Better decision making through responsible research evaluation

Research Administration Briefing Material
Why is responsible evaluation important?

- Ensure staff well-being and manage reputational risk
- Make better decisions
- Ensure return on investment
- Establish operational readiness
- Maintain institutional autonomy
Before starting any evaluation, expertise is needed

It’s surprising how academic rigour is so often abandoned when it comes to management practice.

► Remember that producing and interpreting metrics requires expertise.

► Understand and accept the limitations of both the indicators and the data used as basis for analyses.

► Make sure that metrics only support expert judgement, never replace it.

► Understand that using metrics has consequences: you get what you measure.
SCOPE: 5 stages for doing evaluation responsibly

START with what you value

EVALUATE your evaluation

PROBE deeply

CONTEXT considerations

OPTIONS for evaluating

VALUATE your evaluation
S Start with what you value

E Evaluate your evaluation

C Context Considerations

P Probe deeply

O Options for evaluating
Do we value what is important for us or for others? Or both? Or none?

“What are they worth? Well, it’s difficult to put a price on such extraordinary works of art. How much you got?”
HEIs sometimes focus on what OTHERS value (i.e. rankings, national assessment systems, etc.) rather than what THEY value.

Institutional autonomy is important. Use it or lose it.

If introducing performance indicators, what sort of ‘performance’ are you looking for?

- How do you prioritize your values?

Inherent tensions between what you value and staying alive (compliance)

- Where those tensions exist, ask yourself what you can do about them.

So often we value what we think we can measure rather than measure what we value.
The Streetlight effect:
Measuring by available data not by mission
Start with what you value

Context Considerations

Evaluate your evaluation

Probe deeply

Options for evaluating
Why are you evaluating?

- **Measure to analyse.** “Science of science” activities that study phenomenon for the sole purpose of understanding them better.

- **Measure to advocate.** “Pick me!” activities. The use of metrics to market an individual, group or university on promotional materials or grant applications.

- **Measure for accountability.** Plotting progress against an objective whether internally or externally set.

- **Measure to acclaim.** The use of indicators to compare one entity with another.

- **Measure to adapt.** The use of indicators to incentivise certain behaviours.

- **Measure to allocate.** Any activity that results in some kind of reward for the entity being measured, be this a job, promotion, grant, prize or award of any description.
Why and what are you measuring? Balancing the risks.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Country</th>
<th>HEI</th>
<th>Group</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>To understand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advocacy</td>
<td>To show off</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accountability</td>
<td>To monitor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acclaim</td>
<td>To benchmark</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptation</td>
<td>To incentivise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allocation</td>
<td>To reward</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Risks associated with evaluation activities in various settings.
The disparity of scholarly disciplines

- Differences in scientific disciplines, and in the size or structure of units of assessment have to be considered → all indicators are not suitable for all disciplines
- Metrics are neither objective nor neutral

What works in one context won’t work in another
Do you need to evaluate at all?

- “You don’t fatten a pig by weighing it”
- Measuring is not the only way to incentivise behaviour
- Changing behaviour requires:
  - Understanding - Why do I need to do this? (Messaging)
  - Capability - How do I do this? (Training and support)
  - Opportunity - Where can I do this? (Tools, services, making it easier to do it than not to do it)
  - Motivation - What will happen if I do/don’t? (Reward/sanctions)
Start with what you value

Evaluate your evaluation

Probe deeply

Context Considerations

Options for evaluating
Rule of thumb

- Is your measure a suitable proxy for what you are measuring?
- Quantitative measures are for quantifiable things...
  - Citations, publications, money, students
- Qualitative measures for qualifiable things...
  - Quality, excellence, value
- Be careful if 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.
Quantitative measures

- Always imperfect
- Always looking back
- Made better by:
  - Checking validity (does this measure what you value)
  - Using a ‘basket’ of indicators
  - Combining with qualitative assessment
Unpacking some common metrics

- **Journal Impact Factor**
  - Mean cites-per-paper for a journal over the past two years
  - Problems:
    - Citation data is always skewed - should never use the mean
    - Often used as a metric for researchers or papers. It’s a journal metric.

- **H-index**
  - Number of papers (n) with at least n citations
  - Problems:
    - Disadvantages early career researchers, and those with non-standard career paths
    - Only ever increases. (Often called the ‘Age-Index’)
    - Correlates with total citations.
Qualitative measures (peer review)

- Always imperfect
- Made better by:
  - Ensuring appropriate expertise (there is peer review and EXPERT peer review)
  - Unconscious bias training
  - Double-blind OR Open review
  - More than one reviewer
  - Diverse reviewers
  - Triangulate with metrics where appropriate
Evaluate WITH the evaluated

- Engage with communities under evaluation
  - Understand what the unit of assessment values, what are their aims
- Co-produce evaluative approaches where possible
  - Consider the scientific fields under evaluation to choose relevant indicators
- Jointly interpret the results
  - Openness and transparency increases the legitimacy of evaluation results
Start with what you value

Context Considerations

Evaluate your evaluation

Options for evaluating

Probe deeply
P - Probe deeply

- Who does this discriminate against?
  - E.g. early-career researchers, females, non-journal based disciplines
  - Adjust accordingly

- How might this be gamed?
  - Where there is a prize there is a game - people will optimize

- What might the unintended consequences be?
  - On institutional or individual level

- Does the cost of measuring outweigh the benefit?
  - The cost, including the workload should be proportional to the aims and anticipated outcomes of evaluation
Indonesia recently introduced a new researcher ranking system that uses the volume of articles and citations from GS and Scopus for deciding who are the BEST researchers. And for allocating funding to these best researchers...
... number of women = 1
... number of researchers in social sciences and humanities = 0

Are we really confident that these are the best researchers in the country? Are we really sure that funding those who do well on this metrics is going to be good for science?
Unintended consequences: institutional level

- Poor decision-making
- Neglecting all activities that aren’t measured
- Goal displacement, mirroring e.g. government level assessment practice
- Short termism
- Transactional cost of metrics which may not actually add any value to the organisation
- Discourage initiative, innovation and risk-taking
  - Negative influence on untraditional ways of doing research, such as interdisciplinarity
Unintended consequences: individual level

- Academic burden - academics leaving the profession - or worse
- Focusing on research areas that are more visible in e.g. publication databases
- Narrowing the types of publication according to assessment criteria
- Focusing on tasks that are favourable in regards to assessments
- Aiming at journals with high impact - research results are published later rather than sooner, the editorial system gets clogged...
Start with what you value

Context Considerations

Options for evaluating

Probe deeply

Evaluate your evaluation
E - Evaluate your evaluation

- Go back to S - did the evaluation approach bring new insights to what you value?
- Be open to adjustments, it is always possible to do the evaluation better.
- The area of evaluation (range of data sources and indicators available as well as institutional missions and strategies) is subject to constant change
  - Just because an evaluation approach worked previously, does not mean it will work forever
  - Building in a regular review of evaluation approaches and doing so with units under evaluation is an essential part of the evaluation process
SCOPE: 5 stages for doing evaluation responsibly

1. START with what you value
2. EVALUATE your evaluation
3. PROBE deeply
4. OPTIONS for evaluating
5. CONTEXT considerations

Evaluate your evaluation deeply, considering what you value. Use options for evaluating carefully within the context of your evaluation.
INORMS Research Evaluation Working Group web pages:
https://inorms.net/activities/research-evaluation-working-group/
Subscribe to the INORMS-RES-EVAL discussion list:
https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=INORMS-RES-EVAL