Purpose and principles of review

- Addresses the most popular global university rankings
- Providing universities with analysis of the methodologies
- Only publicly accessible information was used
- Efforts were made to discover
  ✓ what is actually measured,
  ✓ how the scores for indicators are calculated
  ✓ how the final scores are calculated, and
  ✓ what the results actually mean.
Types of rankings

**Academic rankings producing league table**
- Academic Ranking of World Universities (Shanghai)
- Times Higher Education World University Ranking
- Quacquarelli Symonds (QS) World Universities ranking
- US News & World Report with Quacquarelli Symonds

**Rankings concentrating on research only**
- Leiden Ranking (Leiden University)
- Taiwan Ranking of Research Papers (HEEACT)

**Multirankings – without producing league tables**
- CHE/die Zeit University Ranking (CHE, Germany)
- U-Map classification (CHEPS)
- European Multidimensional University Ranking System (UMultirank) – EU funded project

Global rankings cover not more than 3-5% of world’s universities
Decrease of scores within the Top 500 universities

How big can be the scores of the remaining for 16’500 universities?

Indicators covering elite research universities only

- “Quality of faculty” = staff winning Nobel prizes (Shanghai-ARWU)
- “Highly Cited” = belonging to worlds Top 200 in 21 areas, i.e. 4200 altogether (ARWU)
- “Peer review” = nominating 30 best universities from pre-selected list (QS-based rankings)
- Teaching reputation survey(s) = nominating 30 best (QS-based, THE-TR)
- Universities considered: pre-selection from elite group of universities: ARWU, THE, Leiden
Indicator scores are usually not the indicator values themselves

Each indicator has a dimension or denominator, e.g.: articles count, staff numbers, citations per academic

To make indicator scores dimensionless, either
- values are expressed as percentage of the result of the “best” university
  \[
  \text{Score} = \frac{R_x}{R_{\text{best}}} \times 100
  \]
- \(Z\)-score is calculated as being the difference between the measure \(x\) and the mean value \(X\) divided by standard deviation \(\sigma\):
  \[
  Z = \frac{x - X}{\sigma}
  \]

Simple arithmetics have huge influence on scores

- Where a composite score is calculated from several indicators, ranking providers assign weights to each indicator in the overall score.
- If a ranking predominantly uses absolute values (ARWU, Webometrics), its scores are size-dependent, i.e. the ranking favours large universities.
- This means that the ranking provider’s subjective judgement determines which indicators are more important (e.g. citations – 10%, reputation – 40%).
### Shanghai ARWU: transparent, elitist and preferring big

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Indicator</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Education</td>
<td>Alumni winning Nobel Prizes and Fields Medals</td>
<td>10%</td>
</tr>
<tr>
<td>Quality of Faculty</td>
<td>Staff Nobel Prizes and Fields Medals</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Highly cited researchers in 21 areas</td>
<td>20%</td>
</tr>
<tr>
<td>Research Output</td>
<td>Papers in Nature and Science*</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Papers indexed in Science Citation Index-expanded and Social Science Citation Index</td>
<td>20%</td>
</tr>
<tr>
<td>Per Capita Performance</td>
<td>Per capita academic performance of an institution</td>
<td>10%</td>
</tr>
</tbody>
</table>

### QS ranking: ever changing, reputation dominated

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Weight</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic reputation survey</td>
<td>40%</td>
<td>Without visits, selecting from a pre-selected list online, 9'000 of 180'000 academics answer over 3 years</td>
</tr>
<tr>
<td>Employer reputation survey</td>
<td>10%</td>
<td>Big, international, often QS clients, hard to find out how it is actually organized just over 2900 responses</td>
</tr>
<tr>
<td>Stud/staff</td>
<td>20%</td>
<td>Actually - staff/student ratio</td>
</tr>
<tr>
<td>Citations</td>
<td>20%</td>
<td>(WoS + Scopus) /FTE (staff)</td>
</tr>
<tr>
<td>Int.faculty</td>
<td>5%</td>
<td>Int faculty = FTE (international)</td>
</tr>
<tr>
<td>Int. stud</td>
<td>5%</td>
<td>Int students/ FT home stud.</td>
</tr>
</tbody>
</table>
Coverage of the research mission of universities

Indicators:
- Publication count SCI & SSCI, Scopus: - production
- Publication count in *Nature* & *Science* - excellence
- Publications per staff - staff research productivity
- Citations (count) – overall force of HEI
- Citations - per paper or per staff - impact
- Citations to articles in the top impact journals – excellence
- Research income
- Research reputation surveys

*But there are also biases and flaws ...*

Rankings and the teaching. Indicators:
- Alumni who have been awarded a Nobel Prize
- Staff/Student ratio
- Reputation surveys (academics, students, employers)
- Teaching income
- Dropout rate
- Time to degree
- PhD/ undergraduate ratio

*All of the above are distant proxies, some questionable*

- Learning outcomes – *are we there yet?*
BIASES AND FLAWS

Natural sciences and medicine vs. social sciences and humanities bias

Bibliometric indicators primarily cover journal publications and conference proceedings

- Natural and life scientists primarily publish in journals,
- Engineering scientists - in conference proceedings,
- Social scientists and humanists – in books

Several indicators count by 22 broad areas as defined by ISI:

1. Agricultural Sciences
2. Biology & Biochemistry
3. Chemistry
4. Clinical Medicine
5. Computer Science
6. Ecology/Environment
7. Economics & Business
8. Engineering
9. Geosciences
10. Immunology
11. Materials Science
12. Mathematics
13. Microbiology
14. Molec. Biol. & Genetics
15. Neuroscience
16. Pharmacology
17. Physics
18. Plant & Animal Science
19. Psychology/Psychiatry
20. Social Sciences, General
21. Space Sciences
22. Multi-disciplinary
## Different publication and citation cultures in different fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Papers per faculty</th>
<th>Citations per faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological Sciences</td>
<td>7.62</td>
<td>59.62</td>
</tr>
<tr>
<td>Physical Sciences and Mathematics</td>
<td>6.39</td>
<td>31.94</td>
</tr>
<tr>
<td>Engineering</td>
<td>6.04</td>
<td>17.83</td>
</tr>
<tr>
<td>Social and Behavioral Sciences</td>
<td>2.14</td>
<td>5.47</td>
</tr>
<tr>
<td>Arts and Humanities</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

*Source: presentation of Cheng at IREG 2010 conference in Berlin*

## Field normalisation - solutions and issues

Field-normalised citations per publication indicator (Leiden ‘Crown indicator’)

\[
\frac{CPP}{FCSm} = \frac{C_1 + C_2 + C_3 \ldots}{e_1 + e_2 + e_3 \ldots}
\]

- \(C_i\) is the number of citations of the publication \(i\)
- \(e_i\) is the expected number of citations of publication \(i\) given the field and the year

Criticisms:
- Prefers older publications,
- Blurs the picture
Mean-normalisation - solutions and issues

- New attempt (2010) - mean-normalised citation score (MNCS)
  \[
  MNCS = \frac{1}{P} \left( \frac{C_1}{e_1} + \frac{C_2}{e_2} + \frac{C_3}{e_3} + \cdots \right)
  \]
- Good idea, **but**: now the results are unstable for the very newest publications
- To avoid the new flaw MNCS indicator is used which **leaves out** publications of the last year
- And then it appears that a single publication may substantially change university’s score
- But after all, it just improves mathematics, not the issue that WoS and Scopus insufficiently cover books

'Peer review' biases and flaws

- Why calling reputation surveys “Peer reviews”?
- 'Peers' are influenced by previous reputation of the institution (including positions in other rankings)
- Limiting the number of universities nominated (THE, QS based rankings) makes approach elitist – and strengthens previous reputation dependence
- Using pre-selected lists rather than allowing ‘peer’s’ free choice results in leaving out huge numbers of institutions
- Is 5-10 % response rate a sufficient result?
- How does opinion survey work when used internationally?
Other issues

- Language bias
- Regional biases
- Calculation of total score reflects ranking provider’s understanding of quality
- The secondary effects of previous ranking positions
- Global league tables do not cover ‘third mission’

DO YOU AGREE /DISAGREE WITH THE FOLLOWING STATEMENTS?

Percentages of «strongly agree» and «somewhat agree combined»
Basis: 350 respondents in 30 countries

Some Institutions manipulate their data to move up in the rankings
Makcs institutions focus on numerical comparisons rather than on educating students
Methodologies and data used are neither transparent nor reproducible
Appropriate metrics are not included when compiling institutional comparisons
Quantitative information mislead institutional comparisons
Only institutions that have always been ranked highly continue to be ranked highly

Source: Thomson-Reuters, Survey, 2011
The risks of overdoing

- Rankings encourage universities to improve their scores
- Universities are tempted to improve performance specifically in areas measured in rankings
- There are risks that universities will concentrate funds and efforts to the above aspects and pay less attention to issues that are not rewarded in ranking scores such as: quality of teaching, regional involvement, widening access, lifelong learning, social issues of students and staff etc.

Can rankings be improved?

- There will be no improvement from extending 5 distant proxies to 25 – they will still remain proxies...
- Improve coverage of teaching – most probably through measuring learning outcomes,
- Lift biases, eradicate flaws of bibliometric indicators: field, language, regional, but first of all – address non-journal publications properly!
- Change rankings so that they in reality help students to make their choices.
- Addressing elite only, ranking results impact life all universities – it is time to produce rankings that cover more universities!
The new developments: classifications, multi-indicator tools and comparing learning outcomes

U-map classification profiles (EU)
Neutral process indicators – not for value judgments or ranking

- Teaching and learning – levels and orientation of degrees, subject range
- Student profile - mature, distance, part-time,
- Research activity,
- Knowledge exchange,
- International orientation,
- Regional engagement.
Multi-indicator tool U-Multirank (EU)

- Combination of performance indicators
- Ranking based on one indicator, scores in other indicators displayed
- No overall score
- Default-15 indicators
- One can create own set of indicators

### Multirank: default set of 15 indicators

<table>
<thead>
<tr>
<th>Code of Institution</th>
<th>Teaching &amp; Learning</th>
<th>Research</th>
<th>Knowledge Transfer</th>
<th>International Orientation</th>
<th>Regional Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>148</td>
<td>Student staff ratio</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>293</td>
<td>Evaluation rate</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>196</td>
<td>Satisfaction rate</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>111</td>
<td>Quality of teaching</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>222</td>
<td>Reputation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Impact</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>98</td>
<td>Innovation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>152</td>
<td>Outreach</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Source: Multirank presentation, 2011
U-Multirank

- Teaching/learning – same distant proxies – but many 😊

Still to be seen

- How well self-reported will work in international context
- How well student satisfaction data will work in international context,
- Whether (or when?) other parties will turn Multirank into a league table and what will be the consequences

The new developments: AHELO

- OECD’s AHELO project is an attempt to compare HEIs internationally on the basis of actual learning outcomes.
- Three testing instruments will be developed within AHELO: one for measuring generic skills and two for testing discipline-specific skills, in economics and engineering.
- Questions yet to be answered are: whether it is possible to develop instruments to capture learning outcomes that are perceived as valid in diverse national and institutional contexts.
Who will open possibility to involve greater number of universities?

- Global rankings? – not without changing methodology
- Subject rankings? – may be but not if they are barely ‘side products’ of global league tables and not without extending data collection
- National/regional rankings?
- Multi-indicator profiling and ranking tools?

New visualisations of global rankings incl. classifications and multi-indicator tools

- ARWU «Rankings Lab»: possibility to chose indicators and asign various weights
- ARWU GRUP Global Research University Profiles self-submitted data collection, 231 universities
- ARWU «Ranking by indicator (22)»: resembles Multirank
- Times Higher Education - subject rankings
New visualisations of global rankings incl. classifications and multi-indicator tools

- Thomson-Reuters The Profiles Project (reputation, funding, faculty characteristics)
- QS subject rankings – 25 of 52 subjects
- QS Classifications (size, subject range, research intensity, age of university)
- QS Ranking by indicator
- QS Stars: (8 criteria)
- QS Country Guides

Main conclusions

1. Since arrival of global rankings then universities cannot avoid national and international comparisons, and this has caused changes in the way universities function.

2. Rankings so far cover only some of university missions. Lack of suitable indicators is most apparent when measuring teaching performance. The situation is better when evaluating research, but even the bibliometric indicators have their biases and flaws.

3. Publication of the yearly result of any ranking should be accompanied with a «health warning» explaining all the «side effects» as a result of which the scores are skewed.

4. Higher education policy decisions should not be based solely on rankings data.
Thank you for attention!

Andrejs Rauhvargers

http://www.eua.be