the collaborative culture that exists during the design and construction phase will not apply to fundamental scientific discoveries. The respondent noted that knowledge gained from fundamental research studies will be guarded, and shared through academic publications.

The nature and complexity of modern astronomy demands cross-sector, multinational, and multidisciplinary collaborations, hereafter referred to as multidimensional interrelationships, which is a new concept introduced here. Respondents provided examples that highlighted the multidimensional nature of interrelationships at SKA SA. Eight respondents mentioned that SKA SA works in conjunction with partner countries in multinational teams and that SKA SA has representation on SKA International's board of directors. All respondents noted that SKA SA had initiatives across the full spectrum of knowledge generators (knowledge generators refers to high schools, universities and further education and training colleges) and that project initiatives were not limited to universities. In addition, five respondents noted that collaboration was not limited to local businesses, but included collaboration with multinational enterprises, such as IBM.[57] Based on these responses, it may be inferred that triple helix interrelationships are less comprehensive than multidimensional interrelationships. The former is limited to cross-sector interrelationships, while the latter includes cross-sector, multinational and multidisciplinary interrelationships, and should be promoted for their more comprehensive role in developing the knowledge economy. Additionally, multidimensional relationships incorporate the full spectrum of knowledge generators. This study confirmed the conjecture made by Kose and Ozturk[58] that technological advances spur interrelationships and inter-dependencies.

Innovation

This study revealed that SKA SA is fuelling innovation through its big data and commercialisation initiatives. Innovation presents the opportunity for developing economies to leapfrog, and to close the divide between developed and developing economies.[1,18,19] Respondents 3 and 10 described big data[59,60] management as the greatest challenge that SKA SA is expected to face. Respondent 10 explained that the challenge will increase as more antennae are added to the telescope (refer to Supplementary figure 2 for SKA SA's anticipated data volume)[43]; however, the Inter-University Centre for Data Intensive Astronomy has been established in response to the big data challenge presented by MeerKAT, which is expected to be exacerbated by the SKA telescope. The Inter-University Centre for Data Intensive Astronomy recently became the first African institute to launch a cloud data centre, that is the African Research Cloud.[61]

SKA SA collaborates with local industry partners to increase the likelihood of local business winning tenders during the construction phase, developing local capacity in once declining technical areas, and to exploit commercialisation opportunities.[62] A strengthened innovation pillar increases SKA SA's chances of being awarded the post-construction operations contract. The respondents noted that the ongoing support and maintenance of the telescope's post-construction operations, with an expected duration of 50 years, presents a significant opportunity to contribute to the South African economy. In addition, all 11 respondents noted that SKA SA has developed several products suitable for commercialisation. Interestingly, several products developed may be used in various permutations with other products that are also developed by SKA SA, thus creating product value chains, which further strengthens the innovation pillar. Therefore, product value chains are considered a new sub-theme for fostering a knowledge economy, as they were not identified in the literature review. Importantly, creating such value chains promotes market sustainability of the products developed. However, Respondents 1 and 3 noted that SKA SA lacked a coherent policy for commercialising innovation, and Respondent 9 noted that there was no dedicated budget available for promoting commercialisation. A coherent plan is crucial to ensure that resources are optimised.[31] Notably, the respondents also mentioned that a commercialisation manager had recently been appointed by SKA SA, and was likely to address these gaps. Armed with fruitful symbiotic collaborations with industry partners, SKA SA appears poised to enable the transformation from innovation to commercialisation.

Individuals

This study revealed that upgrading human capital and attracting and retaining suitable individuals are essential sub-themes for strengthening the 'individuals' pillar. Suitable individuals refers to individuals representative of the diversity of South Africa who have the desired skill set. Diversity includes gender, age and racial considerations. In addition, a new sub-theme was revealed – that is, the significance of politically astute leadership on publicly funded projects, such as SKA SA. The responsibility for upgrading human capital in the astronomy pipeline, and attracting and retaining suitable individuals, rests mainly with the Human Capital Development (HCD) Programme and is discussed below. Notably, upgrading the skills of SKA SA's staff is the responsibility of the project's human resources function, while upgrading the skills of local industry partners is the responsibility of the design consortia.

The HCD initiatives are primarily directed at the youth, considered by Ross[1] to be more adept in using technology proficiently. Respondents 3 and 7 explained that the initiatives target the full spectrum of knowledge generators (universities, further education and training colleges, and schools) and includes a young professional development programme. Respondent 7 explained that by adopting a responsive and evolutionary approach, the programme has flourished and has become increasingly structured and focused. Ten respondents noted the impact that the HCD programme continues to have in universities in South Africa through

funding bursaries, grants and research chairs for the study of STEM subjects (refer to Supplementary figures 3 and 4 for the number of bursaries, fellowships and grants awarded). Five respondents mentioned that students created by the astronomy pipeline who are not absorbed by the astronomy community, will be absorbed by other sectors of the economy. SKA SA's initiatives are aligned with recommendations in the literature, in that universities contribute by training astronomers and attracting good students to study STEM subjects.[31,35] Astronomy-trained graduates assume senior roles in industry, thus further boosting the knowledge economy.[37] The upgrading of human capital is an essential requirement of a knowledge economy.[8] In a knowledge economy, highly skilled individuals are trained for highly skilled jobs.[9]

The respondents identified key enablers and inhibitors for attracting and retaining suitable individuals. All 11 respondents noted that the interesting and exciting nature of the SKA project has contributed significantly to attracting and retaining suitably skilled individuals within the project. Other factors credited for attracting individuals to SKA SA were the vision underpinning the organisation's goals, the professional management of the organisation, good leadership, healthy team dynamics, and working among high-quality peers. However, respondents also noted several challenges to attracting and retaining suitable individuals, which include the lengthy duration to obtain work permits for foreign nationals, and keen competition for the best skilled individuals who have scarce skills. Therefore, SKA SA places emphasis on the HCD programme to develop the skills of targeted groups of individuals. A significant investment in world-class infrastructure and innovation attracts and retains excellent researchers[37], and the visa processes should facilitate attracting individuals with desired skills, so that interrelationships and innovation can be promoted[20]. There was no consensus regarding the influence of remuneration in attracting and retaining individuals; three respondents considered SKA SA's remuneration to be market-related, while another three considered that it was not market-related.

A new finding that emerged from the primary data collected is that the political astuteness of leaders is crucial for ensuring sustained support for publicly funded projects. The political astuteness of leaders strengthens the key pillar of individuals. Therefore, the role of political astuteness of leaders is considered a sub-theme for fostering a knowledge economy. Four respondents noted the importance of SKA SA's leaderships' political astuteness in obtaining and retaining government support and funding. The respondents mentioned that the SKA SA leadership had strong professional relationships with government leaders, but also understood that delivering on commitments was essential for ongoing support.

Additional pillars and the 4I model

To test the robustness of the 4I model, respondents were asked if there were any factors, other than the four themes under investigation, which either impede or enable SKA SA's success. Six respondents noted that they were not aware of any additional factors or pillars, other than the four themes being investigated. Two respondents repeated their earlier contention that bureaucracy impeded SKA SA's progress. The importance of stable funding, the challenge of big data, and the challenge of South Africa having a limited pool of science graduates available for recruitment, are factors that were raised by each of three respondents. All the additional factors suggested by the respondents may be mapped into one of the four main themes that were already identified. Bureaucracy and stable funding are associated with institutions. The limited pool of science graduates is associated with individuals, and the big data challenge is associated with innovation. Therefore, this study did not reveal any additional pillars that enabled or inhibited SKA SA's contribution to a knowledge economy, thereby providing evidence for the robustness of the 4I model.

Recommendations and conclusion

Institutions, interrelationships, innovation and individuals were identified, through a literature review, as the four main pillars for a knowledge economy and investigated further. The findings of this study corroborate the validity of the proposed 4I model and no new pillars relevant to SKA SA were identified. Furthermore, this study revealed that the four pillars of the 4I model have indeed supported SKA SA's contribution to the knowledge economy. Although additional pillars were not identified, new sub-themes were identified from the interview data, which were not identified in the literature review. These sub-themes strengthened the 4I model, as each of the sub-themes is associated with one of the main themes. Furthermore, this study highlights that by strengthening each of the pillars through interventions identified below, SKA SA's impact on the knowledge economy will be further boosted. For each of the pillars, the main themes, sub-themes, limitations and recommendations, in relation to SKA SA's contribution to a knowledge economy, are summarised below.

It was found that open and inclusive institutions – one of the four main themes – are critical for engendering a knowledge economy. In the case of SKA SA, open and inclusive institutions were represented by protective legislation, and stable and consistent policies and funding. These institutions provide the foundation for SKA SA to flourish. Furthermore, SKA SA complies with institutional requirements, such as employment equity and broad-based black economic empowerment, that promote broad participation of South Africa's citizenry. However, this study also found limitations of institutions that hinder SKA SA from promoting a knowledge economy. Stringent and ill-suited processes, policies and key performance indicators can impede the project's progress. It was recommended that the structures, policies, procedures and key performance indicators for development projects such as SKA SA should be reformulated to optimise the impact of such projects. This study revealed a new sub-theme relevant to institutions: South Africa's inherent competitive advantage, that is, its geo-location was considered when determining the strategic direction of fundamental research that the country should pursue.

This study highlights that the nature and sociology of modern astronomy demands multidimensional interrelationships, rendering open institutions crucial for SKA SA's success. Multidimensional interrelationships are more complex than triple helix relationships. Knowledge is shared through multidimensional collaborations to find innovative solutions to complex problems. SKA SA's collaborations extend to SKA partner countries, government departments, industry partners and knowledge generators. Although not optimised, the project enjoys widespread support from all tiers of government. Furthermore, SKA SA's collaboration with industry ranges from small, medium and micro-sized enterprises to multinationals. Through collaboration with industry partners, SKA SA not only upgrades local human capital, but also places itself at the cutting-edge of technological breakthroughs. It was recommended that SKA SA involve additional industry partners, who already have a need for big data technology, during the early stages of product innovation so that commercialisation opportunities may be fast-tracked. This study also revealed another sub-theme – that is, product innovations may also be inter-related, thus creating product value chains. Product value chains improve the likelihood of market sustainability of products in value chains. The project successfully creates and leverages multidimensional interrelationships, which favourably impact South Africa's endeavours for fostering a knowledge economy.

It was found that the multidimensional collaborations often yield innovations, which may have commercial impact. Innovative products that are commercialised are also critical for a thriving knowledge economy. SKA SA has established and participates in numerous multidimensional collaborations, which promote knowledge sharing and joint problem solving. Local industry partners have been upskilled and are poised to commercialise product innovations resulting from the collaboration. However, SKA SA's commercialisation strategy remains outstanding, and is crucial for maximising economic benefit. It was recommended that SKA SA implement a commercialisation strategy. Overall, SKA SA has been successful with innovation, and once armed with a coherent commercialisation strategy, the project may have a greater impact.

It was also found that prestigious science projects, such as SKA SA, not only attract and retain suitably skilled individuals, but that these individuals are often amongst the best in their field globally. Furthermore, the professional management of the organisation, good

leadership, and healthy team dynamics contribute to the project's low attrition rate. Additionally, the HCD programme has fostered the astronomy pipeline, growing the astronomy community in South Africa. Students upskilled through the astronomy pipeline initiatives and not absorbed by the astronomy community could be absorbed by other sectors in the economy, thus further boosting the knowledge economy. Although diversity challenges persist, SKA SA has made progress with transformation. It was recommended that SKA SA continue to focus on transformation. Importantly, this study revealed the significance of politically astute leaders. These leaders are instrumental in ensuring ongoing support for the publicly funded project.

Overall, the results of this study show that SKA SA is contributing to growing the South African knowledge economy. The project is in its infancy and its full impact is not yet known. Future studies may include longitudinal studies of SKA SA, testing the validity of the 4I model on other fundamental research initiatives, and formulating the underlying economic theory for knowledge as a resource. Longitudinal studies may involve tracking the progress of SKA SA's bursars. Such studies may investigate the industries in which the bursars are employed, or the percentage of bursars retained in the South African economy. Another key subject to investigate is the extent and impact of fundamental research conducted on data from the MeerKAT and SKA telescopes. Here it would be important to determine how many of the big science research initiatives are being led by South African based scientists, which would provide a measure of the success of the HCD programme.

Acknowledgements

I acknowledge Kavilan Moodley and Thomas Koelble for their guidance, Mary Lister for bibliographic support, and Bernard Fanaroff for critical feedback on earlier drafts. I acknowledge Russ Taylor and Kim de Boer for providing figures included in the supplementary material. In addition, my sincere appreciation goes to the SKA SA Director, Rob Adam, the SKA SA business manager, Carla Sharpe, and members of the SKA SA team who participated in this study.

References

1. Ross A. The industries of the future. London: Simon & Schuster; 2016.
2. Statistics SA. Economic growth [document on the Internet]. c2016 [cited 2016 Apr 21]. Available from: http://www.statssa.gov.za/?page_id=735&id=1&paged=1
3. Statistics SA. The economy: Winners and losers of 2015 [document on the Internet]. c2016 [cited 2016 Apr 21]. Available from: http://www.statssa.gov.za/?p=6233
4. Bleak outlook for SA in 2016 and 2017. Business Tech. 2016 Jan 17 [cited 2016 Mar 21]. Available from: http://businesstech.co.za/news/business/109145/bleak-outlook-for-sa-in-2016-and-2017-report/
5. Department of Science and Technology (DST). Innovation towards a knowledge-based economy: Ten-year plan for South Africa (2008–2018) Pretoria: DST; 2008. Available from: http://unpan1.un.org/intradoc/groups/public/documents/CPSI/UNPAN027810.pdf
6. University of Pretoria. The SKA is helping to bring Africa into the knowledge economy. Innovate. 2015;10:57–58. Available from: http://www.up.ac.za/media/shared/404/Articles/innovate_10_2015_the-ska-is-helping-to-bring-africa-into-the-knowledge-economy.zp73375.pdf
7. Shie VH, Meer CD, Shin N. Locating China in the twenty-first-century knowledge-based economy. J Contemp China. 2012;21(73):37–41. https://doi.org/10.1080/10670564.2012.627669
8. Organisation for Economic Co-operation and Development (OECD). The knowledge-based economy. Paris: OECD; 1996. Available from: http://www.oecd.org/dataoecd/51/8/1913021.pdf
9. De La Paz-Marín M, Gutiérrez-Peña PA, Hervás-Martínez C. Classification of countries' progress toward a knowledge economy based on machine learning classification techniques. Expert Syst Appl. 2014;42:562–572. https://doi.org/10.1016/j.eswa.2014.08.008
10. Fowler SW, Lawrence TB, Morse EA. Virtually embedded ties. J Manage. 2004;30(5):647–666. https://doi.org/10.1016/j.jm.2004.02.005
11. Johannessen JA, Olsen B. The future of value creation and innovations: Aspects of a theory of value creation and innovation in a global knowledge economy. Int J Inf Manage. 2010;30:502–511. https://doi.org/10.1016/j.ijinfomgt.2010.03.007
12. Ondari-Okemwa E. The strategic importance of identifying knowledge-based and intangible assets for generating value, competitiveness and innovation in sub-Saharan Africa. S Afr J Libr Inf Sci. 2011;77(1):138–154. https://doi.org/10.7553/77-2-56
13. Kaplan RS, Norton DP. Measuring the strategic readiness of intangible assets. Harv Bus Rev. 2004;82(2):52–63.
14. Etzkowitz H. MIT and the rise of entrepreneurial science. London: Routledge; 2002. https://doi.org/10.4324/9780203216675
15. Etzkowitz H, Leydesdorff L. Universities in the global knowledge economy: A triple helix of academic-industry-government relations. London: Cassell; 1997.
16. Acemoglu D, Robinson JA. Why nations fail: The origins of power, prosperity and poverty. New York: Crown Business; 2012.
17. Johnson C. MITI and the Japanese miracle: The growth of industrial policy: 1925-1975. Stanford, CA: Stanford University Press; 1982.
18. Friedman TL. The world is flat: The globalized world in the twenty-first century. London: Penguin; 2006.
19. Yunus M, Weber K. Creating a world without poverty: Social business and the future of capitalism. New York: Public Affairs; 2007.
20. NAS, NAE, IMNA. Rising above the gathering storm: Energizing and employing America for a brighter economic future. Washington, DC: National Academies Press; 2007. Available from: http://www.nap.edu/catalog/11463.html
21. Saad CS, Damian RI, Moons WG, Benet-Martınez V, Robins RW. Multiculturalism and creativity: Effects of cultural context, bicultural identity, and ideational fluency. Soc Psychol Personal Sci. 2012;4(3):369–375. https://doi.org/10.1177/1948550612456560
22. McCuiston VE, Wooldridge BR, Pierce CK. Leading the diverse workforce: Profit, prospects and progress. Leadersh Organ Dev J. 2004;25(1):73–92. https://doi.org/10.1108/01437730410512787
23. The World Bank. GINI index (World Bank estimate) [document on the Internet]. No date [cited 2016 Apr 11]. Available from: http://data.worldbank.org/indicator/SI.POV.GIN
24. Fin24. Most top managers still whites. Fin24. 2014 April 10. Available from: http://www.fin24.com/Economy/Whites-still-top-management-posts-report-20140409-2
25. Wall M. China finishes building world's largest radio telescope. Space.com 2016 July 6. Available from: http://www.space.com/33357-china-largest-radio-telescope-alien-life.html
26. Mazzucato M. The entrepreneurial state: Debunking public vs. private sector myths. London: Anthem Press; 2015.
27. Janeway WH. Doing capitalism in the innovation economy: Markets, speculation and the state. Cambridge: Cambridge University Press; 2012. https://doi.org/10.1017/CBO9781139381550
28. Carlucci D. Grasping knowledge-based value creation dynamics in 21st century organizations. Meas Bus Excell. 2014;18(1). https://doi.org/10.1108/MBE-11-2013-0056
29. Dzisah J. Institutional transformations in the regime of knowledge production: The university as a catalyst for the science-based knowledge economy. Asian J Soc Sci. 2007;35:126–140. https://doi.org/10.1163/156853107X170196
30. Wallace K. America's brain drain crisis: Why our best scientists are disappearing, and what's really at stake? Reader's Digest. 2005 December 8. Available from: http://americanbraindrain.blogspot.co.za/
31. Astronet. The Astronet infrastructure roadmap: A strategic plan for European astronomy. Paris: Astronet; 2008. Available from: http://www.astronet-eu.org/IMG/pdf/Astronet-Brochure_light.pdf
32. Hoskin M, editor. The Cambridge concise history of astronomy. Cambridge, UK: Cambridge University Press; 1999.
33. Menzies G. 1421: The year China discovered the world. London: Random House; 2003.

34. NASA. Sputnik and the dawn of the space age [homepage on the Internet]. c2007 [cited 2016 Jun 17]. Available from: http://history.nasa.gov/sputnik/

35. CDSAA, NRC. New worlds, new horizons in astronomy and astrophysics. Washington DC: The National Academies Press; 2010. Available from: http://www.nap.edu/catalog.php?record_id=12951

36. CONICYT. Astronomy, technology and industry: Roadmap for fostering technology development and innovation in the field of astronomy in Chile [Internet]. Santiago: Government of Chile; 2012. Available from: http://www.conicyt.cl/astronomia/files/2013/11/Roadmap_Astronomia_v3.pdf

37. NCAAS. New horizons: A decadal plan for Australian astronomy 2006–2015. Canberra: Australian Academy of Science; 2005. Available from: https://www.science.org.au/support/analysis/reports/decadal-plan-australian-astronomy-2006-2015

38. Wild SE. Belt-tightening hits SA'S SKA budget. Wild on Science. 2016 February 24. Available from: http://wildonscience.com/2016/02/belt-tightening-hits-sas-ska-budget/

39. The global competitiveness report 2014–2015. Geneva: World Economic Forum; 2015. Available from: http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf

40. 18 Infrastructure projects for SA. Moneyweb. 2012 October 19. Available from: https://www.moneyweb.co.za/archive/18-infrastrucutre-projects-for-sa/

41. SKA. Everything you wanted to know about the SKA [homepage on the Internet]. c2015 [cited 2016 Apr 04]. Available from: http://www.ska.ac.za/qa/

42. SKA. The SKA project [homepage on the Internet]. c2016 [cited 2016 Apr 04]. Available from: http://www.ska.ac.za/about/project.php

43. SKA. Frequently asked questions about the SKA [homepage on the Internet]. No date [cited 2016 Apr 04]. Available from: https://www.skatelescope.org/frequently-asked-questions/

44. SKA SA. SKA News. 2015 February. Available from: http://www.ska.ac.za/wp-content/uploads/2016/11/24_ska_newsletter_feb2015.pdf

45. O'Leary Z. Researching real world problems: A guide to method of enquiry. London: SAGE; 2005.

46. Halcomb EJ, Davidson PM. Is verbatim transcription of interview data always necessary? Appl Nurs Res. 2006;19:38–42. https://doi.org/10.1016/j.apnr.2005.06.001

47. Tessier S. From field notes, to transcripts, to tape recordings: Evolution or combination? Int J Qual Methods. 2012;11(4):446–460. https://doi.org/10.1177/160940691201100410

48. Neo BS, Chen G. Dynamic governance: Embedding culture, capabilities and change in Singapore. Capabilities and change in Singapore. Singapore: World Scientific Publishing; 2007. https://doi.org/10.1142/6458

49. Leong HK, Ho KL. Shared responsibilities, unshared power: The politics of policy-making in Singapore. Singapore: Marshall Cavendish Academic; 2003.

50. Astronomy Geographic Advantage Act, no. 21 of 2007. Government Gazette 516(31157), 17 June. Government notice no. 666.; 2007. Available from: http://www.gov.za/sites/www.gov.za/files/gg31157_nn666a_pg1-30.pdf

51. Intellectual Property Rights from Publicly Financed Research and Development Act, no. 51 of 2008. Government Gazette 522(31745), 22 December. Government notice no. 1402.; 2008. Available from: http://research.ukzn.ac.za/Files/IPR_from_PFR_D_Act_No_51_of_2008_1.sflb.pdf

52. Department of Science and Technology (DST). Corporate strategy 2009/10. Pretoria: DST; 2009. Available from: http://www.gov.za/sites/www.gov.za/files/DST-CorporateStrategy2009-10_23062009.pdf

53. Department of Arts, Culture, Science and Technology (DACST). White paper on science and technology: Preparing for the 21st century. Pretoria: DACST; 1996. Available from: http://www.gov.za/sites/www.gov.za/files/Science_Technology_White_Paper.pdf

54. South Africa. South Africa's national research and development strategy. Pretoria: The Government of South Africa; 2002. Available from: http://www.gov.za/sites/www.gov.za/files/rd_strat_0.pdf

55. Landes DS. The wealth and poverty of nations: Why some countries are so rich and some so poor. New York: W.W. Norton & Company; 1998.

56. Tan CY. Organisational legitimacy of the Singapore Ministry of Education. Oxford Rev Educ. 2013;39(5):590–608. https://doi.org/10.1080/03054985.2013.830098

57. SKA SA. SKA News. 2016/2017. Available from: http://www.ska.ac.za/wp-content/uploads/2017/03/ska_newsletter_2017_01.pdf

58. Kose MA, Ozturk EO. A world of change. Finance & Development. 2014;51(3):6–11. Available from: http://www.imf.org/external/pubs/ft/fandd/2014/09/kose.htm

59. Laney D. 3D Data management: Controlling data volume, velocity and variety [document on the Internet]. c2001 [cited 2016 Jul 01]. Available from: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

60. Rouse M. 3Vs: Volume, variety, and velocity [homepage on the Internet]. c2013 [cited 2016 Nov 10]. Available from: http://whatis.techtarget.com/definition/3Vs

61. UCT. Researchers aim for the clouds: UCT and NWU launch first stage of big-data African Research Cloud [homepage on the Internet]. c2016 [cited year month day]. Available from: https://www.news.uct.ac.za/article/-2016-11-04-researchers-aim-for-the-clouds-uct-and-nwu-launch-first-stage-of-big-data-african-research-cloud

62. Wild SE. SKA gives high-tech firms a boost. Mail & Guardian. 2015 Dec 04. Available from: http://mg.co.za/article/2015-12-03-ska-gives-high-tech-firms-a-boost

Evaluating 'homegrown' research networks in Africa

AUTHORS:
Camilla Adelle[1]
Nico Elema[2]
Ereck Chakauya[3]
David Benson[4]

AFFILIATIONS:
[1]Centre for the Study of Governance Innovation, University of Pretoria, Pretoria, South Africa
[2]AU/NEPAD SANWATCE, Stellenbosch University, Stellenbosch, South Africa
[3]AU/NEPAD SANBio, Council for Scientific and Industrial Research, Pretoria, South Africa
[4]Department of Politics and the Environment and Sustainability Institute, University of Exeter, Penryn, United Kingdom

CORRESPONDENCE TO:
Camilla Adelle

EMAIL:
camilla.adelle@up.ac.za

DATES:
Received: 07 Mar. 2017
Revised: 24 Oct. 2017
Accepted: 07 Dec. 2017
Published: 27 Mar. 2018

KEYWORDS:
research networks; policy formulation; science–policy interface; evidence-based policymaking; research collaboration

HOW TO CITE:
Adelle C, Elema N, Chakauya E, Benson D. Evaluating 'homegrown' research networks in Africa. S Afr J Sci. 2018;114(3/4), Art. #2017-0070, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170070

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
British Academy (Newton Mobility Grant no. NG150251); Leverhulme Trust International Networks Scheme (project number IN-2014-014)

Attempts to improve the policy environment have led to a growing pressure on governments in Africa to embark on policymaking that is more evidence based and considers a wide spectrum of scientific and indigenous knowledge. Local – or 'homegrown' – research networks on the continent can help strengthen the role of scientific knowledge in policymaking by increasing the capacity of researchers and by enhancing the visibility and communication of the research produced. While a large number of regional and sub-regional research networks have sprung up in Africa, the mere existence of networks does not guarantee their success. In reality, the impact of research networks on the science–policy interface depends on how well the networks operate in practice. We present a framework for evaluating the effectiveness of research networks in a way that is comparable across networks. The evaluation framework was used to evaluate two sub-regional research networks: the NEPAD Southern African Networks of Water Centres of Excellence (SANWACTE) and the NEPAD Southern African Network for Biosciences (SANBio). The evaluation revealed some shared constraints limiting the effectiveness of both networks, including uneven regional representation, asymmetry between network members, and difficulties in securing sufficient, diverse and sustainable resources. Further research into network design and funding models is suggested in order to enhance the role of these networks in providing locally appropriate knowledge for policymaking on the continent.

Significance:

- While a large number of research networks have sprung up in Africa, the mere existence of networks does not guarantee success.
- Uneven regional representation, power asymmetries, and limited funding constrain the effectiveness of research networks.

Introduction

Decision-makers in Africa are facing increasingly complex development challenges. Solving these challenges requires innovative solutions based on sound research and evidence.[1] While it has been recognised that homegrown policy solutions have a higher probability of success than those developed off the continent, the lack of local research capacity has been identified as a major stumbling block to locally designed policy measures.[2] Even when capacity is available, research findings are not always transferred to policymakers who may instead find themselves under pressure to use information and policy options promoted by international consultants and donors.[3]

A number of research networks are being established on the continent that could help strengthen local research capacity and offer an alternative source of evidence for policymaking in Africa. These research networks link researchers and research institutions working on similar issues from across the continent and/or at a sub-continental level. They aim to enhance collaboration and synergies between existing research programmes in order to reach a critical mass of highly trained and qualified researchers across the continent. Such a critical mass can increase the scientific quality (and hence legitimacy) of the research as well as its visibility to policymakers.

However, the mere existence of research networks does not guarantee such positive outcomes. The impact of networks depends on how (and how well) these networks function in practice. Yet there is an acute lack of literature evaluating the effectiveness of these networks. This deficit is partly because of difficulties in linking specific research to policy decisions, which invariably have multiple influences. There is also a lack of evaluation frameworks with which to analyse and compare networks.

We attempt to address this gap by presenting an evaluation framework for assessing the effectiveness of research networks and using it to evaluate two New Partnership for Africa's Development (NEPAD) research networks: the NEPAD Southern African Networks of Water Centres of Excellence (SANWACTE) and the NEPAD Southern African Network for Biosciences (SANBio). Both research networks respond to the African Union's 2005 Science and Technology Consolidated Plan of Action, which forms the basis for implementing NEPAD's Science and Technology programmes.

The 'policy paradox' in Africa and the role of research networks

Policy inappropriateness in Africa

There is an urgent need to improve what is broadly termed as 'the policy environment' on the continent so that policies are more effective from a managerial and delivery point of view.[4] As a result, there is growing pressure on governments in Africa to embark on policymaking that is more evidence based and considers a wide spectrum of scientific and indigenous knowledge.[1] The ultimate objective is to overcome perceived shortcomings in the policy formulation and implementation with a view to making policies more effective. The logic behind this thinking is that Africa's economic crisis was, and still is, partly the product of accumulated policy distortions built up by inappropriate policies since independence in the 1960s.[5] Inappropriate policies hamper economic growth, discourage private initiative, squander natural capital and can even cause widespread famine and civil wars.[4]

The temptation for African governments when formulating policies can be to rely on 'best practice' models developed outside the continent. In many cases these models are 'pushed' by international donors in a type of 'ideas aid'.[2] In recent decades, development assistance has shifted away from investment financing to supporting policy reform[6] largely because of the growing recognition that a precondition of economic growth is a good policy environment.[7] However, the critical issue is not the recognition and copying of available knowledge but rather the ability to integrate this knowledge with local specific knowledge. This integration sets limits on the application of internationally available knowledge and determines its usefulness or lack thereof.[8] Many development projects have been thought to fail because of a lack of adequate understanding of the local conditions on the continent – something African research institutions should be able to provide.[8]

The Structural Adjustment Programmes of the 1980s and 1990s, for example, were an ill-fated attempt by the Bretton Woods Institutions (World Bank and the International Monetary Fund) to improve policymaking in the continent by instituting severe economic reforms.[4] The implementation of Structural Adjustment Programmes received widespread condemnation from both the international community and African intellectuals.[4] A major criticism of the Structural Adjustment Programmes was that they did not bring about the expected positive results because they invoked policies that did not take into account the social, economic and political realities of the countries concerned.[2] During this period (and still today), most countries in sub-Saharan Africa suffered from a common weakness in their institutions that meant there was a lack of capacity to carry out economic reforms.[9] As a consequence, African governments relied heavily on external agencies and foreign consultants for research that was to be the foundation for policy decisions. According to Dollar and Svensson[6], over a third of the Structural Adjustment Programmes failed because they were implemented in a poor policy environment, which was not fully understood by external researchers.

The role of local research networks

The 'policy paradox' in Africa is that at the same time as there is an acute need for local and appropriate research to feed into decision-making, there has been a stark lack of locally driven research conducted on the continent. In addition, what research does exists, is being underutilised in decision-making.[10] Many policymakers and political leaders are not cognisant of the research ongoing in their countries or across the continent when they formulate policies.[2] Research networks could help strengthen the research–policy relationship in Africa in two main ways: firstly, by increasing the capacity of researchers through collaboration and specialisation; and, secondly, by increasing the visibility and communication of the research produced. Furthermore regional and sub-regional research networks are well placed to tackle an increasing number of complex policy problems that do not stop at national borders.

Partly because of this trend, and because of broader patterns of globalisation and technological progress, cooperation in transnational networks has become an increasingly important feature of the interface between research and policy.[8] This trend is also apparent in sub-Saharan Africa, which, while it lags behind much of the world in both the size and number of networks, is establishing some high-profile research networks[8] such as the African Economic Research Consortium and the Council for the Development of Economic and Social Research in Africa. A recent list of African research networks includes over 50 such networks on the continent.[11] Regional research networks aim to strengthen collaboration in particular fields in order to contribute to applied research and quality teaching. The members of these networks are usually higher education institutions and/or research institutes across specific sub-regions, such as southern Africa, but sometimes also across the whole of Africa.

According to Bhorat[12], these networks offer a range of advantages, which include opportunities to engage with peers and institutions outside one's own country and to learn from their experiences; useful cross-pollination of ideas in diverse areas; shared operational, financial and human resource management; and collective dissemination and fund-raising activities. Ideally, a network – if strong and well-functioning – can assist weaker partners in the network in a way that does not draw too heavily on its own internal time and resources.[12] Bhorat cautions that the existence of regional and sub-regional research networks does not necessarily mean that the research policy relationship is strengthened:

> *Research networks confront a number of difficulties. Member institutes may have no common thematic expertise, or have uneven levels and quality of expertise. They may by grouped because of their common region rather than shared research themes, and their work may also have varying degrees of policy relevance. All of these obstacles can make it very difficult to run and sustain an effective network.*[12(p.194)]

So while research networks in principle can help strengthen the research policy relationship, the mere existence of networks does not guarantee this effect. The actual impact of research networks on the science–policy relationship depends on how well the networks operate in practice, i.e. network effectiveness.

Analytical framework and methodology

The literature on networks is vast and stretches across numerous disciplines including health sciences, management and organisational studies, political science, social science and computer science. For the purposes of this paper, we define networks as a set of autonomous organisations that come together in a bid to reach shared goals. Despite the sizable literature on networks, few studies have focused on examining the relationship between networks structure and activities and effectiveness. However, evaluating network effectiveness is critical to understanding if networks are improving the science–policy interface. This feature is important not only to the network members and funders but also to the wider community served by these networks, such as policymakers and the public that will ultimately benefit by the improved problem-solving and policymaking. Criteria for network effectiveness vary, but generally three distinct elements of networks are identified in the literature: structure, process and outcomes. These elements are outlined below based primarily on a framework developed by the Network Impact and Center for Evaluation Innovation[13] with some adjustments inspired by Hill[14] as well as Provan and Milward[15].

An analytical framework for evaluating networks

Network structure

The connections between the network members are the essential glue that binds a network together. Therefore, it is important to know if efforts within the network to weave members' ties to each other are resulting in efficient and effective pathways for collective learning and action. Network structure therefore has three components that can be assessed: (1) network membership (i.e. who are the network members); and (2) membership cohesion (i.e. how connected are the members) and (3) network centrality (i.e. the importance and influence of the partnership/network/collaboration within the power structure and organisational context of its wider community, e.g. in policymaking circles).

Network process

A crucial factor for a network's well-being is its capacity to sustain the active enthusiasm and commitment of voluntary members and enable their ability to work together to achieve shared goals. Network process has three dimensions that can be assessed: (1) resources (i.e. what resources are shared in the network); (2) infrastructure, or the internal systems and structures that support the network (i.e. how is the communication, rules and processes organised and is there a central administrative structure); and (3) added value, or the network's capacity for joint value creation (i.e. are the members working together to achieve shared goals).

Network outcome

Ultimately, most networks have a goal of achieving a particular type of social change. They come together for a purpose, and while network structure and process dimensions are important to their ability to

achieve those results, it is important to know if the network itself is making a difference. Network results have two dimensions that can be assessed: (1) interim outcomes, or the results achieved as the network works toward its ultimate goal or intended impact; and (2) the goal or intended impact itself (e.g. a policy outcome was achieved).

Much of the literature on networks tends to focus on network composition and function rather than on the elements crucial to meeting network goals or indeed ascertaining if and how these goals were achieved. In some ways, the literature implies that establishing a network is a huge success in itself; however, this is not necessarily the case, as our consideration of two NEPAD research networks illustrates.

Research methodology

In our research, the analytical framework set out above was operationalised into a set of specific evaluation questions for each of the three network elements, namely 'network structure', 'process' and 'outcome'. These questions are set out in Table 1 and were used to guide the analysis. The evaluation took place between September 2016 and January 2017. The first stage of the evaluation primarily focused on the qualitative desktop documentary analysis of the many documents charting the history, strategy and achievements of the two networks including annual reports, business plans, brochures, presentations at conferences, webpages, as well as internal and external reviews of the networks. The specific documents used are cited in the text to follow. The desktop documentary analysis was followed by semi-structured discussions with the two network coordinators (N.E. and E.C.). Additional information in the form of direct participant observations of the two networks was provided through these discussions by the authors based on their interactions with the networks over several years.

Two NEPAD-affiliated networks were evaluated: The NEPAD Southern African Network of Water Centres of Excellence (SANWACTE) and the NEPAD Southern African Network for Biosciences (SANBio). Both networks are part of continent-wide networks established in all five African Union regions under the framework of NEPAD Centres of Excellence for Science and Technology. These cases were selected partly because their similar geographical scope and common mandate make them broadly comparable.

Table 1: Evaluation framework for the three network elements (structure, process and outcomes)

Element	Focus	Evaluation question
Structure	Membership	Who participates in the networks and what is their role?
		Has the network assembled members with the capacities needed to meet network goals?
		Who is not included/connected that should be?
	Cohesion	How are the members connected?
		How efficient are these connections?
		How dependent is the network on a small number of actors?
		What is flowing through the network – e.g. information and/or other resources?
		Is the structure adjusted to meet changing network needs and priorities?
	Centrality	To which forums and actors is the network itself connected?
		From where did the mandate for the network come?
Process	Resources	Has the network secured needed material resources?
		How diverse and dependable are these resources?
		How are the members contributing resources to the network?
		Is the network adapting its business plan over time?
	Infrastructure	What infrastructure is in place for network coordination and communications?
		Are these systems efficient and effective?
		What are the network's governance rules and how are they followed (and who developed them)?
		How are the network's internal systems and structures adapting over time?
	Collective value-added	Do all the members share a common purpose for the network?
		Are the members working together to achieve shared goals, including those that evolve over time?
		Are all the members contributing to the networks efforts?
		Are the members achieving more together than they could alone?
Outcomes	Interim outcomes	What signs of progress/outputs are there?
		Do these outputs show a clear progression towards the ultimate goals of the network?
	Ultimate goals	What impacts are expected at a community level (e.g. more input of research into policy)?
		What impacts are expected at an organisational level (e.g. benefits for the members' organisations)?
		If these impacts can be observed, can a plausible case be made that the network contributed to them?

Sources: Based on Network Impact and Center for Evaluation Innovation[13]; Hill[14]; Provan and Milward[15]

The networks were also selected because of the authors' knowledge of these networks, which facilitated a more intimate evaluation than the selection of other similar networks would have. Further information on the mandate and history of the two networks is given below.

Network 1: An evaluation of the Southern African Network of Water Centres of Excellence

Network background

The mandate to establish SANWATCE came from a joint declaration in 2006 from the African Ministerial Council on Science and Technology and the African Ministers' Council on Water to establish Networks of Water Centres of Excellence across the African continent. NEPAD was to act as the facilitating agent. The intention was to establish networks in all five AU regions of the African continent. So far two networks have been set up: SANWATCE covering the Southern African Development Community (SADC); and the NEPAD Western African Network of Water Centres of Excellence (WANWATCE) covering the Economic Community of West African States. SANWATCE's mission is to 'contribute to the improved human and environmental well-being through research and development in water and sanitation'[16] through pursuing the following objectives: facilitate and conduct selected research on water issues; serve as a higher education soundboard to the SADC region on regional water matters; collaborate with other networks and institutions in specialised areas; set the SADC water research agenda; and establish a continental water research agenda which is populated from the SADC regional water agenda.[16]

Network structure

SANWATCE consists of 11 higher education and research institute members (Table 2). The network Secretariat is hosted by Stellenbosch University in South Africa. Members of the network were selected according to a set of transparent criteria posted on the network's website.[17] One of the main challenges has been to draw in network members from across all 15 SADC countries. Despite ongoing efforts to recruit new members, SANWATCE is made up of members from just eight countries with four members in South Africa. Countries currently without a network member are Angola, Democratic Republic of the Congo, Tanzania, Seychelles, Madagascar, Swaziland and Lesotho. The over-representation of South Africa as a relatively well-resourced country with strong tertiary institutions is unsurprising, as is the lack of participation of smaller states such as Lesotho and Swaziland. However, in general, wider coverage of the networks remains a challenge that the network has not yet been able to adequately address.

Table 2: SANWATCE Members

Institution	Country
Stellenbosch University	South Africa
Eduardo Mondlane University	Mozambique
University of KwaZulu-Natal	South Africa
University of the Western Cape	South Africa
University of Malawi	Malawi
University of Zambia	Zambia
University of Botswana	Botswana
CSIR Natural Resources and Environmental Unit	South Africa
Polytechnic of Namibia (recently renamed Namibian University of Science and Technology)	Namibia
National University of Science and Technology, Zimbabwe	Zimbabwe
University of Mauritius	Mauritius

The members are connected through an annual face-to-face Steering Committee Meeting of members as well as additional online meetings. Ad-hoc meetings between members also occur in conjunction to (relatively rare) project meetings. However, it is the Secretariat that plays the most prominent role in maintaining network cohesion through communication with the members via email and the website. In principle, the member institutions are to coordinate with other universities and institutes in their country but in practice this does not happen systematically.

The high degree of centrality of SANWATCE is one of the network's main strengths. The joint endorsement from both the African ministers responsible for water and also ministers of science and technology (see above), provide a strong political mandate from regional institutions, as does the positioning of the network as a NEPAD flagship programme. In addition, SANWATCE has a close relationship with SADC (which approves its business plans) and reports its research findings regularly to its Technical Advisory Committee.[18] Furthermore, the network is the implementing agent for the Water Research Fund for Southern Africa (WARFSA), a programme of SADC Water Division.[19] The network is also a key partner in the Joint Africa–EU Strategy, which provides a high-level cooperation platform for the EU and African partners, as well as the European Union Water Initiative, and its Africa Working Group. Therefore, SANWATCE occupies not only a central position amongst regional decision-making but also inter-regional planning and cooperation.

Network process

The most fundamental challenge for the network has been to access sufficient and sustainable sources of funding. Although the network was initiated and promoted by regional institutions (i.e. AU, NEPAD, SADC), these institutions (and their member states) have not provided the necessary resources to set up and maintain the network. Funding was eventually agreed by the South African Department of Science and Technology (DST) in 2009[18] for maintaining the core activities of the Secretariat. DST has continued to provide, and even increase, this funding over the lifetime of the network. In the financial years 2016/2017 and 2017/2018, DST provided ZAR2.5 million each year.

Wider research and capacity-building activities across the network, albeit fairly modest in scale, have been funded by a single international donor, the EU.[20,21] Between 2009 and 2018, the European Commission allocated approximately EUR2.3 million[22] for SANWATCE activities, which is considerably less than the EUR11.5 million requested in the 2013–2018 Business Plan. Therefore, funding is extremely scarce. While it is realised that in the long term SANWATCE must find ways to raise more resources and be sustainable,[18] so far there has only been limited support from other member states outside of South Africa – usually through in-kind contributions. In addition, in the current fiercely competitive environment, funding applications submitted to international donors and research funding bodies by the Secretariat on behalf of the network have so far been unsuccessful.

In the absence of substantial projects and research activities, the infrastructure, especially the Secretariat, has become particularly important in maintaining the network. The financial support from DST (as well as support in kind from Stellenbosch University) has helped ensure a strong secretariat run by two full-time and three part-time staff members. Besides maintaining communication with the network members, the Secretariat performs a number of functions, including: representing the network at international forums and in discussions with external partners; liaising with regional partners, especially in SADC; coordinating the legal affairs of the network as well as the day-to-day functioning of the network; preparing and submitting joint research proposals on behalf of the network; and creating inventories of expertise, skills and high end equipment. The strong role of the Secretariat, coupled with the dominance of South African actors and funding in the network, has, however, created tensions surrounding the governance of the network; a review of the network in 2012 reports that 'a perception may exist among other participating countries that they are less on the leading side on SANWATCE decisions and more as followers and recipients of approaches from the members within South Africa'[23].

The main collective advantage of the networks is branding as a NEPAD flagship network, which adds considerable legitimacy to joint funding proposals and approaches to donors. However, although it is perceived to be a significant advantage by stakeholders, it has not yet translated into a significant flow of resources into the network. The network has also helped focus attention (and indeed create) the SADC research agenda, which acts as a common vision in future research. However, with limited funding to carry out research and even less from sources within the SADC region, it is not clear how far the network will be able to implement this agenda. One recent promising initiative funded by the DST provides limited funding for postgraduate students in each member institution to work on projects within this research agenda, but this amount is insignificant compared to the research needs identified.

Network outcome

It is too early by far to assess the network in terms of ultimate outcomes (i.e. achieving their goals of contributing through research to improving societal and environmental challenges). However, various interim outcomes can be assessed. For example, the network has contributed to improving the collaborations and contact between the institutions and researchers involved.[24] The network has established a strong presence within relevant forums in SADC and internationally. SANWATCE has also fulfilled one of its objectives to contribute to the creation of a SADC research agenda in the sector.

Case 2: An evaluation of the Southern Africa Network for Biosciences

Network background

The NEPAD Southern Africa Network for Biosciences (SANBio) was established in 2005 within the framework of NEPAD's Centres of Excellence for Science and Technology. It covers the SADC region as one of five sub-regional networks established in the AU under the African Biosciences Initiative.[25] This initiative is directed at facilitating the establishment of state-of-the-art research and development facilities that can enable institutions to pool resources to address common biosciences challenges. The focus areas include human health, agriculture productivity, sustainable water resource use management, biodiversity management and sound environmental management. SANBio's objectives are to: support an effective and dynamic regional research network; enhance human and infrastructure capacity; and develop and commercialise innovation products in health and nutrition.[26]

Network structure

SANBio brings together partners from five southern African countries. The network is structured on a 'hub and node' model facilitated by a secretariat. The hub is located at the Council for Scientific and Industrial Research (CSIR) in South Africa and provides financial management and operational support to the Secretariat (which is also hosted at the same location). Country nodes located at research and/or higher learning institutions throughout the region (Table 3) participate in regional research, development and innovation activities in different technical themes, including inland fisheries; indigenous knowledge systems; mushrooms; livestock; bioinformatics; traditional medicines for HIV/AIDS; and plant genetics. As with SANWATCE, securing commitment and participation from across the SADC region continues to be a challenge and despite ongoing efforts, several SADC states (e.g. Seychelles, Madagascar, Angola) have not been very active in the network.

Cohesion in the network is mainly created through the research projects conducted within the nodes. SANBio projects must include three or more SADC countries and so provide opportunities for creating links between other network partners as well as international collaborating partners. Eight projects were initially funded from SANBio funds (see below) in the first phase of the network (i.e. 2009–2012) and broadly covered the topics of the six nodes. In the second phase of the network (2013–2018), 5 flagship projects and 10 smaller seed projects are being directly funded by the network (with matching national funding). Cohesion across the network (and between nodes) is less evident but the role of the Secretariat is critical here (as it was in SANWATCE). The nodes report back on progress on projects to the Secretariat which also maintains communication across the network via email and through the website. The nodes are also supposed to coordinate and engage with other universities and institutes that work in their areas of specialism across the SADC region; however, there is limited evidence available on the extent to which this occurs.

As with SANWATCE, the centrality of the network is relatively high, which in principle puts SANBio in a strong position to impact on regional decision-making. The African Biosciences Initiative – of which SANBio is a part – responds to the AU's 2005 Science and Technology Consolidated Plan of Action, which forms the basis for implementing NEPAD's Science and Technology programmes. SADC ministers also endorse the business plans of the network. However, an evaluation of the network after its first phase questions whether these top-down linkages between the network and the regional political institutions were enough to link the policies and strategies across the region with the research on the ground. The evaluation also argues that connections between the network and the private sector have also been neglected.[27]

Network process

SANBio has been able to raise significantly more resources for projects than SANWATCE and is perhaps making more headway in securing diverse and sustainable sources of funding for the future. However, attracting sufficient funds is still the main challenge of the network and substantially limits its activities vis a vis its business plan. Similar to SANWATCE, funding for the network has been provided by both DST and international donors. In 2005–2006, a grant of CAD450 000 was made available to SANBio by the Canadian International Development Agency and about ZAR1.3 million was provided by DST. These funds were to prepare a business plan, which was marketed to donors for funding support. In 2009, the Finnish Ministry of Foreign Affairs and the government of South Africa launched the Finnish–Southern African Partnership Programme to Strengthen NEPAD/SANBio (BioFISA).

In the first phase (2009–2011), the South African government, through DST, provided ZAR9 million and the Finnish Government EUR3 million.[28] The purpose of BioFISA was to strengthen the operational capacity of SANBio and to support it in rolling out its business plan, which included the eight initial research projects mentioned above.[28] The Finnish and South African governments agreed to renew the BioFISA programme for a second period (2015–2019) with a total budget of EUR7.82 million.

Table 3: SANBio Members

Institution	Country	Node
CSIR	South Africa	Traditional medicine/HIV and HIV/microbicides
Bunda College of Agriculture (now LUANAR), University of Malawi	Malawi	Fish biodiversity
University of Namibia	Namibia	Technology transfer on mushroom farming
SADC Plant Genetic Resources Centre (SPGRC)	Zambia	Conservation of plant genetic resources (gene banks)
University of Zambia (UNZA)/National Institute for Scientific and Industrial Research (NISIR)	Zambia	Livestock production
University of Mauritius	Mauritius	Bioinformatics
North-West University	South Africa	Indigenous knowledge systems

In the second phase, several new initiatives are aimed at generating diverse and sustainable sources of income – which was seen as a weakness of the first phase of the network. For example, member states must pay a 40% contribution toward projects in which they participate – 20% in cash and 20% in kind (e.g. office space, electricity, water, transport and staff time)[29]; and more emphasis is placed on commercialisation and 'bring-to-market' objectives involving the private sector[30]. While there are some signs that these efforts are beginning to pay off (e.g. five large externally funded projects are expected to bring in USD1.7 million of investments in the financial year 2016/2017),[31] it seems unlikely that the ambitious targets in the business plan for leveraging external funds will be met. In the 2006–2011 business plans, the target was to raise USD50 million (while EUR6 million in cash and ZAR14 million in kind was raised). In the 2013–2018 business plan, the target is to raise USD80 million.[30]

SANBio's infrastructure is composed of three main components: a central Hub, a Secretariat, and also a Steering Committee consisting of one representative from each member state. The Hub is intended to provide financial management and operational support to the Secretariat. In practice, however, various challenges have arisen from the location of the Hub in a national institution like CSIR, including that its processes and payment procedures are designed to work with projects within South Africa and not internationally.[31] The Secretariat also has had internal challenges. For example, the Secretariat staff are on short-term contracts without certain benefits which makes it difficult to attract and retain the best talent to SANBio. Furthermore, the role of the Steering Committee was criticised after the first phase for contributing to an overly top-down and bureaucratic governance structure. Changes were consequently made in the second phase of the network. In 2015, the SANBio Secretariat received a boost in numbers as the team both changed and expanded to six staff members in 2015/2016.[31] In addition, CSIR appointed a Hub Manager who may be able to better promote interaction between the network and the Hub.[31]

The main activities of the network involve joint research and development projects. In the evaluation of the first phase of the projects, several stakeholders noted that network creation had been achieved within each project.[27] However, other stakeholders emphasised that beyond these specific projects, the 'network of networks' had yet to be established. Therefore, although there appears to be an indication that network members are working together to achieve shared research goals within projects and nodes, it is not yet clear how effective this collaboration is in practice.

Network outcome

While it is still early to assess the network in terms of its ultimate goal of addressing common African bioscience challenges, it is possible to evidence achievements within the projects and nodes, in terms of training workshops, products development, publications and students trained etc.[27,30] The network is also participating in a number of technical working groups under the AU Research for Health Strategy and is involved in SADC Science, Technology and Innovation and Health Sector programmes. Other interim outcomes achieved so far have been the increasing level of funds being raised outside of the network and its core funders.

Network constraints and facilitators

The evaluation framework allows us to identify common constraints and facilitators across the two networks that constrain and/or enhance their effectiveness. Firstly, both networks struggled to overcome issues of representation across the regions. Although research networks are designed to build capacity and pool resources so that weak members can benefit from the capacity of stronger members, certain basic criteria need to be reached before joining the network is feasible. Secondly, certain network members inevitably have more resources than others, which can create tensions and asymmetries in the network. In both networks, South African institutes were over-represented. The South African government was also a major donor of the networks. It is therefore perhaps not surprising that, at times, network members from other countries felt that they were not equal network members. Thirdly, funding was the key constraint in both networks. International donors are becoming increasingly interested in funding research networks but there remains a question about whether this approach focuses the agenda away from local priorities. Increasing funding from SADC members would overcome this danger, but SADC members tend to be cash-strapped and focused on more immediate priorities within their own political agendas. In addition, only South Africa has a dedicated department for science and technology, while in other countries there are also strongly competing areas for science support.

The two networks evaluated also share some common strengths that facilitated them in their day-to-day activities despite these challenges. Firstly, both networks benefit from strong secretariats. In the absence of sufficient project funding, the infrastructure of a network becomes critical in maintaining the network. Strong secretariats are therefore essential because top-down hierarchical structures are unlikely to be accepted by network members in what are essentially flat governance structures. Secondly, both networks have a high level of centrality in terms of their influence over regional research agendas and links to decision-makers at sub-regional and regional levels. This is connected to a third strength of both networks: there was a clear advantage for the network members in joining together under the banner of the NEPAD networks in terms of visibility with both funders and policymakers. This shared branding enhanced the legitimacy and credibility of the institutions when operating under the network. While this advantage has yet to be fully realised, especially in terms of attracting funding, it does seem to be a real prize on the horizon.

Conclusions

Demand for more evidence-based policy is increasing across Africa as most political systems undergo a process of liberalisation, thus putting to an end the monopoly of policy analysis and increasing the public scrutiny of policy action.[8] In addition, policy problems are becoming more complex and transnational, prompting a greater need for policymaking at a regional level.[8] If African researchers and research institutions are to play an effective role in these changing dynamics, they need to grow and increase their capacity.[7] Regional research networks aim to pool resources, share experiences and build capacity with the aim of increasing the quality and the impact of research. We argue that the mere existence of networks, however, is not enough to make the science–policy interface more porous to homegrown policy solutions: these networks must function effectively. The evaluation of the SANWATCE and SANBio networks here, however, reveals some shared challenges which are limiting their effectiveness, including achieving adequate regional representation; unequal contribution and power relations between members; and securing sufficient, diverse and sustainable resources.

While the relationship between evidence and policymaking still needs to be better understood around the world, there is a particular gap in the literature on this process in developing countries. This article was aimed at helping to fill this research gap by presenting an evaluation framework that can be used to evaluate other research networks on the continent. The objective of the evaluation framework is not to provide the type of in-depth review of the networks that donors and other stakeholders might require to report and monitor progress, but instead to provide a broad-brush outline of the networks in a way that can be compared across networks. This exercise will help build up a picture of how (and how well) research networks on the continent are operating in practice. Most importantly, the comparative evaluation of research networks on the continent should help to promote critical thinking about these networks that could help to develop recommendations on how these networks can be improved and better designed in future. With over 50 such research networks listed on the continent,[11] we would argue that this type of research is long overdue.

However, at this stage (i.e. with the evaluation of only two research networks) many more research questions are raised than are answered. For example, how can research networks be designed in such a way as to include policymakers and so better bridge the divide between the 'two communities' model of policymaking? Related to this question, we might ask: Can networks be designed to better incorporate indigenous knowledge into science and so help break down perceptions that science is the

domain of the imperialist West while Africa can only contribute cultural experiences? Can research networks help overcome the dichotomy in countries that are more globally connected and those that are not? How does international donor funding influence these networks? Does it risk missing the most relevant issues for African countries? Answering these questions will be critical to understanding how localised research networks can better input into future policymaking in Africa.

Acknowledgements

This research was funded by the British Academy Newton Mobility Grant (grant no. NG150251) and the Leverhulme Trust International Networks Scheme (project number IN-2014-014).

Authors' contributions

C.A. was the lead author and team coordinator; N.E. was responsible for the concept design and gave input into the empirical section on SANWATCE; E.C. gave input into the empirical section on SANBio; and D.B. gave input into the overall research design and draft text.

References

1. Godfrey L, Funke N, Mbizvo C. Bridging the science–policy interface: A new era for South African research and the role of knowledge brokering. S Afr J Sci. 2010;106(5–6):44–51. https://doi.org/10.4102/sajs.v106i5/6.247
2. Marouani AM, Ayuki ET. Introduction. In: Ayuk ET, Mohamed AM, editors. The policy paradox in Africa: Strengthening links between economic research and policymaking. Trenton, NJ/Ottowa: African World Press/ International Development Research Centre; 2007. p. 3–18. http://dx.doi.org/10.13140/2.1.2539.1043
3. Jones N, Jones H, Walsh C. Political science? Strengthening science–policy dialogue in developing countries. ODI Working Paper 294 [document on the Internet]. c2008 [cited 2017 Oct 24]. Available from: https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/474.pdf
4. Banyouko Ndah A. Public policy and policy inappropriateness in Africa: Causes, consequences and the way forward [document on the Internet]. c2010.[cited 2017 Oct 24]. Available from: http://www.academia.edu/318501/Public_policy_and_policy_inappropriateness_in_Africa_Causes_consequences_and_the_way_forward
5. Olokoshi A. Democratization, globalization and effective policy making in Africa. Ottowa: International Development Research Centre; 2000.
6. Dollar D, Svensson J. What explains the success or failure of structural adjustment programs? Econ J. 2000;110:894–917. https://doi.org/10.1111/1468-0297.00569
7. Ayuk ET, Jones B. From myth to reality: Building capacity for economic policy research in Africa. In: Ayuk ET, Mohamed AM, editors. The policy paradox in Africa: Strengthening links between economic research and policymaking. Trenton/Ottowa: African World Press/International Development Research Centre; 2007. p. 117–130. http://dx.doi.org/10.13140/2.1.2539.1043
8. Hansohm D, Naimhwaka E. Joining forces in policy research networks for policy-making in Africa: The SEAPREN experience. In: Ayuk ET, Mohamed AM, editors. The policy paradox in Africa: Strengthening links between economic research and policymaking. Trenton/Ottowa: African World Press/ International Development Research Centre; 2007. p. 131–149. http://dx.doi.org/10.13140/2.1.2539.1043
9. Toye R. Structural adjustment: Context, assumptions, origin and diversity. Paper presented at: The DGIS/ISSAS Conference on Structural Adjustment and Beyond; 1993 June 1–3; The Hague, the Netherlands.
10. Ajakaiye O. Levelling the playing field: Strengthening the role of African research in policy-making in and for sub-Saharan Africa. In: Ayuk ET, Mohamed AM, editors. The policy paradox in Africa: Strengthening links between economic research and policymaking. Trenton/Ottowa: African World Press/International Development Research Centre; 2007. p. 19–39. http://dx.doi.org/10.13140/2.1.2539.1043
11. COHRED. Directory of Selected African Research Networks. 2nd ed [document on the Internet]. [cited 2017 Oct 24]. Available from: http://www.cohred.org/downloads/564.pdf
12. Bhorat H. The nine commandments – a South African case study identifies roles for applied research institutions. In: Ayuk ET, Mohamed AM, editors. The policy paradox in Africa: Strengthening links between economic research and policymaking. Trenton/Ottowa: African World Press/ International Development Research Centre; 2007. p. 185–200. http://dx.doi.org/10.13140/2.1.2539.1043
13. Network Impact and Center for Evaluation Innovation. Framing paper: The state of network evaluation [document on the Internet]. c2014 [cited 2017 Oct 24]. Available from: http://www.networkimpact.org/wp-content/uploads/2014/09/NetworkEvalGuidePt1_FramingPaper.pdf
14. Hill C. Network literature review: Conceptualizing and evaluating networks. Calgary: The Southern Alberta Child and Youth Health Network; 2002.
15. Provan KG, Milward HB. Do networks really work? A framework for evaluating public sector organizational networks. Public Adm Rev. 2001;61(4):414–423. https://doi.org/10.1111/0033-3352.00045
16. NEPAD SANWATCE. Our vision [webpage on the Internet]. No date [cited 2016 Dec 20]. Available from: http://nepadwatercoe.org/aboutus/our-vision/
17. NEPAD SANWATCE. Expression of interest: 2nd call for membership [webpage on the Internet]. No date [cited 2016 Dec 20]. Available from: http://nepadwatercoe.org/aboutus/expanding-the-network/expression-of-interest-2nd-call-for-membership/#Point4
18. Botha AP, Gruenewaldt G, Botha TC. A review of DST support provided to the NEPAD Science and Technology Flagship Programmes. Report TS-109-001-2012-REP 1.0. prepared by Technoscene for the Department of Science and Technology; February 2013. Unpublished report.
19. NEDPAD SANWATCE. Our mandate [webpage on the Internet]. No date [cited 2016 Dec 20]. Available from: http://nepadwatercoe.org/aboutus/our-mandate/
20. Dennis R, Parida BP, Mannel D, Kenabatho P, Braune E. Delivering capacity across the SADC region's water sector – JLP 1.4. Report prepared for the European Commission Joint Research Centre [document on the Internet]. c2013 [cited 2016 Dec 20]. Available from: http://nepadwatercoe.org/wp-content/uploads/JLP1-4_UB_approved.pdf
21. NEPAD SANWATCE. Current projects [webpage on the Internet]. No date [cited 2016 Dec 20]. Available from: http://nepadwatercoe.org/current-projects/using-iwrm-best-practices-to-develop-appropriate-capacity-and-training-for-the-benefit-of-sub-saharan-africa-water-security-act4ssaws/
22. NEPAD SANWATCE. Administration and coordination of activities. The AU/ NEPAD Southern African Network of Water Centres of Excellence. Progress report December 2014–2015 [document on the Internet]. c2016 [cited 2016 Dec 20]. Available from: http://nepadwatercoe.org/wp-content/uploads/SANWATCE-Progress-report-Dec2014-Nov2015FINAL.pdf
23. Botha AP, Gruenewaldt G, Botha TC. A review of DST support provided to the NEPAD Science and Technology Flagship Programmes. Report TS-109-001-2012-REP 1.0. prepared by Technoscene for the Department of Science and Technology; February 2013. Unpublished report. p. 54.
24. Botha AP, Gruenewaldt G, Botha TC. A review of DST support provided to the NEPAD Science and Technology Flagship Programmes. Report TS-109-001-2012-REP 1.0. prepared by Technoscene for the Department of Science and Technology; February 2013. Unpublished report. p. 52.
25. SANBio. Welcome to SANBio [webpage on the Internet]. No date [cited 2016 Dec 20]. Available from: http://www.nepadSANBio.org
26. SANBio. About us [webpage on the Internet]. No date [cited 2016 Dec 20]. Available from: http://www.nepadSANBio.org/about-us.html
27. Finnish Consulting Group. Mid-term evaluation of BioFISA. Report presented to the Ministry for Foreign Affairs of Finland unit for Southern Africa; March 2012. Unpublished report.
28. SANBio. Celebrating four years of BioFISA [document on the Internet]. c2012 [cited 2017 Oct 24]. Available from: www.nepadsanbio.org/.../sanbio-newsletter-celebrating-four-years-biofisa-2009-2012
29. SANBio BioFISA II. BioFISA II Business meets biosciences [document on the Internet]. No date [cited 2016 Dec 10]. Available from: http://www.nepadSANBio.org/sites/default/files/BioFISA%20II%20Brochure%20for%20Web.pdf
30. SANBio. Business plan 2014–2018 [document on the Internet]. No date [cited 2016 Dec 10]. Available from: http://www.nepadSANBio.org/sites/default/files/SANBio_Bus_Plan_Proof_1_(1)_Original.pdf
31. SANBio. Annual report 2015. Pretoria: SANBio; 2016.

Temporal ranges and ancestry in the hominin fossil record: The case of *Australopithecus sediba*

AUTHORS:
Chris Robinson[1]
Timothy L. Campbell[2]
Susanne Cote[3]
Darryl J. de Ruiter[2]

AFFILIATIONS:
[1]Department of Biological Sciences, Bronx Community College (CUNY), Bronx, New York, USA
[2]Department of Anthropology, Texas A & M University, Texas, USA
[3]Department of Anthropology and Archaeology, University of Calgary, Calgary, Canada

CORRESPONDENCE TO:
Chris Robinson

EMAIL:
chris.robinson@bcc.cuny.edu

DATES:
Received: 07 Oct. 2017
Revised: 17 Dec. 2017
Accepted: 15 Jan. 2018
Published: 27 Mar. 2018

KEYWORDS:
sampling bias; hominin temporal ranges; FAD; LAD; genus *Homo* ancestry

HOW TO CITE:
Robinson C, Campbell TL, Cote S, De Ruiter DJ. Temporal ranges and ancestry in the hominin fossil record: The case of *Australopithecus sediba*. S Afr J Sci. 2018;114(3/4), Art. #2017-0327, 7 pages. http://dx.doi.org/10.17159/sajs.2018/20170327

ARTICLE INCLUDES:
× Supplementary material
× Data set

FUNDING:
Texas A&M University; University of Calgary

In attempting to resolve the phylogenetic relationships of fossil taxa, researchers can use evidence from two sources – morphology and known temporal ranges. For most taxa, the available evidence is stronger for one of these data sources. We examined the limitations of temporal data for reconstructing hominin evolutionary relationships, specifically focusing on the hypothesised ancestor–descendant relationship between *Australopithecus sediba* and the genus *Homo*. Some have implied that because the only known specimens of *A. sediba* are dated to later than the earliest fossils attributed to *Homo*, the former species is precluded from being ancestral to the latter. However, *A. sediba* is currently known from one site dated to 1.98 Ma and, thus, its actual temporal range is unknown. Using data from the currently known temporal ranges of fossil hominin species, and incorporating dating error in the analysis, we estimate that the average hominin species' temporal range is ~0.97 Myr, which is lower than most figures suggested for mammalian species generally. Using this conservative figure in a thought experiment in which the Malapa specimens are hypothesised to represent the last appearance date, the middle of the temporal range, and first appearance date for the species, the first appearance date of *A. sediba* would be 2.95, 2.47 and 1.98 Ma, respectively. As these scenarios are all equally plausible, and 2.95 Ma predates the earliest specimens that some have attributed to *Homo*, we cannot refute the hypothesis that the species *A. sediba* is ancestral to our genus based solely on currently available temporal data.

Significance:

- We correct a common misconception in palaeoanthropology that a species currently known only from later in time than another species cannot be ancestral to it.
- On temporal grounds alone one cannot dismiss the possibility that *A. sediba* could be ancestral to the genus *Homo*.

Introduction

In evaluating competing phylogenetic hypotheses there are two primary sources of data available to palaeontologists for most fossil taxa: the morphology of the taxa under investigation and their known temporal ranges. For many taxonomic groups, however, the quality of these two sources of data differs substantially. Some taxa are well known morphologically, yet are spatially and temporally restricted. Other taxa are well sampled across sites and through time, yet are represented by limited and/or fragmentary anatomical elements. While some researchers have argued that the incomplete nature of the fossil record makes temporal information unreliable for reconstructing phylogenetic relationships[1,2], and that using 'age to define...ancestry is eminently circular'[3(p.439)], morphological evidence regarding evolutionary relationships among fossil taxa can also be equivocal, such as when there are multiple equally parsimonious phylogenetic trees or when there is the potential that homoplasy has substantially influenced phylogenetic interpretations (see Wood and Harrison[4] for a discussion of the latter in hominins). Consequently, it is important to consider the relative strengths of these sources of data when evaluating phylogenetic hypotheses.

Here, we focused on the limitations of temporal data for reconstructing evolutionary relationships in the hominin fossil record, using *Australopithecus sediba* as a case study. Originally proposed as the probable ancestor of the genus *Homo*[5,6], some have contended that this scenario is unlikely based on both morphological[7-9] and, either directly or implicitly, temporal grounds (e.g. see comments by White in Balter[10], Grine in Cherry[11] and Richmond in Gibbons[12]). Leaving aside the morphological arguments for others to debate, here we examine the suggestion that the currently understood first appearance date (FAD) for *A. sediba* of ~1.98 Ma (million years ago) [5,6], in and of itself, negates it as a potential ancestor of the genus *Homo* because putative fossils of early *Homo* appear earlier in the geological record[13,14]. As these critiques derive from news pieces rather than scholarly articles it is possible that the quotes have been taken out of context and do not reflect what the researchers intended to say. However, they give the impression that at least some in the field of palaeoanthropology, like many in the general public and popular press (see Gibbons[15] for a recent example), think that if all representatives of one taxon are found later in time than at least one specimen attributed to another taxon, it implies that the former cannot be members of the ancestral lineage from which the latter evolved. This issue is especially relevant as these misconceptions are currently being presented in college anthropology textbooks.[16(p.154)]

As noted by Spoor[7], two scenarios have been proposed that are consistent with the hypothesis that *A. sediba* is ancestral to *Homo*. First, Berger et al.[5] hypothesised that the Malapa specimens represent late surviving members of the species that gave rise to *Homo* earlier in time. Second, Pickering et al.[6] questioned the validity of specimens attributed to *Homo* that had been recovered from strata predating the Malapa deposits, and suggested that *A. sediba* cannot be precluded as a potential ancestor of *Homo* based on the age of the fossils from Malapa. Notwithstanding the difficulties in recognising early members of the genus *Homo*, we concur with Spoor[7] that the first scenario is more likely. Therefore, in this paper we explore the question of whether it is plausible for *A. sediba* to be the ancestor of the genus *Homo* based on the FADs of specimens currently attributed to these two taxa.

Research methods and data

Contemporaneous ancestors and descendants in the fossil record

Depending on the mode of speciation, it may or may not be possible for ancestral and descendant taxa to coexist in the fossil record. Speciation resulting from bifurcating cladogenesis or anagenesis (Figure 1: Modes 1 and 2) precludes ancestors from being contemporaneous with their descendants because in both cases the entire ancestral species evolves into one or more descendant species. Alternatively, under a budding cladogenesis model of speciation (Figure 1: Mode 3) ancestors and descendants do co-occur.[17] In budding cladogenesis, a subset of a species, usually a geographically isolated population, differentiates from the rest of the species and forms a new descendant taxon. Within the fossil record this is seen as a change in morphology in one population, while the remaining populations retain the ancestral form. Thus, as far as can be perceived, the ancestral species persists after giving rise to its descendant. Recent studies of animal and plant biogeography suggest that this mode of speciation is relatively common[18], and there is genetic evidence that some ancestral species are extant, living contemporaneously with their descendants[19]. Some have even argued that budding cladogenesis is the primary mode by which species arise, with most ancestral taxa existing contemporaneously with their descendants for some time.[20,21]

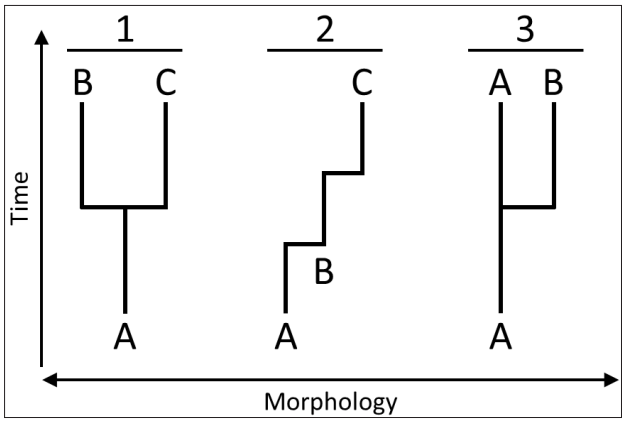


Figure 1: Modes of speciation: (1) bifurcating cladogenesis, (2) anagenesis and (3) budding cladogenesis. In both Modes 1 and 2, ancestors cannot be contemporaneous with their descendants as the entire ancestral species evolves into one or more descendant species. In Mode 3, ancestral and descendant taxa can coexist as a, typically geographically isolated, population speciates from the main population, which retains its species integrity.

The coexistence of ancestral and descendant species, which would imply that the latter species evolved via budding cladogenesis, has been hypothesised to be present in the fossil record for a variety of invertebrate groups including bryozoans, ostracods and mollusks (see citations in Gould[22]). Additionally, budding cladogenesis has been suggested, either explicitly or implied through the hypothesised coexistence of ancestors and descendants, in phylogenetic reconstructions of numerous mammalian groups spanning almost the entire range of body sizes including rodents[23], suids[24,25], equids[25], hippopotamids[26] and proboscideans[25,27].

Among hominins, examples of putative ancestral and descendant species existing contemporaneously include *Homo habilis* and *Homo erectus*, *H. erectus* and most later *Homo* species, and *Australopithecus afarensis* and *Australopithecus africanus*.[4,28,29] Even the transition of *Australopithecus anamensis* into *A. afarensis* via anagenesis – generally considered to be the strongest example of this mode of speciation in the early hominin fossil record[30,31] – has been questioned based on a critical reevaluation of the morphological differences between the older *A. afarensis* material from Laetoli and specimens from the younger Hadar deposits[32]. Moreover, fossils tentatively assigned to *A. afarensis* from the site of Fejej in southern Ethiopia (e.g. Fleagle et al.[33,34] but see Ward[35] for an alternative view) overlap with the currently known temporal range of *A. anamensis*.[36] If this diagnosis holds, then the evolution of *A. afarensis* from *A. anamensis* must have also occurred via budding cladogenesis. Although some have argued that the evidence for hominins rarely meets the criteria for budding cladogenesis, specifically that there are few examples of ancestors and descendants overlapping in time[30], if we accept that speciation in hominins can occur via this mechanism, as appears to be common in many other taxa, then contemporaneity in and of itself cannot be used to refute a potential ancestor–descendant relationship between *A. sediba* and the genus *Homo*, just as the proposed ~250 kyr of temporal overlap between *H. erectus* and *H. habilis* in East Africa does not preclude the latter taxon from being the progenitor of the former.[29,37]

Recognising budding cladogenesis requires documenting that ancestors and descendants co-occur in the fossil record. It is of course implicit that any specimens utilised are correctly identified to taxon. As an analysis of hominin alpha taxonomy is beyond the scope of this paper, we rely on the analyses of the experts working on the species in question for the identification of the earliest and latest examples of each taxon. Detailed criteria for identifying ancestors in the fossil record are set out by Delson[3]. If, for example, one seeks to test whether it is plausible that *A. sediba* is the ancestor of the genus *Homo* it would first be necessary to provide evidence that *A. sediba* is the sister taxon of the genus *Homo* (i.e. that it shares synapomorphies with *Homo* that other hominin taxa do not possess). A sister group relationship has been suggested in an extensive recent phylogenetic analysis of hominins[38] and in the original description of *A. sediba*[5]; however, as with any phylogenetic hypothesis, it must withstand further testing by other researchers, especially when additional evidence is obtained. For *A. sediba* to be ancestral to the genus *Homo* it would also have to exhibit more primitive hominin features (plesiomorphies) than *Homo* and cannot exhibit any uniquely derived characters (autapomorphies) as these would indicate that it is also a descendant of the 'true ancestor' of both groups. If these criteria are met then one could argue that *A. sediba* corresponds to the ancestral morphotype of the node shared with *Homo* that links them as sister taxa. Moreover, Delson[3(p.440)] cogently argues that only after morphological criteria are met should one then consider other data such as a taxon being 'widespread, polytypic, anatomically well known and perhaps of "correct" geographic and chronological age'.

The incomplete nature of the fossil record

The known fossil record likely represents fewer than half of the species that have lived[39], albeit with large differences in representation among taxonomic groups, as a result, at least partly, of differential preservation[40]. With respect to primates, it is estimated that between about 3.8% and 7% of taxa have been sampled in the fossil record.[41,42]

Although it is implicit in evolutionary theory that some portion of an ancestral taxon necessarily preceded its descendants[3], palaeontologists have long recognised that ancestors can potentially be found in strata dated to later in time than those of their descendants as a consequence of the incomplete nature of the fossil record[3,43]. For example, even after accounting for the known temporal ranges of the 'abundant and heavily studied' North American fossil hipparionines, Alroy[44(p.167)] notes that the most parsimonious phylogenetic hypothesis has two ancestral species that, based on the available evidence, arose one million years after their descendants, and two additional ancestral taxa that have the same FAD as their descendants. In other examples, the possibility of descendant taxa preceding their ancestors was explored by researchers studying graptoloids and echinoids.[45,46] Although these researchers ultimately judged those scenarios as less likely than alternative phylogenetic hypotheses, they did not dismiss them based solely on descendants being present in the known fossil record prior to their ancestors.

One might argue that the intensive focus on recovering hominins over the past ~100 years at sites in East and South Africa would have resulted in a relatively complete fossil record. However, hominins make up very small percentages of most Plio-Pleistocene faunal assemblages

when compared to other medium- and large-sized mammals. Although published percentages are not fully comparable as different research teams include different taxonomic groups and body size subsets of the overall mammalian assemblages in their data sets, hominins are nonetheless relatively rare at most African sites from this time period. For instance, hominins make up only 5.3% of the reported mammalian assemblages at sites yielding robust australopiths in South Africa,[47] and even this figure is likely to be an overestimate given the enhanced scrutiny that they receive relative to non-hominins. Hominins are even less common (<2%) at sites in East Africa such as Omo, Hadar and Laetoli[48-50], although they make up a greater percentage (9% of the vertebrate fauna) of the smaller sample of mammalian fossils from Kanapoi[51].

If we accept that hominin fossils are rare, their recovery is likely to be more strongly influenced by stochastic factors than is the case for other mammals (e.g. bovids or suids). Consequently, current FADs and last appearance dates (LADs) for hominins are likely not representative of their actual temporal ranges[52,53], and the FADs of hominins that are not known from long stratigraphic sections in particular are 'subject to substantial error' [54(p.10375)]. Moreover, some have noted that the hominin record has 'a disproportionate contribution from the East African Rift Valley' and, accordingly, have contended that this 'precludes firm conclusions regarding immigration or speciation events'[50(p.178)]. It is entirely possible that earlier (or later) populations of hominin species that are currently only known from one or a few localities and from a limited temporal range will be identified in under-sampled regions of Africa. Depending on how many distinct hominin species one recognises, there are as many as seven that are currently known from only one locality (i.e. they are arguably 'single hits') (Table 1). As Foote and Raup[55(p.136)] argue, 'a very high frequency of single hits suggests the possibility of a poor fossil record which should be analyzed with caution'. The implications of the above are that hominins are not as well known as might be expected based on the number of published articles on this group, and that a substantial amount of the hominin fossil record may be unknown. This claim is bolstered by the relatively large number of new hominin taxa named over the past 25 years, which has nearly doubled the number of putative hominin species (see citations in Wood and Boyle[56]). As such, it is likely that even the more generous estimates for hominin species durations significantly underestimate the true temporal ranges of these taxa, and this needs to be considered when evaluating hypotheses of ancestor–descendant relationships.

Species temporal ranges

Numerous methods have been developed to estimate 'true' temporal ranges for fossil taxa,[45,57-59] but none of these methods can be applied to *A. sediba* because they require that specimens be known from more than a single stratum. Thus, calculating confidence intervals for the FAD and LAD of *A. sediba* using these methods is not possible and we must use indirect methods for estimating its temporal range.

One million years (Myr) has been cited as the typical mammalian species longevity (e.g. Martin[41] and references therein), a value that can be traced back through several studies to Kurtén's[60] analysis of the Pleistocene mammals of Europe. In contrast, recently compiled average species durations from a survey of published data sets of Cenozoic mammals ranged from 0.8 to 6.3 Myr.[61] Although these types of estimations are highly dependent on the group of mammals under consideration (e.g. large versus small mammals) and the data set used, most of the studies yielded average species durations between 2 and 4 Myr.[61] Taxonomic practices (e.g. tendencies to 'lump' or 'split') also influence estimates of species longevity and vary between groups. Given that related taxonomic groups tend to share similar preservation potential[17,58], and similar risks of extinction[40], it seems most appropriate to use the temporal ranges of fossil primate species in general, and hominins in particular, as models for hominin species durations.

Unfortunately, few estimates of overall primate species longevity can be found in the literature (e.g. Martin[41]). Based on published hominin species temporal ranges from a recent study,[56] the estimated average hominin species longevity is 0.43 Myr, which is substantially lower than that of other mammalian groups. However, these estimates include taxa that are recorded from only one locality. Given that species known from a single locality cannot provide an estimate of the temporal range of that species, unless that locality samples a range of time within well-defined strata, removing them from these calculations seems appropriate. By removing these taxa, and *Homo sapiens*, which lacks an LAD at the time of writing, the average hominin species duration would be 0.50 Myr using published hominin species range data, and 0.80 Myr when the estimates of dating error from Wood and Boyle[56] are incorporated (Table 1). Furthermore, if we group hominin species that many researchers 'lump' together (e.g. those that are listed as 'low confidence' in Table 2 of Wood and Boyle[56]), the average published, and with dating error, species durations for hominins rise to 0.62 and 0.97 Myr, respectively (Table 1). We acknowledge that the choice of which taxa to retain may not be agreed upon by all researchers, but note that the larger 0.97 Myr average hominin species duration utilising the data set with error is still on the low end of the ranges reported for other groups of mammals,[61] and is very close to the 1.0 ± 0.25 Myr range suggested by Wood and Boyle[56]. As such, we will use this estimate to assess the possibility that, based on their currently estimated temporal ranges, *A. sediba* could be ancestral to the genus *Homo.*

Using 0.97 Myr as the average species duration for hominins, we estimate three temporal distributions for *A. sediba* by assuming that the Malapa specimens represent either the LAD, midpoint or FAD of the species (Scenarios 1, 2, and 3) (Figure 2). If we assume the Malapa specimens represent the LAD (Scenario 1), then the species would have originated around 2.95 Ma and gone extinct at 1.98 Ma. Alternatively, if Malapa is at the midpoint of the temporal range for the species (Scenario 2), then *A. sediba* can be estimated to have arisen at around 2.47 Ma and gone extinct at around 1.49 Ma. Finally, the possibility that the Malapa specimens represent early members of the species, chronologically closer to or at the speciation event (FAD) (Scenario 3), yields a potential temporal range of 1.98–1.01 Ma.

Thus, Scenarios 1, 2 and 3 would date the FAD of *A. sediba* to 2.95, 2.47 and 1.98 Ma, respectively, with all three being equally plausible based on currently available temporal data. The earliest specimen that some have attributed to the genus *Homo* is the partial mandible LD 350-1 from Ledi-Geraru, which is dated to 2.75–2.8 Ma.[14] Under Scenarios 2 and 3 in which the Malapa specimens represent the midpoint of the temporal range or the FAD of *A. sediba* it would not be possible for that species to be the ancestor of the genus *Homo* based on the estimate we are using for the average hominin temporal range. However, under Scenario 1 (i.e. the Malapa specimens represent the LAD for *A. sediba*), this ancestor–descendant relationship would be possible even if we incorporate the 'with error' FAD estimates from Wood and Boyle[56] for LD 350-1 (Figure 2).

It is important to carefully examine whether the first scenario is plausible, because if it is not, then the possibility that *A. sediba* is ancestral to the genus *Homo* would be considered unlikely given the parameters discussed above. In this regard, we make several observations. First, if the Malapa specimens are correctly identified as part of an 'australopith adaptive grade'[5] then these deposits likely contain some of the latest surviving members of the gracile form of this grade of early hominin. As such, the dates for the Malapa deposits may be close to the LAD for *A. sediba*. Second, given the mosaic nature of the morphology of *A. sediba*[5] and, thus, the difficulty of determining whether the taxon is represented by other, more incomplete specimens in the hominin fossil record, it may be that researchers have already recovered, or will recover at a later date, other fossils from earlier (or later) in time that should be attributed to this taxon. Furthermore, we note that some researchers have questioned the taxonomic attribution of LD 350-1.[62] If the specimen does not belong to the genus *Homo*, the earliest putative specimens of our genus would be dated to ~2.4 Ma using published hominin species range data,[13] and 2.6 Ma utilising the 'with error' data (Table 1). As such, both Scenarios 1 and 2 would be possible if 2.4 Ma accurately reflects the FAD for *Homo*, although again, only Scenario 1 remains possible at a 2.6 Ma FAD for *Homo*. Recall, however, that the 0.97 Ma average hominin duration used to generate a FAD for *A. sediba* is only an estimate based on current temporal range data from other hominins that also suffer from incomplete sampling. It is entirely possible that this figure underestimates the true temporal range for *A. sediba*.

Table 1: Data from Wood and Boyle[56] used to calculate average hominin species duration estimates. The first series represent the 'conservative' data with corresponding first appearance date (FAD), last appearance date (LAD) and temporal ranges, while the second series represents the 'with error' data. Calculated average hominin durations are provided in bold, with those in brackets generated using a lumping approach. Taxa that were grouped together are indicated by footnotes. Taxa below the dashed line are considered single hits and were not used in average hominin species duration calculations.

	Observed range[a]			Dating error incorporated[b]		
Taxon	**FAD**	**LAD**	**Range**	**FAD**	**LAD**	**Range**
Orrorin tugenensis	6	5.7	0.3	6.14	5.52	0.62
Ardipithecus kadabba	6.3	5.2	1.1	6.7	5.11	1.59
Ardipithecus ramidus	4.51	4.3	0.21	4.6	4.262	0.338
Australopithecus anamensis	4.2	3.9	0.3	4.37	3.82	0.55
Australopithecus afarensis	3.7	3	0.7	3.89	2.9	0.99
Kenyanthropus platyops	3.54	3.35	0.19	3.65	3.35[c]	0.3
Australopithecus deyiremeda[d]	3.5	3.3	0.2	3.596	3.33	0.266
Australopithecus africanus	3	2.4	0.6	4.02	1.9	2.12
Paranthropus aethiopicus	2.66	2.3	0.36	2.73	2.23	0.5
Paranthropus boisei	2.3	1.3	1	2.5	1.15	1.35
Paranthropus robustus	2	1	1	2.27	0.87	1.4
Homo habilis	2.35	1.65	0.7	2.6	1.65[c]	0.95
Homo rudolfensis	2	1.95	0.05	2.09	1.78	0.31
Homo erectus	1.81	0.027	1.783	1.85	0.027[c]	1.823
Homo ergaster[e]	1.7	1.4	0.3	2.27	0.87	1.4
Homo antecessor[f]	1	0.936	0.064	1.2	0.936[c]	0.264
Homo heidelbergensis	0.7	0.1	0.6	0.7[c]	0.1[c]	0.6
Homo helmei[f]	0.26	0.08	0.18	0.26[c]	0.08[c]	0.18
Homo neanderthalensis	0.13	0.04	0.09	0.197	0.03922	0.15778
Homo rhodesiensis[f]	0.6	0.3	0.3	0.6[c]	0.3[c]	0.3
Average			**0.501 (0.620)**			**0.800 (0.969)**
Sahelanthropus tchadensis[g]	7.2	6.8	0.4	7.43	6.38	1.05
Australopithecus bahrelghazali[g]	3.58	3.58	–	3.85	3.31	0.54
Australopithecus garhi[g]	2.5	2.45	0.05	2.5[3]	2.488	0.012
Australopithecus sediba[g]	1.98	1.98	–	2.05	1.91	0.14
Homo georgicus[g]	1.85	1.77	0.08	1.85[c]	1.77[c]	0.08
Homo floresiensis[g]	0.074	0.017	0.057	0.108	0.016	0.092
Homo sapiens[g]	0.195	0	0.195	0.2	0	0.2
Homo naledi[h]	0.286	0.286	–	?	?	?

[a]*Conservative estimates reported in Wood and Boyle[56(table 1)].*

[b]*Estimates with dating error reported in Wood and Boyle[56(table 1)].*

[c]*No 'with error' date provided in original publication; this value represents those reported in the 'conservative estimate'.*

[d]*Taxon and associated dates lumped with* A. afarensis *in calculation of temporal range.*

[e]*Taxon and associated dates lumped with* H. erectus *in calculation of temporal range.*

[f]*Taxon and associated dates lumped with* H. heidelbergensis *in calculation of temporal range.*

[g]*'Single hit' taxa not considered in calculating average hominin duration.*

[h]*Mid-range of most parsimonious age estimates reported in Dirks et al.[67]*

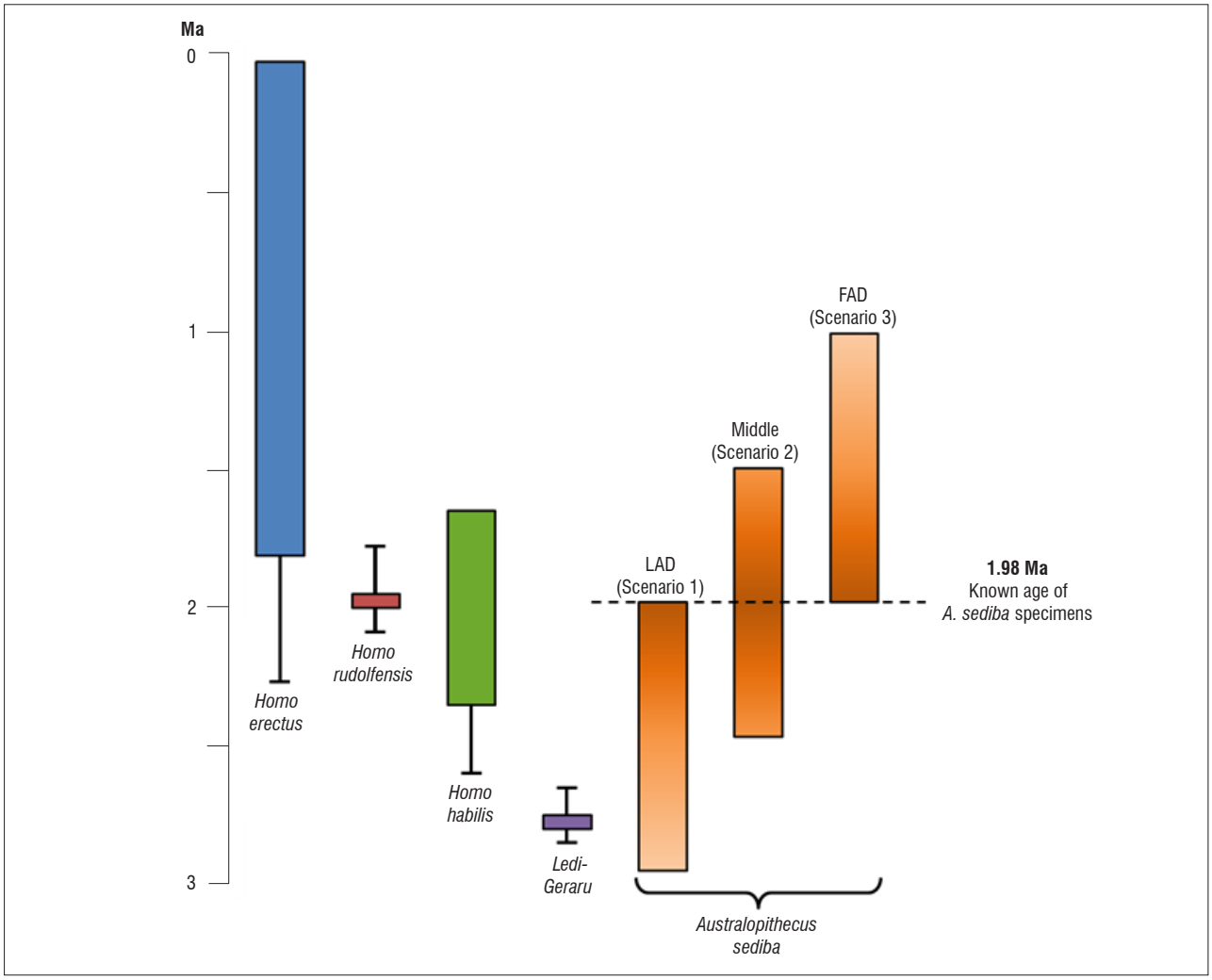


Figure 2: Estimates of the temporal ranges of the earliest species attributed to the genus *Homo* and potential durations for *Australopithecus sediba* assuming the Malapa specimens represent the last appearance date (LAD; Scenario 1), the middle of the species actual temporal range (Scenario 2) or first appearance date (FAD; Scenario 3) using an average hominin temporal duration of 0.97 Myr. Data used to generate the figure are from Table 1. Solid bars represent ranges using the 'conservative' data set while error bars incorporate both the 'with error' data and a lumping of the 'low confidence' taxa from Wood and Boyle[56]. Graded shading on the bars for *A. sediba* indicates the greater degree of uncertainty that the taxon existed at the time indicated.

Thus, although arguments can be made against Scenarios 2 and 3, we find insufficient evidence to refute Scenario 1, and, as a result, would argue that it is not implausible that *A. sediba* is the ancestor of the genus *Homo* based on our analysis of the dates of specimens currently attributed to these and other hominin taxa.

Conclusions

While some researchers have critiqued the hypothesis that *A. sediba* is ancestral to the genus *Homo* on morphological grounds, others have based their criticism, at least partly if not largely, on the date of the Malapa specimens. Although the known temporal range of a fossil species can be an important piece of evidence in testing ancestor–descendant hypotheses, palaeontologists do not typically dismiss the possibility that a fossil species is the ancestor of another based solely on the two species' currently recognised FADs.[44-46] Echoing previous researchers[3,63,64], Foote[17(p.147)] argues, 'whether species are preserved in the 'wrong' order does not affect the facts of their genealogical relationships, which we must attempt to reconstruct regardless of where we find the species stratigraphically'. This does not mean that temporal data cannot be informative, particularly for taxa that have a well-sampled fossil record, such as deep-sea microplankton.[43] However, if a taxon is not well known in the fossil record and/or is only known from a highly localised area, such as is the case for *A. sediba* and many other hominins (Table 1), the likelihood that its known record is an accurate reflection of the entire temporal range during which that taxon lived is substantially reduced. This limitation is demonstrated by the specimens attributed to *Homo floresiensis* and *Homo naledi* that are both suggested to have ghost lineages extending back much earlier in time based on their more primitive morphological features.[65,66] FADs and LADs are especially likely to be inaccurate for taxa like early hominins that had low population densities and 'relatively sparse fossil records'[53(fig.1)]. Therefore, it is difficult to have confidence in hypothesised evolutionary relationships that are based on the dates attributed to a handful of specimens.

Criticisms of the putative relationship of *A. sediba* and *Homo* based on their relative temporal ranges may be related, in part, to a dispute over the mode by which hominin species arose, with some taking the view that budding cladogenesis occurred rarely, if at all, in hominin evolution (e.g. White et al.[30]). However, as discussed above, this mode of speciation appears to be relatively common in mammalian, including hominin, evolution. Given that there is no theoretical reason to suspect that human evolution was any different than the evolution of any other mammal, we would argue that it would not be unusual to find evidence for contemporaneous ancestors and descendants in the human fossil record (e.g. Spoor et al.[29]).

Finally, we stress that the date for the Malapa deposits containing the *A. sediba* specimens should be interpreted for what it is – evidence of one particular moment in time when the species existed, but which cannot provide an accurate estimate of the lineage's temporal range (i.e. absence of evidence is not evidence of absence). In this paper, we are neither advocating for the position that *A. sediba* is the ancestor of the genus *Homo,* nor are we addressing questions about the taxonomic attribution of the Malapa specimens. We are merely pointing out that the critiques of *A. sediba* as a potential ancestor of the genus *Homo* based on temporal criteria are at the very least premature and are prone to misinterpretation by the media and general public. Until such time as additional data on its temporal range are available for *A. sediba,* any inferences about the evolutionary relationship between it and *Homo* should be based primarily on morphological data. While the arguments presented here have specifically revolved around one taxon, they are germane to studies of all hominin taxa in the fossil record, particularly 'single hit' taxa (Table 1) for which we have no clear understanding of the species' true temporal range. As such, we urge caution for all scientists involved in studies of human evolution to carefully think about how temporal data can, and should be, used in assessing phylogenetic hypotheses.

Acknowledgements

C.R. thanks Eric Delson and Lauren Halenar and T.L.C. thanks Kersten Bergstrom for their helpful comments during the development of this paper. Funding was generously provided by the Texas A&M University Cornerstone Faculty Fellowship and the University of Calgary.

Authors' contributions

C.R. and D.J.d.R. conceptualised the project with contributions to the research goals later provided by T.L.C. and S.C. T.L.C. developed the methodology and collected and analysed the data on fossil hominin first and last appearances. C.R. and T.L.C. wrote the initial draft with the table and figures created by T.L.C. Later contributions and revisions of the text were provided by D.J.d.R. and S.C. Revisions to the first submission were primarily written by T.L.C. and C.R.

References

1. Forey PL. Time as arbiter. In: Smith A, moderator. Is the fossil record adequate? Nature debates [online]. c1998 [cited 2017 Dec 17]. Available from: http://www.nature.com/nature/debates/fossil/fossil_11.html
2. Siddall ME. Stratigraphic fit to phylogenies: A proposed solution. Cladistics. 1998;14:201–208. https://dx.doi.org/10.1111/j.1096-0031.1998.tb00333.x
3. Delson E. Catarrhine phylogeny and classification: Principles, methods and comments. J Hum Evol. 1977;6:433–459. https://doi.org/10.1016/S0047-2484(77)80057-2
4. Wood B, Harrison T. The evolutionary context of the first hominins. Nature. 2011;470:347–352. https://dx.doi.org/10.1038/nature09709
5. Berger LR, De Ruiter DJ, Churchill SE, Schmid P, Carlson KJ, Dirks P, et al. *Australopithecus sediba*: A new species of *Homo*-like australopith from South Africa. Science. 2010;328:195–204. https://dx.doi.org/10.1126/science.1184944
6. Pickering R, Dirks P, Jinnah Z, De Ruiter DJ, Churchill SE, Herries AIR, et al. *Australopithecus sediba* at 1.977 Ma and implications for the origins of the genus *Homo*. Science. 2011;333:1421–1423. https://dx.doi.org/10.1126/science.1203697
7. Spoor F. Paleoanthropology: Malapa and the genus *Homo*. Nature. 2011;478:44–45. https://dx.doi.org/10.1038/478044a
8. Lordkipanidze D, Ponce de León MS, Margvelashvili A, Rak Y, Rightmire GP, Vekua A, et al. A complete skull from Dmanisi, Georgia, and the evolutionary biology of early *Homo*. Science. 2013;342:326–331. https://dx.doi.org/10.1126/science.1238484
9. Kimbel WH, Rak, Y. *Australopithecus sediba* and the emergence of *Homo*: Questionable evidence from the cranium of the juvenile holotype MH 1. J Hum Evol. 2017;107:94–106. https://doi.org/10.1016/j.jhevol.2017.03.011
10. Balter M. Candidate human ancestor from South Africa sparks praise and debate. Science. 2010;328:154–155. https://dx.doi.org/10.1126/science.328.5975.154
11. Cherry M. Claim over 'human ancestor' sparks furore. Nature News [online]. 2010 April 08. https://dx.doi.org/10.1038/news.2010.171
12. Gibbons A. A human smile and funny walk for *Australopithecus sediba*. Science. 2013;340:132–133. https://dx.doi.org/10.1126/science.340.6129.132
13. Kimbel WH, Johanson DC, Rak Y. Systematic assessment of a maxilla of *Homo* from Hadar, Ethiopia. Am J Phys Anthropol. 1997;103:235–262. https://dx.doi.org/10.1002/(SICI)1096-8644(199706)103:2<235::AID-AJPA8>3.0.CO;2-S
14. Villmoare B, Kimbel WH, Seyoum C, Campisano CJ, DiMaggio EN, Rowan J, et al. Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. Science. 2015;347:1352–1355. https://dx.doi.org/10.1126/science.aaa1343
15. Gibbons A. A famous 'ancestor' may be ousted from the human family. Science News [online]. 2017 April 23; Africa, Anthropology. https://dx.doi.org/10.1126/science.aal1099
16. Haviland WA, Prins H, Walrath D, McBride B. Anthropology: The human challenge. 15th ed. Belmont: Wadsworth Cengage; 2017.
17. Foote M. On the probability of ancestors in the fossil record. Paleobiology. 1996;22:141–151. https://dx.doi.org/10.1017/S0094837300016146
18. Hörandl E, Stuessy TF. Paraphyletic groups as natural units of biological classification. Taxon. 2010;59:1631–1653.
19. Wagner PJ, Marcot JD. Probabilistic phylogenetic inference in the fossil record: Current and future applications. In: Alroy J, Hunt G, editors. Paleontological Society Papers: Vol. 16. Quantitative methods in paleobiology. Lubbock: The Paleontological Society; 2010. p. 189–211.
20. Hörandl E. Paraphyletic versus monophyletic taxa – evolutionary versus cladistic classifications. Taxon. 2006;55:564–570.
21. Van Valen L. A new evolutionary law. Evol Theory. 1973;1:1–30.
22. Gould SJ. The structure of evolutionary theory. Cambridge: Belknap Press; 2002.
23. Flynn LJ. Species longevity, stasis, and stairsteps in rhizomyid rodents. In: Flanagan KM, Lillegraven JA, editors. Contributions to geology. Vol. 3: Vertebrates, phylogeny, and philosophy. Laramie, WY: University of Wyoming; 1986. p. 273–285.
24. White TD, Harris JM. Suid evolution and correlation of African hominid localities. Science. 1977;198:13–21. https://dx.doi.org/10.1126/science.331477
25. Cooke HBS. Horses, elephants and pigs as clues in the African later Cenozoic. In: Vogel JC, editor. Late Cainozoic palaeoclimates of the southern hemisphere. Rotterdam: A.A. Balkema; 1984. p. 473–482.
26. Boisserie J-R. The phylogeny and taxonomy of Hippopotamidae (Mammalia: Artiodactyla): A review based on morphology and cladistic analysis. Zool J Linnean Soc. 2005;143:1–26. http://dx.doi.org/10.1111/j.1096-3642.2004.00138.x
27. Maglio VJ. Origin and evolution of the Elephantidae. Trans Am Philos Soc. 1973;63:1–149.
28. McHenry H. Tempo and mode in human evolution. Proc Natl Acad Sci USA. 1994;91:6780–6786.
29. Spoor F, Leakey MG, Gathogo PN, Brown FH, Antón SC, McDougall I, et al. Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. Nature. 2007;448:688–691. https://dx.doi.org/10.1038/nature05986
30. White TD, WoldeGabriel G, Asfaw B, Ambrose S, Beyene Y, Bernor RL, et al. Asa Issie, Aramis and the origin of *Australopithecus*. Nature. 2006;440:883–889. https://dx.doi.org/10.1038/nature04629
31. Kimbel WH, Lockwood CA, Ward CV, Leakey MG, Rak Y, Johanson DC. Was *Australopithecus anamensis* ancestral to *A. afarensis*? A case of anagenesis in the hominin fossil record. J Hum Evol. 2006;51:134–152. https://doi.org/10.1016/j.jhevol.2006.02.003

32. Harrison T. Hominins from Upper Laetolil and Upper Ndolanya Beds, Laetoli. In: Harrison T, editor. Paleontology and geology of Laetoli: Human evolution in context: Volume 1: Geology, geochronology, paleoecology and paleoenvironment. New York: Springer; 2011. p. 141–188. https://doi.org/10.1007/978-90-481-9962-4_7

33. Fleagle JG, Rasmussen DT, Yirga S, Bown TM, Grine FE. New hominid fossils from Fejej, Southern Ethiopia. J Hum Evol. 1991;21:145–152. https://doi.org/10.1016/0047-2484(91)90005-G

34. Kappelman J, Swisher III CC, Fleagle JG, Yirga S, Bown TM, Feseha M. Age of *Australopithecus afarensis* from Fejej, Ethiopia. J Hum Evol. 1996;30:139–146. https://doi.org/10.1006/jhev.1996.0010

35. Ward C. Taxonomic affinity of the Pliocene hominin fossils from Fejej, Ethiopia. J Hum Evol. 2014;73:98–102. https://doi.org/10.1016/j.jhevol.2014.05.008

36. Brown FH, McDougall I, Gathogo PN. Age ranges of *Australopithecus* species, Kenya, Ethiopia, and Tanzania. In: Reed KE, Fleagle JG, Leakey RE, editors. The paleobiology of *Australopithecus*. New York: Springer; 2013. p. 7–20. https://doi.org/10.1007/978-94-007-5919-0_2

37. Spoor F, Gunz P, Neubauer S, Stelzer S, Scott N, Kwekason A, et al. Reconstructed *Homo habilis* type OH 7 suggests deep-rooted species diversity in early *Homo*. Nature. 2015;519:83–86. https://dx.doi.org/10.1038/nature14224

38. Dembo M, Natzke NJ, Mooers AØ, Collard M. Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. Proc R Soc Lond B. 2015;282, Art. #20150943, 9 pages. http://dx.doi.org/10.1098/rspb.2015.0943

39. Benton MJ, Hitchin R. Congruence between phylogenetic and stratigraphic data on the history of life. Proc R Soc Lond B. 1997;264:885–890. https://dx.doi.org/10.1098/rspb.1997.0123

40. Benton MJ. Molecular and morphological phylogenies of mammals: Congruence with stratigraphic data. Mol Phylogenet Evol. 1998;9:398–407. https://dx.doi.org/10.1006/mpev.1998.0492

41. Martin RD. Primate origins: Plugging the gaps. Nature. 1993;363:223–234. https://dx.doi.org/10.1038/363223a0

42. Tavaré S, Marshall CR, Will O, Soligo C, Martin RD. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature. 2002;416:726–729. https://dx.doi.org/10.1038/416726a

43. Lazarus DB, Prothero DR. The role of stratigraphic and morphologic data in phylogeny. J Paleontol. 1984;58:163–172.

44. Alroy J. Continuous track analysis: A new phylogenetic and biogeographic method. Syst Biol. 1995;44:152–178. https://dx.doi.org/10.1093/sysbio/44.2.152

45. Paul CRC. The adequacy of the fossil record. In: Joysey KA, Friday AE, editors. Problems of phylogenetic reconstruction. London: Academic Press; 1982. p. 7–117.

46. Melchin MJ, Mitchell CE, Naczk-Cameron A, Fan JX, Loxton J. Phylogeny and adaptive radiation of the Neograptina (Graptoloida) during the Hirnantian mass extinction and Silurian recovery. P Yorks Geol Soc. 2011;58:281–309. https://dx.doi.org/10.1144/pygs.58.4.301

47. De Ruiter DJ, Sponheimer M, Lee-Thorp JA. Indications of habitat association of *Australopithecus robustus* in the Bloubank Valley, South Africa. J Hum Evol. 2008;55:1015–1030. https://dx.doi.org/10.1016/j.jhevol.2008.06.003

48. Reed KE. Paleoecological patterns at the Hadar hominin site, Afar Regional State, Ethiopia. J Hum Evol. 2008;54:743–768. https://dx.doi.org/10.1016/j.jhevol.2007.08.013

49. Su D, Harrison T. Ecological implications of the relative rarity of fossil hominins at Laetoli. J Hum Evol. 2008;55:672–681. https://dx.doi.org/10.1016/j.jhevol.2008.07.003

50. Bobe R, Leakey MG. Ecology of Plio-Pleistocene mammals in the Omo–Turkana Basin and the emergence of *Homo*. In: Grine FE, Fleagle JG, Leakey RE, editors. The first humans: Origin and early evolution of the genus *Homo*. Amsterdam: Springer; 2009. p. 173–184.

51. Bobe R. Fossil mammals and paleoenvironments in the Omo-Turkana Basin. Evol Anthropol. 2011;20:254–263. http://dx.doi.org/10.1002/evan.20330

52. White TD. African omnivores: Global climatic change and Plio-Pleistocene hominids and suids. In: Vrba ES, Denton GH, Partridge TC, Burckle LH, editors. Paleoclimate and evolution with emphasis on human origins. New Haven, CT: Yale University Press; 1995. p. 369–384.

53. Wood B, Lonergan N. The hominin fossil record: Taxa, grades and clades. J Anat. 2008;212:354–376. https://dx.doi.org/10.1111/j.1469-7580.2008.00871.x

54. Wood B. Did early *Homo* migrate "out of" or "in to" Africa? Proc Natl Acad Sci USA. 2011;108:10375–10376. https://dx.doi.org/10.1073/pnas.1107724108

55. Foote M, Raup DM. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology. 1996;22:121–140. https://dx.doi.org/10.1017/S0094837300016134

56. Wood B, Boyle EK. Hominin taxic diversity: Fact or fantasy? Yearb Phys Anthropol. 2016;159(S61):37–78. https://dx.doi.org/10.1002/ajpa.22902

57. Marshall CM. Confidence intervals on stratigraphic ranges. Paleobiology. 1990;16:1–10. https://dx.doi.org/10.1017/S0094837300016766

58. Wagner PJ. Stratigraphic tests of cladistic hypotheses. Paleobiology. 1995;21:153–178. https://dx.doi.org/10.1017/S009483730001318X

59. Foote M, Hunter JP, Janis CM, Sepkoski Jr J. Evolutionary and preservational constraints on origins of biologic groups: Divergence times of eutherian mammals. Science. 1999;283:1310–1314. https://dx.doi.org/10.1126/science.283.5406.1310

60. Kurtén B. Pleistocene mammals of Europe. Chicago, IL: Aldine Publishing Company; 1968.

61. Prothero DR. Species longevity in North American fossil mammals. Integr Zool. 2014;9:383–393. https://dx.doi.org/10.1111/1749-4877.12054

62. Hawkes J, De Ruiter DJ, Berger LR. Comment on "Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia." Science. 2015;348:1326. https://dx.doi.org/10.1126/science.aab0591

63. Elderedge N, Cracraft J. Phylogenetic patterns and the evolutionary process: Method and theory in comparative biology. New York: Columbia University Press; 1980.

64. Patterson C. Significance of fossils in determining evolutionary relationships. Annu Rev Ecol Syst. 1981;12:195–223. https://dx.doi.org/10.1146/annurev.es.12.110181.001211

65. Argue, D, Groves CP, Lee MSY, Jungers WL. The affinities of *Homo floresiensis* based on phylogenetic analyses of cranial, dental, and postcranial characters. J Hum Evol. 2017;107:107–133. https://doi.org/10.1016/j.jhevol.2017.02.006

66. Berger LR, Hawks J, Dirks PHGM, Elliott M, Roberts EM. *Homo naledi* and Pleistocene hominin evolution in subequatorial Africa. eLife 2017;6, e24234, 19 pages. http://dx.doi.org/10.7554/eLife.24234

67. Dirks PHGM, Roberts EM, Hilbert-Wolf H, Kramers JD, Hawks J, Dosseto A, et al. The age of *Homo naledi* and associated sediments in the Rising Star Cave, South Africa. eLife 2017;6, e24231, 59 pages. http://dx.doi.org/10.7554/eLife.24231