World University Rankings

Briefing note by Luke Georghiou and Andrew Walsh

1. Introduction

This note is intended to provide a summary of the increasingly diverse range of world university ranking tables, focussing principally on those which are dominated by research measures. The rankings covered in the paper are distinct from the even greater number of single territory league tables as they seek to assess the quality, impact or standing of institutions on a scale that is applicable internationally. Some league tables use a narrow set of indicators (or even a single indicator) whereas others seek to provide a broader assessment of institutional performance across a range of weighted indicators.

The most prominent world university ranking tables, each of which uses a range of different indicators, are:

1. *Academic Ranking of World Universities (ARWU)* produced since 2003 by Shanghai Jiao Tong University (China)

2. *Times Higher Education World University Rankings (THE - WUR)* produced in association with Thomson Reuters since 2010

3. *World University Rankings*, produced by Quacquarelli Symonds since 2005 (in association with THE until 2010)

The strengths and weaknesses of each of these ranking tables will be considered in turn. Following this, brief information will be provided on eight other ranking tables that have attracted attention in recent years.

2. General Limitations

A very considerable literature now exists discussing the phenomenon of rankings and those who produce them. It is fair to say that while some positive aspects are recognised the tone is overwhelmingly one of criticism of the ways in which ranking tables are produced and to a lesser extent of the use made of them.

It is important firstly to acknowledge the positives. These include the use as an input to benchmarking with the aim of improving an individual institution or even a national system. There are also benefits accruing to potential students, academics and research collaborators who may find them a convenient way to compare key features of institutions they are contemplating joining or working with. Even a relatively poor performance can stimulate some governments to invest resources to improve their national position.

Why then do rankings have such a negative image, particularly among the communities who work on indicators (in the fields of scientometrics and higher education studies). The criticisms may be grouped into four main categories: data, indicators, aggregation and aggregation methods, and governance.

2.1 Data

The data used by ranking organisations are generally derived either from public statistical sources (for example staff numbers, publication and citation databases, research funding) or else collected directly from institutions. Some but not all ranking organisations seek to check these with institutions. The principal problem here lies in the lack of international comparability for all but a few items covered by the OECD Frascati Manual and similar publications. As Barré has pointed out in a well known paper on the difficulties of benchmarking UK and French performance[1] there is no common conception of a full-time member of staff in an academic institution. OECD statistics simply assume 50% of time spent on research – he comments "the real situation may be completely different in different institutions, disciplines, categories of personnel, stage in career", to which may be added the status of people working for other organisations situated in a university and issue if identification for those in affiliated hospitals etc. Funding data are still more problematic as financial models differ widely and in some cases universities are not allowed to receive certain categories of income (for example funding from business). Conventions for treatment of endowments form another source of difference. Issues with publication and citation data are discussed in the next section.

A key issue surrounding data is that of transparency and reproducibility. Ranking organisations rarely reveal the base data that support their calculations (ARWU is particularly opaque and has been frequently criticised by analysts who have been unable to replicate their indices). There is an ethical issue here – some of the ranking organisations seek to exploit their data commercially in follow-up studies and have a vested interest in a black box approach. This issue is returned to in the discussion on governance below.

2.2 Indicators

Ranking tables are normally based upon a composite or aggregation of a series of individual indicators, which may also be ranked. All are proxies for what really needs to be measured. The main approaches each have serious limitations:

Citations

Most rankings rely on inputs from citation data. There are well rehearsed criticisms of this approach including challenges to its accuracy (eg wrong addresses, misspelt names etc.), and biases (US bias, English language bias, greater tendency to cite reviews and methodology papers than original findings, excessive citing of editors etc). Treatment of publications with large numbers of co-authors is a further issue – most analyses do not use fractional counting.

Nonetheless it is reality of academic life that citations are part of the assessment environment at individual, institutional, field and national level. Citations do not measure research quality directly as there may be many reasons for citing a paper, including negative citations – stating that the findings are wrong. However, it is normally accepted that they measure impact, in that other researchers demonstrate their awareness of the research by citing it and peer reviewers of the citing papers accept the validity of the citations.

[1] Barré R., Sense and nonsense of S&T productivity indicators, Science and Public Policy, volume 28, number 4, August 2001, pages 259–266,

The weakest element in citation data is normally the institutional address – a critical issue when dealing with institutions which have affiliated hospitals, have a complex organisation or have been through mergers or name changes. The range of error could be up to 20%.

It is also the case that the importance of citations varies strongly between fields, being important in the sciences and economics, less so in some engineering and social sciences and not measured or used in arts and humanities subjects where books are the main medium of output. Hence, bibliometric analyses favour science and engineering institutions over full service universities.

A frequently used but questionable indicator is that of citations per paper (normalised by field, though not taking account of variations within a field). The recent controversy over the Times Higher Education Rankings (see below) illustrates many of the problems of using bibliometrics. The now infamous projection of the University of Alexandria certainly demonstrates data issues (being largely based on self-citations to one person and citations to the same person in a journal he edits which is itself the subject of scientific controversy.[2] However, the real flaw in this approach is more structural – the idea that a needle peak of performance can be equated to excellence on a broad front. This is effectively an ideological position held by Evidence Ltd, the branch of Thomson Reuters responsible for the analysis. In an earlier analysis for the UK Government the same analysts argued that the UK and German institutions led the USA in life sciences. However, the Max Planck institutes produced a few tens of papers , LMB a few hundred while Harvard, the first of 17 US institutions that followed, produced 10,000. It seems clear that a high volume of sustained excellence has a value in itself that is not recognised by crown indicators.

Reputational surveys

Two of the major ranking organisations use reputational surveys as major components of their activities. This involves asking a sample of academics to rank universities against one or more criteria. The general limitations of reputational assessment were exposed by the US National Research Council in a review of reputational assessment that compared US graduate programmes[3] (A far more feasible objective than comparing universities across the world). The NRC panel concluded:

- Such studies inherently reflect perceptions that may be several years out of date and do not take into account recent changes;

- The ratings of individual programmes are likely to be influenced by the overall reputation of the university – a “halo effect”;

- A disproportionately large fraction of the evaluators are graduates or faculty members of the largest programs which may bias survey results;

To this may be added the criticism that nominations are only likely for the most visible so there is little discrimination lower down the table.

[2] Schiermier Q, Self-publishing editor set to retire, Nature Vol 456, p432 November 2008

[3] Jones LV, Lindzey G and Coggeshall PE (Eds.), 1982, An Assessment of Research-Doctorate Programs in the United States: Mathematical and Physical Sciences, National Academy Press, Washington D.C.

2.3 Aggregation and Aggregation Methods

The broad issue of aggregation includes the basic observation that ranking tables include institutions with very diverse missions and for whom any chosen indicators would vary widely in the weighting that should be attached to them. Taking this point further the validity of comparing very different institutions is itself questionable – not apples and oranges but apples and armchairs…

The problems experienced by ranking organisations in compiling tables is illustrated by instability – horizontally in terms of very wide differences between tables and longitudinally with method or data changes (often without transparency) causing changes in position for institutions. Relatively small differences in the indices can have dramatic effects on rank in some "bunched" parts of the tables. There is a suspicion that these differences may be statistically insignificant but transparency is insufficient to assess this.

The technicalities of aggregation methods have also been criticised in the literature. Weights (scaling constants) should be used to normalise the data not to reflect its relative importance.[4] It is very worrying that the response of the Times Higher to challenges about the validity of some of its indicators is simply to reduce their weighting.

2.4 Governance

A recent polemical article has challenged the motivations and legitimacy of the ranking organisations.[5] Virtually all of these organisations are self-appointed and the majority use the rankings to support commercial operations for analytical services or media purposes. Olds et al point out that annual rankings serve only the interests of the rankers given the slow pace of change in institutional performance (they cite the RAE/REF cycle as a counter-example).

The credibility of rankings has been diminished by the proliferation of tables and disputes between ranking organisations. The Times Higher risks damaging its own reputation by attaching itself to one ranking effort which compromises its ability to take a critical perspective upon the sector it seeks to serve.

Risks to the sector include convergence of behaviour around the indicators resulting in mission creep and loss of diversity. To the extent that reputation becomes driven by tables and has financial consequences there is also a significant risk of litigation.

[4] Billaut J-C, Bouyssou D and Vincke P, Should you believe in the Shanghai Ranking – an MCDM View, Scientometrics (2010) 84:237-263

[5] Olds K, Hazelkorn E and Robertson S, Governing world university rankers: an agenda for much needed reform, Global Higher Ed http://globalhighered.wordpress.com/2010/09/21/governing-world-university-rankers-an-agenda-for-much-needed-reform/

Annex: Strengths and Weaknesses of Ranking Tables

Academic Ranking of World Universities (ARWU)

http://www.arwu.org/

Produced since 2003 by Shanghai Jiao Tong University (China) and from 2009 by the Shanghai Ranking Consultancy.

Indicators:

- Number of alumni winning Nobel Prizes and Fields Medals (10 percent)
- Number of staff winning Nobel Prizes and Fields Medals (20 percent)
- Number of highly-cited researchers in 21 broad subject categories (isihighlycited.com) (20 percent)
- Number of articles published in Nature and Science (20 percent)
- Number of articles recorded in the Science Citation Index and Social Sciences Citation Index (Thomson Reuters) (20 percent)
- Per capita academic performance (on the indicators above) of an institution (10 percent)

Strengths:

- Earliest established world university ranking table
- Most respected table in most countries
- Utilises a range of weighted indicators to reflect different aspects of institutional performance
- Internationally comprehensible set of indicators, all of which are widely considered to be reasonable reflections of institutional standing
- Indicators are reasonably simple, objective and transparent
- Ranking outcomes are not notably counter-intuitive
- Limited year-on-year volatility

Weaknesses:

- Based solely on quantitative indicators of research performance and esteem – no indication of performance in other areas of institutional mission
- No account taken of 'input indicators' – research income, research training, diversity of research population, etc

- Indicators give particular credit to a small number of peaks of exceptional research performance and, despite per capita measure, to sheer volume of research output
- Highly-cited researchers indicator is based on a web resource that is updated infrequently and is frequently challenged as a metric of research impact (it takes no account of author position and favours activity within the confines of a single subject area)
- By contrast with the transparency of the indicators themselves, the data collection and manipulation processes are highly opaque with minimal involvement of or responsiveness to institutions themselves (even when clear errors are discovered)
- Some lack of confidence in the reproducibility of the ranking outcomes[6]
- Some indicators have apparent anglocentic bias

Times Higher Education World University Rankings

http://www.timeshighereducation.co.uk/world-university-rankings/

Produced between 2005 and 2009 in association with Quacquarelli Symonds. For 2010, THE decided to work instead with Thomson Reuters. This summary considers the 2010 methodology and outcome.

Indicators

- Citation impact (normalised average citations per paper) (32.5%)
- Research volume (papers per academic – 4.5%), income (scaled – 6%) and reputation (peer survey – 19.5%)
- Teaching and learning: reputational survey (15%), PhD awards per academic (6%), undergraduates admitted per academic (4.5%), income per academic (2.25%) and PhD awards/bachelor awards (2.25%)
- Industry income and innovation – research income from industry per academic (2.5%)
- International mix – ratio of international to domestic staff (3%) and students (2%)

Strengths:

- Diverse range of quantitative and qualitative indicators covering multiple aspects of institutional mission
- Prominence due to heavy promotion by THE
- Substantial involvement of institutions in data collection and checking for some indicators

[6] Răzvan V. Florian (June 2007). "Irreproducibility of the results of the Shanghai academic ranking of world universities". *Scientometrics* **72** (1): 25–32 http://www.springerlink.com/content/567201224678618j/

- Wide consultation over selection and weighting of indicators
- Intention to overcome previous limitations regarding subject normalisation for bibliometric data (to remove bias towards institutions with substantial biomedical activity) and assessment of smaller and monotechnical institutions

Weaknesses:

- Unreliability of reputational survey components – prone to conservatism and anglocentric bias and open to manipulation. Also unlikely that respondents can discriminate sufficiently below top choices. Concerns also expressed about fairness of sample group and differential response rates.
- Income indicators are hard to compare cross-nationally due to differential contexts for purchasing power, public vs. private support, indirect costs, etc.
- Substantial bibliometric component has substantial 'black box' of complex methodology for output normalisation and compensation for institution type, national context, etc.
- 2010 rankings contain a number of counter-intuitive outcomes. These predominantly result from skewed bibliometric indicators – e.g. placing Alexandria University (Egypt) and Bilkent University (Turkey) in the upper echelons on the basis of small volumes of highly-cited work by as few as one or two member of staff.
- New ranking so year-on-year volatility cannot be gauged

World University Rankings

http://www.topuniversities.com/

Produced by Quacquarelli Symonds since 2005 (in association with Times Higher Education until 2010).

Indicators:

- Academic peer review *via* international reputational survey (40%)
- Recruiter review *via* international reputational survey (10%)
- Faculty (staff):student ratio (20%)
- Citations per faculty – 5 year total citations (from Scopus) divided by total academic staff (20%)
- International orientation – percentage of international staff (5%) and students (5%)

Strengths:

- Range of indicators permit some breadth of coverage of university missions

- Survey attempts to combine qualitative and quantitative measures of performance and standing
- Non survey-based indicators are reasonably simple and objective
- Good involvement of institutions in data collection and checking

Weaknesses:

- Substantial dispute about merits of reputational survey components (see THE above)
- Very high weighting (50%) for disputed reputational survey components
- High general position of UK universities out of line with other ranking exercises
- Crude bibliometric indicator makes no allowance for natural subject variations in citation practice (thus disadvantaging institutions without substantial biomedical output)
- Unreasonable level of year-on-year volatility
- Uncertain future following loss of THE partnership and associated prominence

Other Ranking Tables

1. The *Performance Ranking of Scientific Papers for World Universities* produced since 2007 by the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT). This output-based ranking is based upon a set of eight bibliometric indicators (11 years articles, current articles, 11 years citations, current citations, average citations, H-index, highly cited papers, high Impact journal articles) representing three different criteria of scientific papers performance: research productivity, research impact and research excellence. Data is sourced from Thomson Reuters Web of Science.

   http://ranking.heeact.edu.tw/en-us/2010/TOP/100

2. The *Research Performance Index (RPI)* produced since 2010 by the University of Western Australia Publication. This output-based ranking is based upon an average of five normalised 'faculty' G-index scores, calculated using 10 years of citation data drawn from Elsevier's Scopus.

   http://www.highimpactuniversities.com/

3. The *Global Universities Ranking (GUR)* produced since 2010 by RatER (Rating of Educational Resources- a Russian 'independent rating agency'). This ranking is based upon assessment conducted by a pool of 'experts' formed by project officials and managers to determine the rating scales for every indicator of performance of the universities in seven areas including academic performance, research performance, faculty expertise, resource availability, socially significant activities of graduates, international activities of the university, and international opinion of foreign universities. Each expert performs his/her own evaluation of

performance indicators of all the universities. The final evaluation of each indicator is determined as the average of all the expert evaluations.

http://www.globaluniversitiesranking.org/

4. The Leiden Ranking, produced since 2008 by the Centre for Science and Technology Studies (CWTS) at Leiden University (Netherlands). Five output-based ranking tables are provided, each based on a different bibliometric indicator following analysis of data from Web of Science. The indicators are 'number of publications, citations-per-publication, field-normalized average impact ('crown indicator' - 2 variants) and field-normalized average impact * number of publications ('brute force').

   http://www.socialsciences.leiden.edu/cwts/products-services/leiden-ranking-2010-cwts/

5. *SCImago Institutions Rankings (SIR),* produced since 2009 by SCImago Research Group (Spain/Portugal). This output-based ranking uses the Scopus database to determine five indicators: output volume, citations per document, internationally collaborative authorship, field-normalised citation impact and journal outlet prestige.

   http://www.scimagoir.com/

6. The *Professional Ranking of World Universities* produced since 2007 by the École Nationale Supérieure des Mines de Paris. This ranking uses a single indicator - the number of alumni holding a post of chief executive officer or equivalent in one of the 500 leading international companies (Fortune Global 500).

   http://www.ensmp.fr/Actualites/PR/EMP-ranking.html

7. The *Webometrics Ranking of World Universities*, produced since 2004 by the Cybermetrics Lab (CCHS), a unit of the Spanish National Research Council (CSIC). This ranking is based upon four indicators of institutional web presence and impact: size (number of web pages), visibility (number of external links), rich files (number of text files) and scholar (number of papers and citations from Google Scholar).

   http://www.webometrics.info/index.html

8. U-Multirank – a European Commission supported pilot project, commencing in 2010 with 76 participants, intended to "develop a feasible transparency instrument that can contribute to enhancing the transparency of institutional and programmatic diversity of European higher education in a global context".

   http://www.u-multirank.eu/