Benchmarking research performance

Anne-Wil Harzing
Middlesex University
www.harzing.com

Twitter: @AWharzing
Blog: http://www.harzing.com/blog/
Email: anne@harzing.com
Quick Intro:
Anne-Wil Harzing

• My name?...., Yes Anne-Wil is one name and not part of my family name
• Started at Middlesex in September 2014
  • previously in Melbourne (PhD director 2004-2009, Associate Dean RHD, 2009-2010, Associate Dean Research, 2010-2013)
  • 1991-2001: Bradford (UK), Maastricht, Tilburg & Heerlen (Netherlands)
• Productive and passionate researcher & research mentor
  • 79 international journal articles since 1995 (160+ publications in total)
  • >11,000 Google Scholar citations, h-index 49, ISI citations: >4,000, top 1% most cited world-wide in Economics & Business
• Service to the academic community
  • Editorial board membership of a dozen journals
  • Personal website since 1999, 1000-1500 visitors/day, many free resources
    • Journal Quality List since 2000, 57th edition
    • Publish or Perish since 2006, version 5 launched late October 2016
Presentation overview

• Two key challenges for a Research Dean:
  • Managing “Up” and Managing “Down”
  • How can benchmarking help and give you legitimacy?

• Benchmarking: not a simple choice
  • 5 key decisions: what to benchmark, which measures, which data source, which level, who to benchmark against?
  • More information, readings & resources

• [Only if time]: Metrics vs. peer review
The two key challenges for a Research Dean

- Managing “up”:
  - Keep your Dean with his/her feet on the ground
  - Make your Pro or Deputy Vice Chancellor Research aware of disciplinary differences

- Managing “down”:
  - Motivating individual academics to:
    - Improve their research performance
    - Engage in new initiatives/media
Managing up: Benchmarking provides a frame of reference

- More convincing and face-saving than just telling senior management
  - This can’t be done!
  - You really have no clue don’t you?
- Dean: “all (Associate) Profs should be PI on an ARC grant”
  - Ok, that means our Faculty alone would get double the number of ARC grants that are allocated for the whole of Australia (nearly 50 universities) last year
  - We are already #1 in Australia
- Dean: “every academic should publish at least one SAE A* pub/year”
  - Ok, but the top-20 academics in B&M worldwide on average only publish only 1.5 SAE ISI-listed (not necessarily A*) publication a year
- DVC: “your faculty should double its research output”
  - Ok, that means that we would leave Harvard, Wharton, Stanford, and MIT a long way behind us
  - We are already #18 world-wide in terms of ISI-listed pubs in the last 5 years
Use visuals:
Research funding & productivity

Source: Own creation; simplified, stylised visualisation based on anecdotal experience
Use visuals:
RHD supervision & productivity

Source: Own creation; simplified, stylised visualisation based on anecdotal experience
Managing Down: Benchmarking provides a frame of reference

- More convincing and face-saving than just telling academics
  - You should really do this!
  - You are just making up excuses for your poor performance
- Good way to counter prejudices
  - Some Australian academics *do* publish in 4*/A*/North American journals, you are not necessarily a victim of discrimination
  - A* publication doesn’t always mean high citation counts
  - High quantity ≠ Low quality, Low quantity ≠ High quality, you *can* be productive and do high quality work
- Provide evidence of how e.g. engagement in:
  - Social media (blogging, twitter) helps downloads and citations
  - The world outside academia can lead to research funding
Journal quality & citation impact
2008-2013 (Scopus Scival)

20 Professors in the same School, size circle # of publications:
Journal quality doesn’t “determine” citations
Number of pubs & citation impact
2008-2013 (Scopus Scival)

20 Professors in the same Business School:
Quantity and quality can go hand in hand
Tweeting is for twits only?
No, and nor is blogging

- UCL experiment, for details see: http://blogs.lse.ac.uk/impactofsocialsciences/2012/04/19/blog-tweeting-papers-worth-it/
- Looks at papers in University repository
- Tweeting led to 10-50 fold increase in downloads
- Author had 7 out of 10 most downloaded papers and 1/3 of all downloads of her department
- She combined this with an active blog and was the only one active in social media
Benchmarking: not a simple choice

1. What to benchmark?
   - Publications, citations, societal impact, research funding, students

2. Which measures to use?
   - Journal rankings, ISI JIF, h-index & variants, total cites

3. Which data source to use?
   - Thomson ISI, Scopus, Google Scholar, Microsoft Academic

4. At what level to benchmark?
   - Individuals, Departments, Faculty/School, University

5. Who to benchmark against?
   - Similar entities, aspirational benchmarks?

➢ Most of these decisions are independent, though some are linked
1. What to benchmark?

- Research funding (input), publications (throughput), citations/societal impact (output), research students (input?, output?)
  - Research funding often seen as a target in itself; especially in the Social Sciences and Humanities this is inappropriate, it is an input measure

- Traditionally, the emphasis was on publications, but more and more rankings also (or even mainly) look at citations
  - Bibliometric researchers see citations per paper as compared to the field norm as the only valid measure of research performance for institutional or country comparisons
  - What is the use of publishing if your work is not cited at all? We are expected to be part of a scholarly conversation. Not publishing is similar to being mute. Publishing without citation means you are speaking, but nobody is listening.

- Citations only measure “academic” impact, we have a big responsibility in societal impact as well (Adler & Harzing, AMLE, 2009)
2. Which measures to use?

- For journals: wide range of journal rankings available
  - Most published rankings use ISI/Scopus-listed papers only; fortunately ISI/Scopus are expanding their coverage under pressure of alternative providers

- For impact/cites: h-index & variants, total # cites, citations per paper
  - H-index increasingly seen as a convenient summary of quantity & impact and used in many research assessments
    - H-index of 10 means 10 papers with at least 10 citations each
    - hla: h-index adjusted for co-authorships and career length, average number of single-author equivalent impactful publications an academic publishes a year (usually well below 1.0)
  - Total citations is probably the fairest way to assess impact for individuals, a strict focus on citations per paper might discourage people to publish additional papers

- All these metrics can be calculated in seconds with Publish or Perish
  - [http://www.harzing.com/resources/publish-or-perish](http://www.harzing.com/resources/publish-or-perish)
2. Calculate citation metrics with Publish or Perish
2. Metrics: H-index vs. hla

- hla adjusts h-index for
  - co-authorship (1.87 vs. 6.14)
  - career length (22 vs. 43 years)
3. Which data source (1)?: Commercial databases

- WoS and Scopus might not provide comprehensive paper/citation count for the Social Sciences
  - General search: Citations to books, chapters, working papers, reports, conference papers, journal articles published in non-ISI journals are not included
  - WoS Cited Reference Search
    - Does include citations to non-ISI publications. However, it only includes citations from journals that are ISI-listed
  - WoS: problems with non-Anglo names and limited coverage of non-English journals
  - Percentage of research output submitted in ISI listed journals (Australia, 2004)
    - 24% (Economics)
    - 11% (Management)
3. Which data source? (2):
Free databases

- Google Scholar has most comprehensive paper/citation count
  - Provides fairest comparison between sub-disciplines within the University
  - Only source that has fairly comprehensive book coverage
  - Includes lower quality citations, but vast majority are journal references
  - Cannot search by institution or country, no advanced bibliometric queries

- Microsoft Academic is a very promising compromise between WoS/Scopus and Google Scholar
  - Coverage is (much) better than WoS/Scopus
  - Provides cleaner results than Google Scholar
  - Data access is much better than GS (no restrictions)
  - Advanced bibliometric queries might be possible in the future with Publish or Perish
3. Data sources: difference between disciplines

Sample: 146 (Associate) Professors @ University of Melbourne in 37 disciplines
4. At what level?

- Individual level analysis
  - Most suitable for individual level decisions (e.g. promotion)
  - Probably provides too much detail (and creates too much work) for institutional comparisons
- Departmental level
  - Relevant if you need to have different policies for different departments
  - Difficult to implement if your department structures do not match those of other universities, e.g.
    - IB, Marketing free standing departments, vs. combined with Management
    - Accounting & Business Information Systems combined/separate
- Faculty/School level
  - Probably not the best option for most Business Schools
    - Finance & Accounting very different research traditions
    - Some schools include Economics which is different again
- University level
  - Not that relevant for Research Directors, but many international rankings are university-wide (and thus biased towards those with Medical Schools)
5. Who to benchmark against?

- Universities in the same “mission group”
  - Probably necessary, but not sufficient
  - I would suggest also benchmarking against international universities
- Pick universities who substantially increased their performance and who face similar challenges (i.e. generally non-NA universities)
  - Gather information about their research policies, web presence etc.
- WoS is still used as the data source by many international rankings, so using their Essential Science Indicators for a specific set of benchmarking institutions could give a good picture for papers and citations
- Scopus Scival is worth investigating as it allows for very flexible benchmarking, including both predefined groups (Go8, Russell Group), individual universities, disciplines, and custom-created groups
Papers vs Cites/Paper: National comparison

ISI Essential Science Indicators: Economics & Business
Papers vs Cites/Paper: International comparison

ISI Essential Science Indicators: Economics & Business
More information on citation analysis?

- [http://www.harzing.com/.category/publish_or_perish_tips](http://www.harzing.com/.category/publish_or_perish_tips)
Readings: Google Scholar and citation analysis

Readings: WoS, Microsoft Academic & new metrics/data

- Harzing (2017) Microsoft Academic: Is the Phoenix getting wings, accepted for Scientometrics
Resources

• For Research Deans
  • Publish or Perish
  • Journal Quality List
  • White papers (http://www.harzing.com/publications/white-papers)
  • Classic papers

• For Academics
  • Idem
  • Blog (Academia Behind the Scenes, Academic Etiquette, Publish or Perish tips)
  • For female academics in particular: http://www.harzing.com/.keyword/gender
METRICS VS PEER REVIEW

Anne-Wil Harzing
Increasing audit culture: Metrics vs. peer review

- Increasing “audit culture” in academia, where universities, departments and individuals are constantly monitored and ranked
- National research assessment exercises, such as the ERA (Australia) and the REF (UK), are becoming increasingly important
  - Publications in these national exercises are normally assessed by peer review for Humanities and Social Sciences
  - Citations metrics are used in the (Life) Sciences and Engineering as additional input for decision-making
- The argument for not using citation metrics in SSH is that coverage for these disciplines is deemed insufficient in WoS and Scopus
The danger of peer review? (1)

• Peer review might lead to harsher verdicts than bibliometric evidence, especially for disciplines that do not have unified paradigms, such as the Social Sciences and Humanities
  • In Australia (ERA 2010) the average rating for the Social Sciences was only about 60% of that of the (Life) Sciences
    • This is despite the fact that on a citations per paper basis Australia’s worldwide rank is similar in all disciplines
    • The low ERA-ranking led to widespread popular commentary that government funding for the Social Sciences should be reduced or removed altogether
  • Similarly negative assessment of the credibility of SSH can be found in the UK (and no doubt in many other countries)
The danger of peer review? (2)

• More generally, peer review might lead to what I have called “promise over proof”

• Assessment of publication might be (subconsciously) influenced by “promise” of:
  • the journal in which it is published,
  • the reputation of the author's affiliation,
  • the sub-discipline (theoretical/modeling vs. applied, hard vs. soft)

• [Promise] Publication in a triple-A journal means that 3-4 academics thought your paper was a worthwhile contribution to the field. But what if this paper is subsequently hardly ever cited?

• [Proof] Publication in a “C-journal” with 1,000+ citations means that 1,000 academics thought your paper was a worthwhile contribution to the field.
What can we do?

• Be critical about the increasing audit culture

• But: be realistic, unlikely to see a reversal of this trend. In order to “emancipate” the SSH, an inclusion of citation metrics might help. However, we need to:
  • Raise awareness about:
    • Alternative data sources for citation analysis that are more inclusive (e.g. including books, local and regional journals, reports, working papers)
    • Difficulty of comparing metrics across disciplines because of different publication and citation practices
  • Suggest alternative data sources and metrics
    • Google Scholar or Scopus instead of WoS/ISI
    • hla (Individual annualised h-index), i.e. h-index corrected for career length and # of co-authors
Conclusion

• Will the use of citation metrics disadvantage the Social Sciences and Humanities?
  • Not, if you use a database that includes publications important in those disciplines (e.g. books, national journals)
  • Not, if you correct for differences in co-authorships
• Is peer review better than metrics (in large scale research evaluation)?
  • Yes, in a way…. The ideal version of peer review (informed, dedicated, and unbiased experts) is better than a reductionist version of metrics (ISI h-index or citations)
  • However, the inclusive version of metrics (GS hIa or even Scopus hIa) is probably better than the likely reality of peer review (hurried semi-experts, influenced by journal outlet and affiliation)
• In research evaluation at any level use a combination of peer review and metrics wherever possible, but:
  • If reviewers are not experts, metrics might be a better alternative
  • If metrics are used, use an inclusive database (GS/MA) and career & discipline adjusted metrics