Base notions in metrics-based research assessment

Henk F. Moed
Lecture at Conference on Teaching and Research Evaluation in Europe
Rome, 4-5 Dec 2014
## Short CV Henk F. Moed

<table>
<thead>
<tr>
<th>Year Range</th>
<th>Position/Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981-2009</td>
<td>Staff member at Centre for Science and Technology Studies (CWTS), Leiden Univ.</td>
</tr>
<tr>
<td>2009</td>
<td><strong>Professor</strong> of Research Assessment Methodologies at <strong>Leiden University</strong></td>
</tr>
<tr>
<td>Sept 2012 – 1 Nov 2014</td>
<td><strong>Elsevier</strong>, AGRM Dept; Head of <strong>Informetric Research Group</strong></td>
</tr>
<tr>
<td>Sept-Oct 2014</td>
<td><strong>Visiting professor</strong> at Univ Rome la Sapienza (Scuola Superiore di Studi Avanzati)</td>
</tr>
</tbody>
</table>
Securing a political basis for academic research

- General politics
- Research policy & management
- Research metrics Development

Define targets of assessment and funding

Develop assessment tools and funding formula
1. Research assessment requires the optimal combination of metrics and peer review

- Peer committees may use citation analysis for *initial rankings*
- Peer review may be used to assess cases in which metrics indicate *anomalies*
2. Design of an assessment process very much depends upon the assessment context

- Multi-dimensional Research Assessment Matrix (EC AUBR Expert Group, 2010)
<table>
<thead>
<tr>
<th>Unit of assessment</th>
<th>Objective</th>
<th>Output dimensions</th>
<th>Bibliometric indicators</th>
<th>Other indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>Allocate resources</td>
<td>Research productivity</td>
<td>Publications</td>
<td>Peer review</td>
</tr>
<tr>
<td>Research group</td>
<td>Improve performance</td>
<td>Quality, scholarly impact</td>
<td>Journal citation impact</td>
<td>Patents, licences, spin offs</td>
</tr>
<tr>
<td>Department</td>
<td>Increase multi-discipl. research</td>
<td>Innovation and social benefit</td>
<td>Actual citation impact</td>
<td>Invitations for conferences</td>
</tr>
<tr>
<td>Institution</td>
<td>Increase regional engagement</td>
<td>Sustainability &amp; Scale</td>
<td>Internat. co-authorship</td>
<td>External research income</td>
</tr>
<tr>
<td>Research field</td>
<td>Promotion, hiring</td>
<td>Research infrastruct.</td>
<td>citation ‘prestige’</td>
<td>PhD completion rates</td>
</tr>
<tr>
<td>Unit of assessment</td>
<td>Objective</td>
<td>Output dimensions</td>
<td>Bibliometric indicators</td>
<td>Other indicators</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Individual</td>
<td>Allocate resources</td>
<td>Research productivity</td>
<td>Publications</td>
<td>Peer review</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality, scholarly impact</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patent, licence, spin off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research group</td>
<td>Improve performance</td>
<td>Innovation</td>
<td>Patent, licence, spin off</td>
<td>Invitations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality, scholarly impact</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patent, licence, spin off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department</td>
<td>Increase multi-discipl. research</td>
<td>Innovation</td>
<td>Patent, licence, spin off</td>
<td>Invitations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality, scholarly impact</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patent, licence, spin off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Increase regional engagement</td>
<td>Sustainability</td>
<td>Internat. authorship</td>
<td>External research income</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research field</td>
<td>Promotion, hiring</td>
<td>Research infrastruct.</td>
<td>‘prestige’</td>
<td>PhD completion rates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The choice of assessment methodology and indicators depends upon:

1. Which type of units are to be assessed?
2. Which aspects are being assessed?
3. What are the assessment’s objectives?
4. What is the state of the system under assessment? (Meta-analysis)
Meta assumptions

Typcial examples

Substantial part of professors in this country is not research-active (e.g., too much engaged in teaching)

Through self-selection applicants are research active; their quality level tend to be high

Researchers in this country are not well integrated into the international community; publish mainly in national journals

Young research groups have no good chances to develop in this funding system

Decisions on hiring tend to be made on the basis of political considerations
Wider issues

• Change an assessment method every 5-10 years?

• Focus on top or on bottom of quality distribution?

• What is an acceptable “error rate”?

• Wrong in individual cases ↔ beneficiary for the system as a whole
3. Targeted performance-based funding is feasible without large scale assessment

- Performance based funding may focus on emerging research groups
- **Funding allocated to institutions (partly) based on their number of emerging groups**
Identification of emerging groups (5 yrs)

**Features**
- Young senior
- Successful PhD and post-doc stage
- Expanding competitive funding
- Increasing output and impact
- Growing prestige

**Indicators**
- Age
- Status post-doc institute citations to PhD oeuvre
- Trend in competitive funding
- Trend in citation rate and in diversity journal packet
- Speaker invitations; internat. collaboration
4. Journal impact factors are popular but are often misinterpreted

- TR/JCR rankings of journals by impact factor and subject category are widely available
- Differences and correlations with qualitative measures are often misunderstood
Journal impact factors are no surrogates of actual citation rates of individual articles
Normal vs. skewed distributions

**Boys (Mean length=95 cm)**

**Players (Mean length=185 cm)**

**PAMS (JIF=0.43)**

**TAMS (JIF=0.85)**
<table>
<thead>
<tr>
<th>Question</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>a randomly selected boy is at least as tall as a randomly selected adult?</td>
<td>Almost zero</td>
</tr>
<tr>
<td>Av. Length: Boys 85 cm; Adults: 185 cm</td>
<td></td>
</tr>
<tr>
<td>a randomly selected PAMS paper is cited at least as often as a randomly selected TAMS paper?</td>
<td>62 %</td>
</tr>
<tr>
<td>JIF: PAMS: 0.43; TAMS: 0.85</td>
<td></td>
</tr>
</tbody>
</table>
5. The role of journal impact factors in research and journal assessment must be reduced

• The DORA Declaration makes sense: their role in promoting journals must be reduced

• The genuine alternative is use of metrics of the quality of journals’ peer review system
A positive correlation was found between a journal’s impact factor and its manuscript rejection rate.

This outcome does NOT sufficiently justify the use of impact factors in the assessment of individuals or groups.
Linear / rank correlations - Case 1

Pearson R = 0.0
Spearman Rho = 0.0
Linear / rank correlations - Case 2

Pearson R = 0.85*

Spearman Rho = 0.39
Linear / rank correlations - Case 3

Pearson R = 0.94*
Spearman Rho = 0.76*
Use of journal impact factors: The problem

Discourage publication in national jrnls

DORA Critique

\[
X_1 > X_2 \quad \text{?} \quad Y_1 > Y_2
\]

YES!

\[
X_1 > X_2 \quad \text{?} \quad Y_1 > Y_2
\]

NO!
6. Indicators have not yet fully profited from the computerization of the research process

- A challenge: The *practical realization* of the *ethos of science* and scholarship in a *digital* age
- Notions of *founding fathers* in bibliometrics can be implemented in *new online tools*
Computerization of the research process

Information processing
- Searching, browsing, reading behavior
  - Usage data (e.g. downloads) patterns

Communication & organization
- Online tools amplifying collective intelligence
  - Use of blogs, wikis, issue trackers

Research assessment
- Scholars’ interaction with research outputs
  - Mentions in social media & ref managers

Research data & methodologies
- Big data; crowdsourcing; analysis software
  - Analysis top cited articles by subject field
## Comp Sci-related top-cited articles in Scopus [Halevi, Research Trends, 2014]

<table>
<thead>
<tr>
<th># Cites</th>
<th>Discipline</th>
<th>Article Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>17,171</td>
<td>Agr &amp; Biol Sci, Mol Biol; Medicine</td>
<td>MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (2007)</td>
</tr>
<tr>
<td>4,335</td>
<td>Business, management, social sciences</td>
<td>User acceptance of information technology: Toward a unified view (2003)</td>
</tr>
<tr>
<td>5,325</td>
<td>Chemistry</td>
<td>UCSF Chimera - A visualization system for exploratory research and analysis (2004)</td>
</tr>
<tr>
<td>1,335</td>
<td>Energy</td>
<td>Geant4 developments and applications (2006) [software for simulating passage of particles through matter]</td>
</tr>
<tr>
<td>7,784</td>
<td>Engineering; Math</td>
<td>A fast and elitist multi-objective genetic algorithm: NSGA-II (2002)</td>
</tr>
</tbody>
</table>
7. Online bibliometric tool for self-assessment is needed based on flexible benchmarking

• An **advanced** online tool for bibliometric **self-assessment** is technically feasible....

• ...combining Garfield’s **citation-based** algorithm and Merton’s notion of a **reference group**
1. Select, **verify**, add, correct, finalize the **list of publications** of the assessed researcher (AR) (SET 1)

2. Select a set (SET 2) of articles that are **similar** to those in SET1 in terms of **subject field**

3. Identify **the authors** in SET2 and **all** their articles (SET 3)

4. Select the authors in SET3 most similar to the AR in terms of **subject field** and **academic age** (SET 4)

5. Let the user further select the authors in SET 4 she wants to include in her tailormade benchmark set (SET 5)

6. Calculate **quartile**-based indicators (e.g., in top 25 %) for the AR using the data in the author benchmark SET 5
8. New metrics and online tools of the manuscript peer review process are needed

- Distinguish **two** dimensions:
- Applying **base-line** quality criteria – facilitating manuscript improvement
- Focusing on **potentially groundbreaking** findings – facilitating scientific debate
Metrics of the manuscript peer review process as tools to measure and improve its quality

• **Text mining** of correspondences between editors, referees and authors

• **Make explicit** a journal’s quality standards

• Operationalize the concept of **rigorousness** of the peer review process
9. One should not focus merely on output of research, but also on its input and process

- Econometric efficiency measures relate output to input
- Science of Team Science (SciTS) studies a.o. conditions for successful collaboration
10. Metrics development should be independent and enable wide access

- Metrics development must apply rigorous scientific quality standards.
- Independent of data providers, funders and politicians
- Metrics availability should not be controlled by business or political interests
Thank you for your attention