

THE UNITED KINGDOM'S APPROACH TO EVALUATING RESEARCH

*Thomson-Reuters Conference on National and
Institutional Perspectives on Metrics-based
Research Evaluation, Brisbane, 16/4/09*

Charles Oppenheim
Loughborough University
C.Oppenheim@lboro.ac.uk

MY CREDENTIALS

- Have undertaken research on the links between UK Research Assessment Exercise (RAE) results and bibliometrics since the mid 1990s
- (Token bibliometrician?) member of the Committee advising HEFCE on the use of bibliometrics in the forthcoming Research Excellence Framework (REF), and the pilot use of it to compare to 2008 RAE

ONE IMPORTANT QUESTION

- Is the point of research evaluation to evaluate past performance, predict future performance, or a way of working out how much money to dish out?
- The three are not identical, yet the RAE tries to be all three
- Evaluating past output using metrics or peer review (+ PhD completions, research income achieved, etc.) does the first; evaluating future research plans does the second

THE REF

- Announced by Gordon Brown in 2006 when he was Chancellor of the Exchequer (so it is clear that the motivation was cost-cutting)
- To be metrics **only** – details left to HEFCE et al to sort out
- HEFCE itself was evidently surprised by the announcement

THE REF

- HEFCE commissioned expert advice on the use of bibliometrics and consulted the community on key elements of the REF
- Large number of responses
- Consultation outcomes published on HEFCE website
- Significant modification announced in April 2008:
 - combination of metrics-based indicators, including bibliometrics where appropriate, as well as input from expert panels for all subjects

THE PILOT

- Trial run of the bibliometrics approach using RAE2008 data
- RAE outcomes were published a few months ago
- REF Pilot ongoing right now
- Purpose of pilot is to assess two things: which bibliometrics measures correlate best with actual RAE results? What are the administrative and technical burdens on the 22 participating Universities doing the pilot?
- Broad results will be published; participating Universities will get detailed results, to be retained for a short time period and only for the purpose of feeding back to HEFCE any errors or issues – must not be used for evaluating individuals or Departments – and they must destroy the data after a short period of time

THE REF PILOT IN PRACTICE

- Collect ALL papers written by staff submitted to 2008 RAE by selected HEIs in the selected subject areas
- Assign the papers to somewhere between 100 and 250 subject categories (two runs, one with the smaller and one with the larger number of subject categories)
- Calculate: average no. of citations per article; again, but ignoring top and bottom 25% results; % uncited
- Calculate: world average number of citations per article in chosen subject area over chosen time period
- Calculate: % of articles from the University that are above the world average
- N.B. Subject areas based on where a journal is assigned by Thomson-Reuters; ignore non-journal articles (for the Pilot only hard sciences and life sciences are being examined, so this is not too controversial). Where a journal title is not known to Thomson-Reuters, then look at the citations in the article being studied and based on the journal titles in those citations, decide what subject this article is about

FURTHER CALCULATIONS

- Do the same, BUT:
- Ignore all review articles (identified by algorithm)
- Add in/exclude papers published in any previous employment not in this HEI
- Exclude papers by Category C staff (mainly visiting Fellows or Professors and Honorary Professors)
- Restrict to the 6 papers with the highest number of citations
- Restrict to the four papers returned in the 2008 RAE

FINALLY

- See which of the combinations provides the best correlation with actual RAE results
- HEFCE will digest the results and will then probably follow the best combination in running the real REF
- But it may well be that for different subjects, different bibliometric measures correlate best – then what?
- Most unlikely that there will be NO correlation – early results confirm this

REF: overview

- REF is a unified framework for funding and assessing research
- Its primary focus is still excellence, but will take more explicit account of the impact of research on the economy and society

REF: overview (2)

- Following previous consultations, it has already moved some way from where it started
- REF is not only about basic research, nor is it only about bibliometrics
- It involves a range of assessment approaches which can be combined in different ways to fit particular subject fields and types of research

Tools for assessment

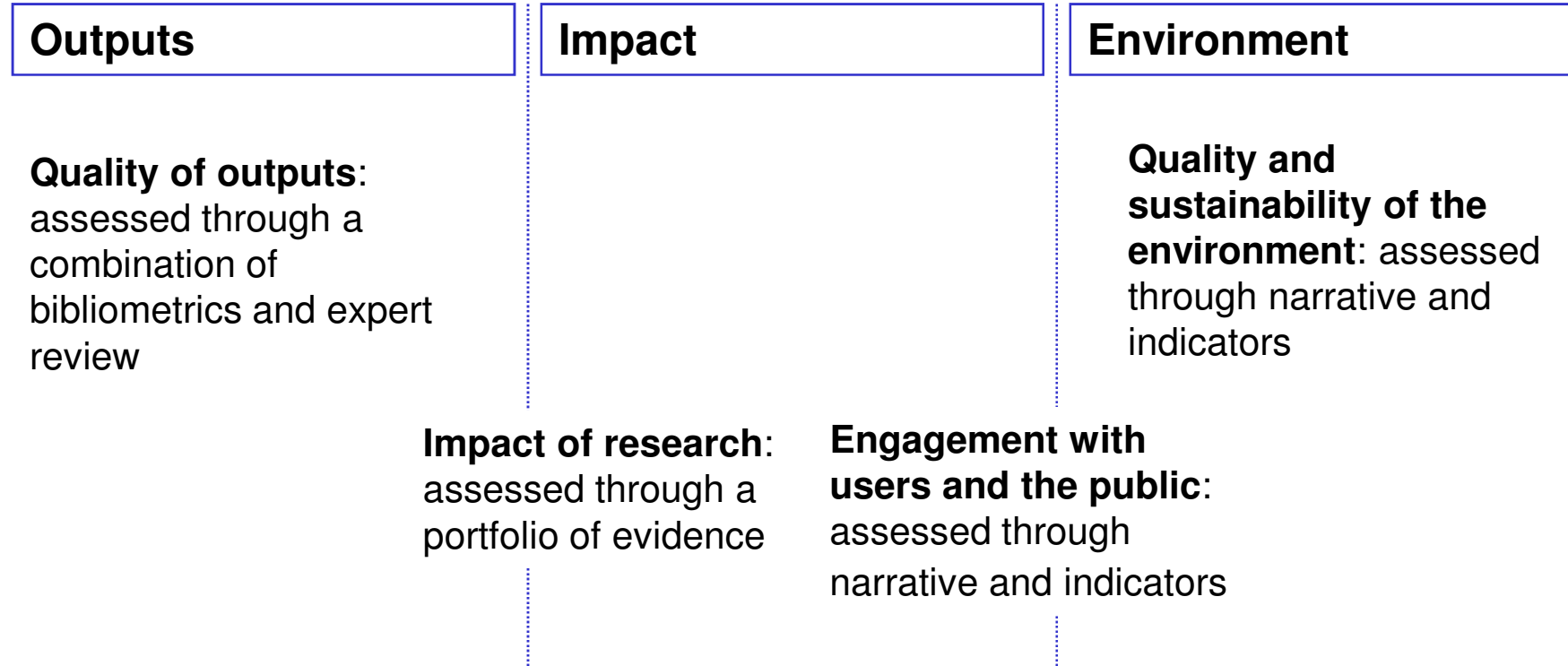
The toolkit:

- Bibliometric analysis
- Expert review of outputs
- Other available indicators
- Submission of information by HEIs
- All of these will be collected and interpreted with input from expert panels

Main elements of assessment

- The quality of research outputs remains the dominant element of assessment
- Explicit account of the impact of research on the economy, society and quality of life – but how??
- The quality and sustainability of the research environment remains important
- Esteem, which figured in the old RAE, is no longer used.

The REF framework



Model 1 – Top down approach

- Citation indicators are produced by associating outputs in WoS/SCOPUS to HEIs through address/affiliation data, and to 'fields' based on journal categories
- Implications:
 - Employers at the time, not current employers, get the credit
 - What is the volume measure?
 - How would expert panels use/interpret the indicators and combine with Departmental commentary?
 - Potentially very little effort for HEIs, but they will want to check the data

Model 2 – Bottom up approach

- For a defined group of staff, citation indicators are produced based on (all their indexed?) outputs
- Implications:
 - Better fit with other aspects of REF which are Department-based
 - Will HEIs select staff?
 - Collecting publications data, which must be linked to staff and to WoS/SCOPUS

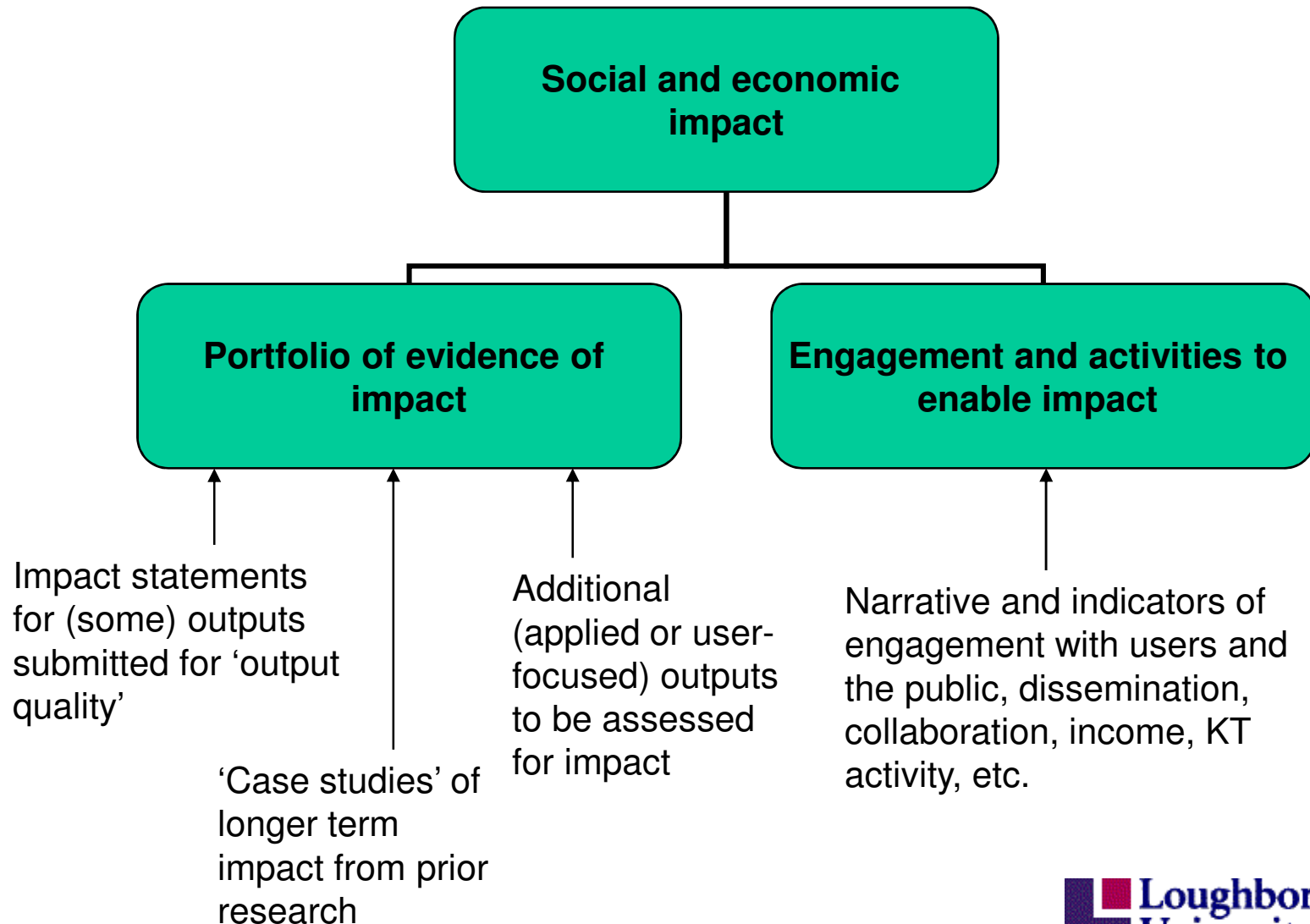
Model 3 – Indicative profiling

- HEIs use citation data to select 'best' outputs which they then submit; expert panels review samples and supplementary (non journal) outputs
- Implications:
 - Fits with other elements of the REF across all subjects
 - How useful are citation rates for individual outputs? How much sampling would be needed?
 - For HEIs, submissions process similar to RAE

WHICH MODEL WILL BE USED?

- No decision yet!

Impact – possible approaches



Criteria for assessment

- Impact could be assessed for the breadth and depth of social, economic or quality of life benefits:
 - while meeting a minimum threshold for rigour and originality
 - assessment by a combination of user and academic experts

BIBLIOMETRIC PART OF THE REF

- HEFCE counts the numbers of citations to all the papers and totals them up using WoS and/or SCOPUS (for pilot, it's just WoS)
- HEFCE assigns papers to subject area
- HEFCE does a world calculation of the average number of citations per paper per year for that subject area
- A profile will then be created of proportion of papers from Unit of Assessment (UoA) that are uncited, below world average, at world average, above world average; maybe by percentiles.
- Decisions yet to be made about excluding certain publications, e.g., in "popular" outlets, review papers (characterised by number of citations in that article) from these calculations
- Followed by a round of peer review ("light touch" for STM, heavier touch for arts/humanities) to amend profiles in light of particular circumstances of Department/subject area
- The profile still forms just one component of final REF assessment of UoA – PhDs, research income, etc., still get considered

Timetable

Now –	Complete bibliometrics pilot
June 2009	Develop proposals for all key elements of the REF
Autumn 2009	Consult on REF proposals
Early 2010	Announce consultation outcomes, including implementation plan and initial guidance to HEIs
2013	Full REF exercise to inform funding from 2014

WHY BIBLIOMETRICS?

- Civil servants clearly felt that this would provide a cheap and reliable method of evaluating research
- But, following up the One Important Question, it is backward looking only and does not evaluate future research strategy
- There are other issues as well, as we shall see!

CHEAP AND RELIABLE?

- I'm partly to blame for this
- In a series of articles published since 1997, I have demonstrated the statistically significant correlation between RAE results and citation counts – and have argued that citation counting could and should be used as a cheap and reliable substitute for expensive and subjective peer review
- It's possible (I don't know) that Treasury civil servants read my articles and were persuaded by them

IF THIS IS WHAT THE CIVIL SERVANTS DID.....

- ...then they were being naïve
- I made it clear that to reliably undertake such studies, you needed **subject experts** to carry out the analyses manually
- Instead, the Treasury instructed HEFCE to go for a purely algorithmic approach

THE EVIDENCE

- **All** studies carried out so far have shown a statistically significant correlation between RAE scores and citation counts
- Subjects evaluated include: archaeology; business studies; genetics; library and information management; engineering; music; psychology
- So, the whole gamut of pure science, engineering, social sciences and humanities – but not medicine yet

THE CORRELATIONS ARE HARDLY SURPRISING

- Citation counts are a measure of **impact**
- And impact is closely related to **quality**
- Nonetheless, the two concepts are not synonymous

BUT IF THE CIVIL SERVANTS WERE NAÏVE, SO ARE CRITICS OF CITATION ANALYSIS

A long familiar catalogue of criticisms, aptly called “fairy tales” by Ton van Raan, head of CWTS in Leiden:

- *Web of Knowledge* has poor coverage of the humanities, computer science, conferences, monographs.....
- Poor coverage of non-English language sources
- Co-authors only included post-2000
- People with the same surname and initials
- Same person using different names, e.g., after marriage
- And more – but all have been shown to be statistically insignificant

A KEY POINT

- No matter how convincing the objective arguments might be, if people don't "buy into" the concept, there will be problems
- Most academics simply don't believe citation counts are an adequate substitute for peer review
- So the current approach to the REF, combining bibliometrics with peer review, makes a lot of sense. It now looks like the Treasury will accept a REF that relies much less on bibliometrics than Gordon Brown's announcement indicated

WHERE WE HAVE ENDED

- Civil servants were naïve to think simple citation counts would do the trick
- Many academics are naïve in believing that citation counts cannot work in their subject area
- The proposed new REF gives us the best of both worlds
- But what weighting for bibliometrics and peer review?
- Will a new Government scrap the REF altogether??

THE POLITICS OF BIBLIOMETRICS V RAE

- Bibliometrics – all data is in the public domain, so anyone can replicate and check if they've been calculated correctly; numbers are "objective"
- RAE – decisions taken behind closed doors
- HEFCE knows use of bibliometrics is controversial, and is determined to involve stakeholders at all stages of the pilot and implementation of the REF

POSSIBLE IMPACT OF OVER-RELIANCE ON CITATION COUNTS ON RESEARCH PRACTICE

- Development of citation clubs
- Pressure to only publish in high IF journals
- Reluctance to jointly author with junior staff/PhDs
- Reluctance to publish blue skies, speculative, interdisciplinary research outputs
- Eventually such game-playing will mean yet another way of evaluating research will be needed

ACKNOWLEDGEMENTS

- I would like to thank Graeme Rosenberg of HEFCE for the use of adaptations of some of his slides in this presentation