Responsible metrics: what’s the state of the art?
Overview

• What are responsible metrics?
• Why should we care?
• How to implement a responsible metrics policy
• How to actually do metrics responsibly
• Who is responsible for responsible metrics?
• A call for research evaluation literacy
Responsible metrics is what?

- A movement that seeks to ensure that the use of metrics in the evaluation of research is done responsibly mitigating against perverse effects and unintended consequences.
Metrics can kill people

Imperial College professor Stefan Grimm ‘was given grant income target’

Emails with manager reveal details of review placed on academic found dead in September

December 3, 2014

By Chris Parr
Twitter: @ChrisParrTHE

A researcher at Imperial College London who was found dead in September had been told he was “struggling to fulfil the metrics” of a professorial post at the institution.
Why the audit culture made me quit

When Liz Morrish opened up to students about the pressures academics are under, disciplinary proceedings culminated in her resignation. She reflects on why she chose to tackle the failings of the neoliberal academy from the outside

March 2, 2017

By Liz Morrish
Twitter: @lizmorrish

In the UK, much of the rush to management by metrics is in response to shifting government incentives and policy changes, which, fed through the mechanism of the research excellence framework, affect institutional priorities, reputations and funding levels. Many of these metrics are quite outside the control of academics. Nevertheless, they have been weaponised as tools of performance management, and the very nature of the scrutiny creates a hostile environment for academic freedom.
Academics ‘face higher mental health risk’ than other professions

Lack of job security, limited support from management and weight of work-related demands on time among risk factors

August 22, 2017

By Holly Else
Twitter: @HollyElse
Responsible metrics lead to better decisions

• Comparing SSH with STEM on citation counts…
• Comparing early & late-career academics on h-index…
• Judging anyone by their ResearchGate score…
• …just isn’t going to lead to a sensible decision, let alone a fair one.
Responsible metrics statements
Response to DORA

No. respondents

- Already signed or likely to sign DORA
- Actively considering DORA but no decision made
- Actively considered DORA and decided not to sign
Own responsible metrics principles

- Created or developing own set of principles
- Actively considering but no decision made
- Actively considered and decided against

No. respondents

2015 2016 2017 2018 2019
Inspiration for RM statements

Leiden Manifesto 2016: 50%
Another University's statement 2017: 40%
Metric Tide 2018: 30%
DORA 2019: 20%
How to implement a responsible metrics policy
The need to accept your policy is just the beginning

This Photo by Unknown Author is licensed under CC BY-NC
The need to consider the advise-police-judge spectrum
The need for ownership at senior level

Senior University Managers involved in developing responsible metrics statements

<table>
<thead>
<tr>
<th>Year</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>45%</td>
<td>40%</td>
<td>30%</td>
<td>35%</td>
<td>25%</td>
</tr>
</tbody>
</table>
The need to manage upwards

From a mailing list:

“…there’s a desire to have…a metric (and they are keen on just one) against which to evaluate the performance of our research…. I’d be very interested to hear anyone else’s experiences …in dealing with the expectations of senior managers with this sort of thing.”
How do you actually DO metrics responsibly?
Q: How has your policy affected your use of metrics in practice?

- Avoid ALL metrics
- Ban certain metrics
- Reduce use of metrics
- Use metrics in line with policy
- Use metrics alongside peer review
- Use metrics in context
- Develop new metrics
- Use a wider range of metrics
Introducing the SCOPE model

1. Start with your values
2. Context
3. Options
4. Probe
5. Evaluate
START with what you value
START with what you value

• Not with the data you have available
  – The Streetlight Effect
• Not what others value
• University autonomy: use it or lose it

“If my h-index is the answer, what is the question?”
The streetlight effect

The cartoon shows a streetlight and a police officer in a dark room. The police officer asks, "This is where you lost your wallet?" The man replies, "No, I lost it in the park. But this is where the light is."
START with what you value

• Not with the data you have available
  – The Streetlight Effect
• Not what others value
• University autonomy: use it or lose it

“If my h-index is the answer, what is the question?”
CONTEXT
Understand who & why you’re evaluating

Figure 1. Risks associated with metric use in various settings

- Low risk
- Medium risk
- High risk
Use of FWCI in measuring to understand

Use of FWCI to identify staff for redundancy...

So, these are the proposed criteria I (a historian, remember) wd have to meet to avoid being among the ~140 of whom ~65 will lose our jobs.

The University will consult with the Trade Unions on a set of criteria to reduce the group of around 627 academic posts ‘in scope’ to a group of around 140 posts that will subsequently be ‘at risk’ of redundancy. The loss of 65 posts will come from this ‘at risk’ pool. The criteria that we are proposing to apply to identify the ‘at risk’ pool are defined below. However, please note that these criteria are subject to consultation with the Trade Unions and therefore may change.

If staff meet or exceed one or more of the proposed criteria below, they will not be at risk.

- Research and other income in the four-year period from 1 August 2012 to 31 August 2016 of £400k, £300k, £200k or more respectively for staff in Grades 9, 8, 7/6; or
- Research awards from 1 August 2015 to 31 March 2017 of £225k, £150k, £75k or more respectively for staff in Grades 9, 8, 7/6; or
- A sum of Field-Weighted Citation Impact greater than 1.5
- Staff on a core, permanent teaching only, teaching focused or teaching scholarship contract

Where staff have had a significant period of absence from work (three months or more) due to maternity leave or sickness absence for example, we will consider the data in these cases and seek to mitigate any adverse impact attributable to the period of absence, e.g. by using a time period that is more relevant to the individual circumstances. Equally we will give consideration as to how the criteria may need to be adjusted in relation to staff who have a disability.
Do we need to evaluate at all?

• Huge growth in incentivising behaviour through measurement
• Campbell’s Law: “The way you measure me is the way I’ll behave”
• Measuring is not always the best way to incentivise behaviour
Open Science and its role in universities:

research. In order for these goals to be achieved, universities should align their assessment, reward and evaluation systems with Open Science developments.
The Hong Kong Manifesto for Assessing Researchers: Fostering Research Integrity

David Moher 1, Lex Bouter 2, Sabine Kleinert 3, Paul Glasziou 4, Mai Har Sham 5

1Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute; School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada; 2Department of Epidemiology and Biostatistics, Amsterdam University Medical Centers, location VUMc, and Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, The Netherlands; 3The Lancet, London Wall Office, London, UK; 4Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Qld, Australia; and 5School of Biomedical Sciences, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
OPTIONS
Options

• Is your measure a suitable proxy for what you’re measuring?
• Quantitative measures are for quantifiable things…
  – Citations, publications, money, students
• Qualitative measures for qualifiable things…
  – Quality, diversity, excellence, value
• Beware using quantitative indicators as a proxy for qualitative things
  – Citations ≠ quality
  – Ranking position ≠ excellence
Faculty/Student Ratio (20%)

Teaching quality is typically cited by students as the metric of highest importance to them when comparing institutions using a ranking. It is notoriously difficult to measure, but we have determined that measuring teacher/student ratios is the most effective proxy metric for teaching quality. It assesses the extent to which institutions are able to provide students with meaningful access to lecturers and tutors, and recognizes that a high number of faculty members per student will reduce the teaching burden on each individual academic.

Faculty/student Ratio constitutes 20 percent of an institution’s final score.
When is Peer Review the Gold Standard, and When is it Only Tin?

Published on October 22, 2016 by Tony Waters

A new gold standard of peer review is needed
metrics are rubbish but ... (far) people are worse

Alan Dix
University of Birmingham and Talis
http://alandix.com/ref2014/
PROBE
Probe for potential negative impacts

1. Who does this discriminate against?
2. How could this be gamed?
3. What might the perverse incentives and consequences be?
4. Do the benefits of measuring outweigh the cost of measuring?
5. Is evaluating research actually going to make it any better?
REF 2014 cost almost £250 million

Accountability review finds cost of assessment equates to 2.4 per cent of funding bodies’ expected spend over next six years

July 13, 2015

By Holly Else
Twitter: @HollyElse
Does the cost outweigh the benefit?

1 citation tool + 1 research metrics post = £90,000 p.a.

1 “FAR” tool + 1 Research Impact Librarian = $175,000 pa
Probe for potential negative impacts

1. Who does this discriminate against?
2. How could this be gamed?
3. What might the perverse incentives and consequences be?
4. Do the benefits of measuring outweigh the cost of measuring?
5. Is evaluating research actually going to make it any better?
You don’t fatten a pig by weighing it
EVALUATE your evaluation…
That’s how we can evaluate responsibly. But who’s ‘we’? Who’s responsible for responsible metrics?
The research evaluation food chain

- Researchers
- Universities
- Funders
- Governments
- World rankings

Data vendors
“nothing will do more to foster change in accordance with the principles set out in this report than concerted work and institutional change in the area of rewards and incentives”
Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations

Erin C. McKiernan¹,*, Lesley A. Schimanski², Carol Muñoz Nieves², Lisa Matthias³, Meredith T. Niles⁴, and Juan Pablo Alperin²,⁵**

¹Departamento de Física, Facultad de Ciencias, Universidad Nacional Autónoma de México
²Scholarly Communications Lab, Simon Fraser University
³John F. Kennedy Institute, Freie Universität Berlin
⁴Department of Nutrition and Food Sciences, Food Systems Program, University of Vermont
⁵School of Publishing, Simon Fraser University

*Corresponding author: emckiernan@ciencias.unam.mx
**Corresponding author: juan@alperin.ca

Abstract

The Journal Impact Factor (JIF) was originally designed to aid libraries in deciding which journals to index and purchase for their collections. Over the past few decades, however, it has become a relied upon metric used to evaluate research articles based on journal rank. Surveyed faculty often report feeling pressure to publish in journals with high JIFs and mention reliance on the JIF as one problem with current academic evaluation systems. While faculty reports are useful, information is lacking on how often and in what ways the JIF is currently used for review, promotion, and tenure (RPT). We therefore collected and analyzed RPT documents from a representative sample of 129 universities from the United States and Canada and 381 of their academic units. We found that 40% of doctoral, research-intensive (R-type) institutions and 18% of master's, or comprehensive (M-type) institutions explicitly mentioned the JIF, or closely related terms, in their RPT documents. Undergraduate, or baccalaureate (B-type) institutions did not mention it at all. A detailed reading of these documents suggests that institutions may also be using a variety of terms to indirectly refer to the JIF. Our qualitative analysis shows that 87% of the institutions that mentioned the JIF supported the metric's use...
Funders – Plan S

03 In cases where high-quality Open Access journals or platforms do not yet exist, the Funders will, in a coordinated way, provide incentives to establish and support them when appropriate; support will also be provided for Open Access infrastructures where necessary;

04 Where applicable, Open Access publication fees are covered by the Funders or research institutions, not by individual researchers; it is acknowledged that all researchers should be able to publish their work Open Access;

05 The Funders support the diversity of business models for Open Access journals and platforms. When Open Access publication fees are applied, they must be commensurate with the publication services delivered and the

08 The Funders do not support the ‘hybrid’ model of publishing. However, as a transitional pathway towards full Open Access within a clearly defined timeframe, and only as part of transformative arrangements, Funders may contribute to financially supporting such arrangements;

09 The Funders will monitor compliance and sanction non-compliant beneficiaries/grantees;

10 The Funders commit that when assessing research outputs during funding decisions they will value the intrinsic merit of the work and not consider the publication channel, its impact factor (or other journal metrics), or the publisher.
Funders

What is the level of understanding of bibliometrics within the panels?

I am not at all an expert in bibliometrics; I just have a general idea of what it is.

Half of the interviewees reported previously sitting on selection panels that used bibliometrics (some of these were in panel settings beyond NIHR). One panel member noted that the experience gained through selection panels over a number of years, had improved their understanding of bibliometrics data and consequently they were able to “make better use of it” during the selection process. The remaining interviewees had never encountered the use of bibliometrics on selection panels prior to being involved with these competitions. These panel members described their understanding of bibliometrics as “rudimentary”, “cursory”, and “limited”. These members recognised that, at best, their understanding of bibliometrics was at a basic level, and certainly not at the detailed statistical level. Some of them were unsure about the details of the normalisation procedure and the comparability of applicants across different research fields. Overall, levels of expertise varied considerably, so some form of introduction or briefing is required to make sure that the information is accessible and useful to all panel members.

Wellcome-funded organisations must sign or publicly commit to the San Francisco Declaration on Research Assessment (DORA), or an equivalent. We may ask organisations to show that they’re complying with this as part of our organisation audits. This is a new requirement to encourage organisations to consider the intrinsic merit of the work when making promotion and tenure decisions, not just the title of the journal or publisher.
Governments
Indonesia’s scientists voice concerns about the country’s researcher ranking system

Critics flag unclear methodology, lack of credit for research contributions other than publications

by Dalmeet Singh Chawla
DECEMBER 31, 2018

Critics say the methodology and reasoning behind the metric, known as the Science and Technology Index (SINTA), are unclear. SINTA takes into account the number of journal and non-journal articles indexed in the database Scopus, the number of citations these documents accumulate in Scopus and Google Scholar, and researchers’ h-index. The h-index is another controversial metric that is designed to measure researchers’ productivity and the impact of their publications.
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Scopus H-Index</th>
<th>Google H-Index</th>
<th>Since</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUHARYO SUMOWIDAGDO</td>
<td>Lembaga Ilmu Pengetahuan Indonesia</td>
<td>91</td>
<td>134</td>
<td>2017</td>
</tr>
<tr>
<td>INDAH SUCI WIDYAHENING</td>
<td>Universitas Indonesia</td>
<td>7</td>
<td>11</td>
<td>2017</td>
</tr>
<tr>
<td>RIYANARTO SARNO</td>
<td>Institut Teknologi Sepuluh Nopember</td>
<td>17</td>
<td>20</td>
<td>2017</td>
</tr>
<tr>
<td>IGEDE WENTEN</td>
<td>Institut Teknologi Bandung</td>
<td>21</td>
<td>30</td>
<td>2017</td>
</tr>
<tr>
<td>ANWAR MALLONGI</td>
<td>Universitas Hasanuddin</td>
<td>12</td>
<td>11</td>
<td>2017</td>
</tr>
<tr>
<td>HANUNG ADI NUGROHO</td>
<td>Universitas Gadjah Mada</td>
<td>10</td>
<td>13</td>
<td>2017</td>
</tr>
<tr>
<td>ACHMAD NIZAR HIDAYANTO</td>
<td>Universitas Indonesia</td>
<td>8</td>
<td>12</td>
<td>2017</td>
</tr>
<tr>
<td>MAURIDHI HERY PURNOMO</td>
<td>Institut Teknologi Sepuluh Nopember</td>
<td>4</td>
<td>17</td>
<td>2017</td>
</tr>
<tr>
<td>MOHAMMAD BASYUNI</td>
<td>Universitas Sumatera Utara</td>
<td>14</td>
<td>16</td>
<td>2017</td>
</tr>
<tr>
<td>TOLE SUTIKNO</td>
<td>Universitas Ahmad Dahlan</td>
<td>13</td>
<td>19</td>
<td>2017</td>
</tr>
</tbody>
</table>
### Indicators and Weights for ARWU

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Indicator</th>
<th>Code</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Education</td>
<td>Alumni of an institution winning Nobel Prizes and Fields Medals</td>
<td>Alumni</td>
<td>10%</td>
</tr>
<tr>
<td>Quality of Faculty</td>
<td>Staff of an institution winning Nobel Prizes and Fields Medals</td>
<td>Award</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Highly cited researchers in 21 broad subject categories</td>
<td>HICl</td>
<td>20%</td>
</tr>
<tr>
<td>Research Output</td>
<td>Papers published in Nature and Science*</td>
<td>N&amp;S</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Papers indexed in Science Citation Index-expanded and Social Science Citation Index</td>
<td>PUB</td>
<td>20%</td>
</tr>
<tr>
<td>Per Capita Performance</td>
<td>Per capita academic performance of an institution</td>
<td>PCP</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

* For institutions specialized in humanities and social sciences such as London School of Economics, N&S is not considered, and the weight of N&S is relocated to other indicators.
Data vendors

Cox, Brian E.

University of Manchester ... Show all affiliations | View this Researcher in Scopus | Why do the metrics look different to those in Scopus?

2014 to >2018 no subject area filter selected

Summary Topics Collaboration Published Viewed Cited Economic Impact

Overall research performance

Scholarly Output
36
View list of publications

Field-Weighted Citation Impact
3.52

Citation Count
966

Citations per Publication
26.8

h-index
70

h5-index
33
The need for Evaluation Literacy
Responsible metrics requires responsible people

- Robust
- Humble
- Transparent
- Diverse
- Reflexive
Universities

29%

The proportion of Librarians surveyed whose LIS degree included bibliometrics
## Bibliometric Competencies

### Competency model for bibliometric work (version 1.1; July 2017)


<table>
<thead>
<tr>
<th>Entry Level</th>
<th>Core</th>
<th>Specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Applications</strong></td>
<td>I. Uses bibliometric knowledge to recommend where to publish and what to read; to increase academic staff bibliometric literacy; to support annual reporting by academic departments; to support grant capture; and to guide library collection and evaluate repository coverage.</td>
<td>I. Uses bibliometric knowledge to evaluate departmental/research centre performance; to evaluate institutional performance; and to support academic bibliometric research.</td>
</tr>
<tr>
<td>I. Explains effectively the concept, potential uses and limitations of bibliometrics to a range of stakeholders.</td>
<td>I. Advises on which are the appropriate tools to calculate a particular metric and explains differences in results between metrics produced by different tools.</td>
<td>I. Monitors national policy changes around research evaluation and advises on institutional responses.</td>
</tr>
<tr>
<td>II. Explains the concept, potential uses and limitations of altmetrics to stakeholders.</td>
<td>II. Explains responsible use as a general set of principles, and applies these principles to specific requests/cases. For example, advises on the applicability of metrics and tools to particular disciplines.</td>
<td>II. Advises on decisions about how the institution should use specific tools and on decisions about institutional Key Performance Indicators.</td>
</tr>
<tr>
<td>III. Explains author identifiers, such as ORCID, and promotes their use.</td>
<td>III. Participates in key debates about how research quality should be measured.</td>
<td>III. Advises on decisions about what a responsible use policy should be.</td>
</tr>
</tbody>
</table>

---

Loughborough University

#InspiringWinners since 1909
Statistics for responsible bibliometrics

Responsible Bibliometrics Analysis, Interpretation, and Reporting – A Statistics Approach

Presenters
Mike Thelwall
Professor of Data Science, University of Wolverhampton

Dr Abigail McBirnie
UK-based Information Specialist

Registration
11th October 2019 9:30 am

Finish
11th October 2019 4:00 pm

Venue
Room Met 2, The Met Hotel Leeds, King Street, Leeds, UK

Phone
0131 357 9854
Meaningful, responsible & effective research evaluation

Dr Elizabeth Gadd
Workpackage 1: Rating the rankers

• What if Rankers are no longer at the top of the food chain?

• Rate the rankers criteria: [https://inorms.net/activities/research-evaluation-working-group/](https://inorms.net/activities/research-evaluation-working-group/)

• Key themes:
  – Responsibility
  – Transparency
  – Measuring what matters
  – Rigour
Workpackage 2: Briefing materials for senior managers

• Set of powerpoint slides with notes to brief senior leaders on responsible research evaluation
• Based on SCOPE model
• Adapted to different settings (CC-BY)
• Translated into many different languages
Join the conversation!

• Start a conversation about responsible metrics in your own setting
• Start with what you/your institution values
• Join the broader conversation:
  – INORMS-RES-EVAL@jiscmail.ac.uk discussion list
  – Lis-Bibliometrics@jiscmail.ac.uk discussion list
Better to light a candle than curse the darkness.
Thank you for listening

Dr Elizabeth Gadd
Research Policy Manager (Publications)
Loughborough University

Skype: lizziegadd
Twitter: @lizziegadd
Email: e.a.gadd@lboro.ac.uk

http://orcid.org/0000-0003-4509-7785
http://about.me/elizabeth.gadd