EUA
European University Association

STRONG UNIVERSITIES FOR EUROPE

GLOBAL UNIVERSITY RANKINGS AND THEIR IMPACT

EUA Rankings Review

Prof. Andrejs Rauhvargers

Bibliothèque Solvay, Brussels,17 June 2011

Introduction

- Society may like to have league tables
- Politicians like information in a business-like manner
- University rankings encourage national and institutional data collection
- However, the results of global rankingsdepend strongly on the choice of indicators and weights.
- It is difficult, if not impossible, to measure and quantify quality itself – and that leads to using proxies
- Rankings push universities to improve performance specifically in those areas that improve ranking scores
- Rankings prefer some areas: research vs. teaching, sciences vs. humanities, English language vs. others

Purpose and principles of review

- Providing universities with analysis of the methodologies of the most popular global university rankings,
  not judging or ranking the rankings themselves
- Only publicly available and freely 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.

Structure of the Report

- Executive Summary
- Methodologies of the most popular global rankings
- Analysis of results (What does it mean for universities)
- Main conclusions
- Guidance to interpreting ranking results

Selection of rankings

Academic rankings with the main purpose of producing university league tables

- Academic Ranking of World Universities (Shanghai)
- Times Higher Education World University Ranking –
  - ✓ with Quacquarelli Symonds (until 2009)
  - ✓ with Thomson Reuters
- Best Universities Ranking – US News & World Report with Quacquarelli Symonds
- Global Universities Ranking – Reitor (Рейтор, Russia)

Selection of rankings (contd.)

Rankings concentrating on research only

- Leiden Ranking (Leiden University)
- Taiwan Ranking of Research Papers for World Universities – (HEEACT)
- Assessment of University-Based Research – EU

Multirankings – using a number of indicators without the (intention of) producing league tables

- CHE/die *Zeit* University Ranking (CHE, Germany)
- CHE Excellence Ranking
- U-Map classification (CHEPS)
- European Multidimensional University Ranking System (U-Multirank) – EU funded project

Selection of rankings (contd.)

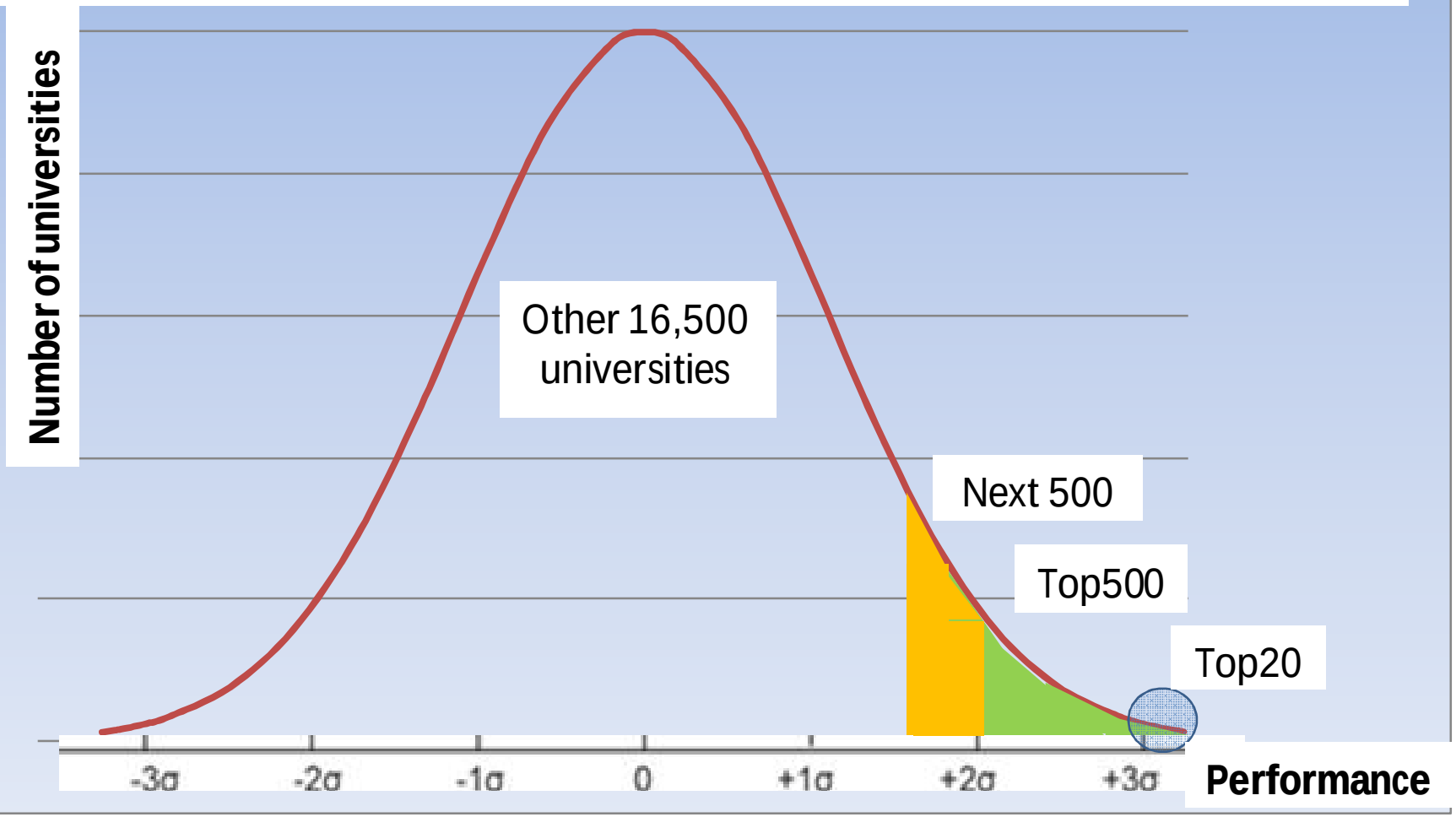
Web rankings

- Webometrics Ranking of World Universities (Webometrics, Spain)

Benchmarking based on learning outcomes

- Assessment of Higher Education Learning Outcomes Project (AHELO – OECD)

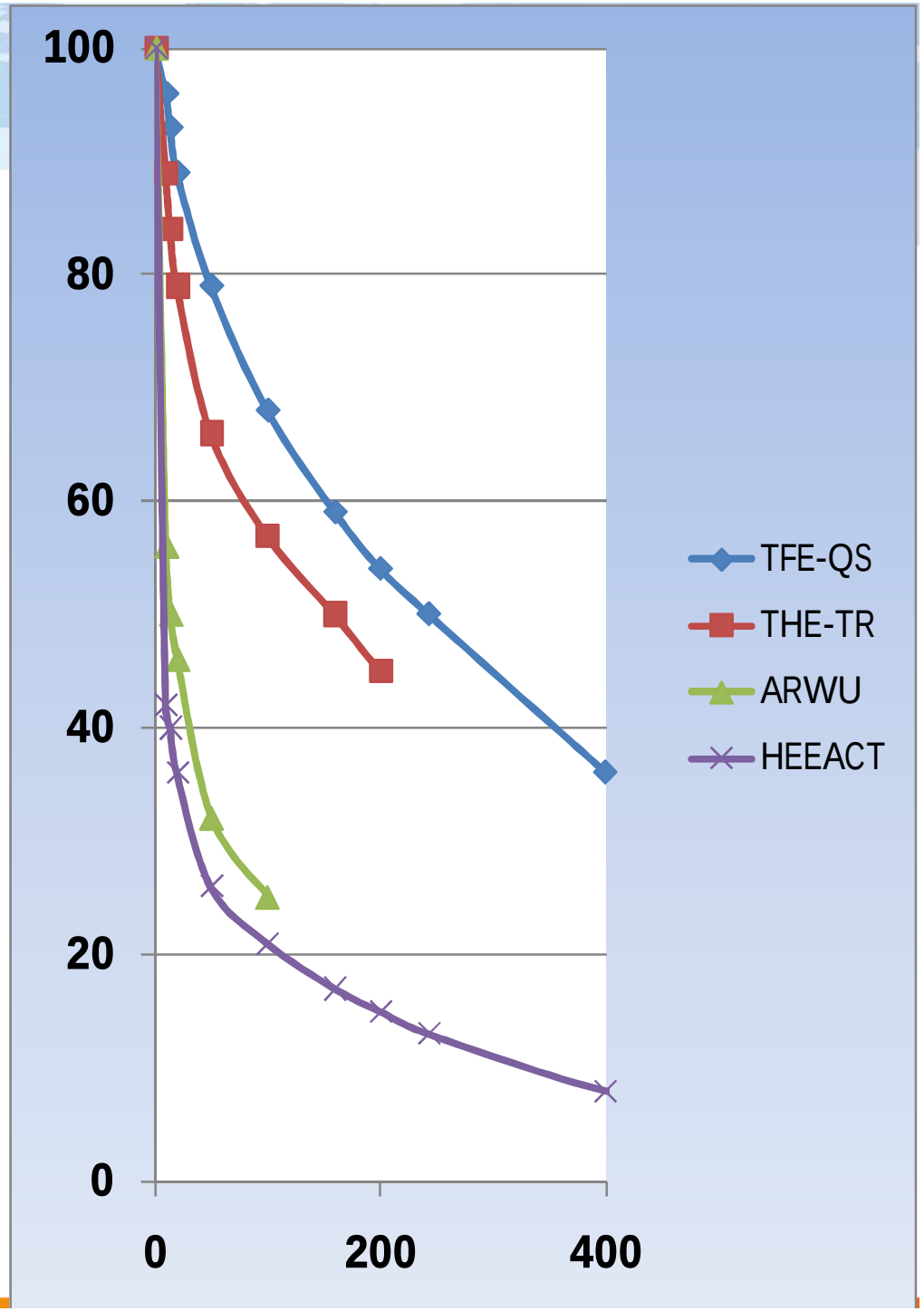
Global rankings cover not more than 3-5% of world's universities



EUA
European University Association

Decrease of scores within the Top 400 universities

How big can be the scores of remaining for 16’600 universities?



Indicators covering elite research universities only

- "Quality of faculty" = staff winning Nobel prizes (ARWU, Reitor)
- "Highly Cited" = belonging to worlds Top 200 researchers in 21 areas, i.e. 4200 researchers in the whole world altogether (ARWU)
- "Peer review" = nominating 30 best universities from pre-selected list (THE-QS and other QS-based rankings, THE-TR)
- Universities considered: selection from elite group of universities: ARWU, THE, Reitor, Leiden

Indicator scores are usually not the indicator values themselves

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

To make indicator scores dimensionless, either

- values are expressed as percentage of the result of the "best" university,

$$\text{Score} = \frac{R_x}{R_{best}} \times 100$$

- *Z-score* is calculated as being the difference between the measure x and the mean value X divided by standard deviation σ:

$$Z = \frac{x - X}{\sigma}$$

Composite scores always contain subjective elements

- In all cases where a composite score is calculated, ranking providers assign weights to each indicator in the overall score.
- This means that the **ranking provider's subjective judgement determines which indicators are more important**

  (e.g. citations – 10%, reputation – 40%)
- In other words, the composite score reflects the ranking provider's concept of quality.

Choosing between simple counts or calculating relative values is not neutral

- If a ranking predominantly uses absolute values, its scores are size-dependent, i.e. the ranking favours large universities.
- If relative values prevail, universities which are more efficient and not necessarily large, will score more highly.
- Predominantly using **absolute numbers** are, e.g. ARWU, Reitor, and also Webometrics
- HEEACT predominantly and THE-QS and THE-TR mainly use relative values (except for reputation surveys).
- The Leiden Ranking, which does not combine indicator scores, offers both size-dependent and size-independent indicators

Rankings and the research indicators used

- Publication count SCI &SSCI, Scopus:
- Publication count (specific): Nature & Science -
- Publications per staff
- Citations (count)
- Citations - per paper
- Citations - per staff
- Citations to articles in the top impact journals
- H – index

Bibliometric indicators are based on reliable data

Other research indicators

- Research reputation surveys
- Research income
- Intensity of PhD production

Teaching mission of universities
Indicators used:

- Alumni that have been awarded a Nobel Prize/ Field medal
- Staff/Student ratio
- Reputation
- Teaching income
- Dropout rate
- Time to degree

All are distant *proxies and some strongly questionable*

BIASES AND FLAWS

Natural sciences and medicine vs. social sciences bias

- Bibliometric indicators primarily cover journal publications
  - ✓ Natural and life scientists primarily publish in journals,
  - ✓ Engineering scientists - in conference proceedings,
  - ✓ Social scientists and humanists – in books

  Several indicators count by *21 broad area*

The "21 broad subject areas"

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. Molecular Biology & Genetics
15. Neuroscience
16. Pharmacology
17. Physics
18. Plant & Animal Science
19. Psychology/Psychiatry
20. **Social Sciences, General**
21. Space Sciences

Different publication and citation cultures in different fields

	Papers per faculty	Citations per faculty
Biological Sciences	7.62	59.62
Physical Sciences and Mathematics	6.39	31.94
Engineering	6.04	17.83
Social and Behavioral Sciences	2.14	5.47
Arts and Humanities	Unknown	Unknown

- Table from 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

Criticism – 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, a modified MNCS2 indicator is used which **leaves out p**ublications of the last year
- But after all, **it just improves mathematics, not the issue that WoS and Scopus insufficiently cover books**

Impact factor – to be used with care

- … especially in social sciences and humanities, expert rankings do not correlate very well with impact factors. (EU WG on assessment of university research, 2010)
- "the impact factor should not be used without careful attention to the many phenomena that influence citation rates, as for example the average number of references cited in the average article. The impact factor should be used with informed peer review" (Garfield, 1994)
- "by quantifying research activity and impact solely in terms of peer-publication and citations, rankings narrowly define 'impact' as something which occurs only between academic 'peers'" (Hazelkorn, 2011).

‘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 rankings) makes approach elitist – and strengthens previous reputation dependence
- Using pre-selected lists ather than allowing ‘peer’s’ free choice results in leaving out huge numbers of institutions
- Is 5% response rate a sufficient result?

Even keeping current position in ranking requires great effort

- J.Salmi attributes this phenomenon to the 'Red Queen effect' (Salmi, 2010). "In this place it takes all the running you can do, to keep in the same place", says the Red Queen in Lewis Carroll's 'Through the Looking Glass'.
- The principle has been also articulated as "for an evolutionary system, continuing development is needed just in order to maintain its fitness relative to the systems it is co-evolving with" (van Heyligen, 1993).

Shanghai ARWU: transparent, elitist and preferring big

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

THE/ QS: ever changing, reputation dominated

Indicator	Weight	Explanation
Peer review	40%	Without visits, selecting from a pre-selected list online, 9'000 of 180'000 academics answer over 3 years
Employer review-	10%	Big, international, often QS clients, hard to find out how it is actually organized just over 2900 responses
Stud/staff	20%	Actually - staff/student ratio
Citations	20%	(WoS + Scopus) /FTE (staff)
Int.faculty	5%	Int faculty = FTE (international)$_x$
Int. stud.	5%	Int students/ FT home stud.

THE – Thomson-Reuters 2010 on

- **Bibliometric indicators** (citations/paper, papers/staff) weight 37%
- **Reputation** of research and teaching – 34.5%
- **Income** (research: overall, from industry, public vs. Total income (all per academic), total weight of 10.75%
- **Importance of PhD** (PhD/ undergrad & PhD/staff) – 8,25%
- **International** staff & student ratios - 5% smaller
- **Undergraduate/ staff ratio** weight of 4.5%.

Taiwan HEEACT: research only, per capita

Criteria	Indicator	Weight	
Productivity	Articles in last 11 y.	10%	20%
	Articles last year	10%	
Impact	Citations last 11 y.	10%	30%
	Citations last 2 y.	10%	
	Av. citations last 11 y.	10%	
Research Excellence	H-index of last 2 y.	20%	50%
	Highly cited papers	15%	
	Articles of current y. in high-impact journals	10%	

Ranking research performance without league table – Leiden ranking

1. Number of publications **P** *(Web of Science + Scopus)*

2. Citations per publication:

   $\boldsymbol{CPP = (C - C_s) / P}$ (C_s- self-citation)

3. field-normalised citations per publication
   $\boldsymbol{CPP_{norm}}$ - takes into account differences in citation among fields

4. "brutal force" indicator:

   $\boldsymbol{P \times CPP_{norm}}$ – shows university's "brute force"

Webometrics: web based, different approach, covering many universities (Top 12,000)

Indicator	Definition	Weight
Visibility	No of unique inward links	50%
Size of web	No of pages in university web	20%
Rich files	No of *pdf, PostScript, doc & ppt*	15%
Scholar	No of papers & citations in Google Scholar	15%

Quite good correlation with global 'academic' rankings

The risks of overdoing

- Rankings encourage universities to improve their scores.
- Universities are tempted to improve performance specifically in areas measured
- Risk: 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.

After all: **how many universities are in Top 500?**
(Answer – exactly 500) *Jamil salmi*

How 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 all universities!

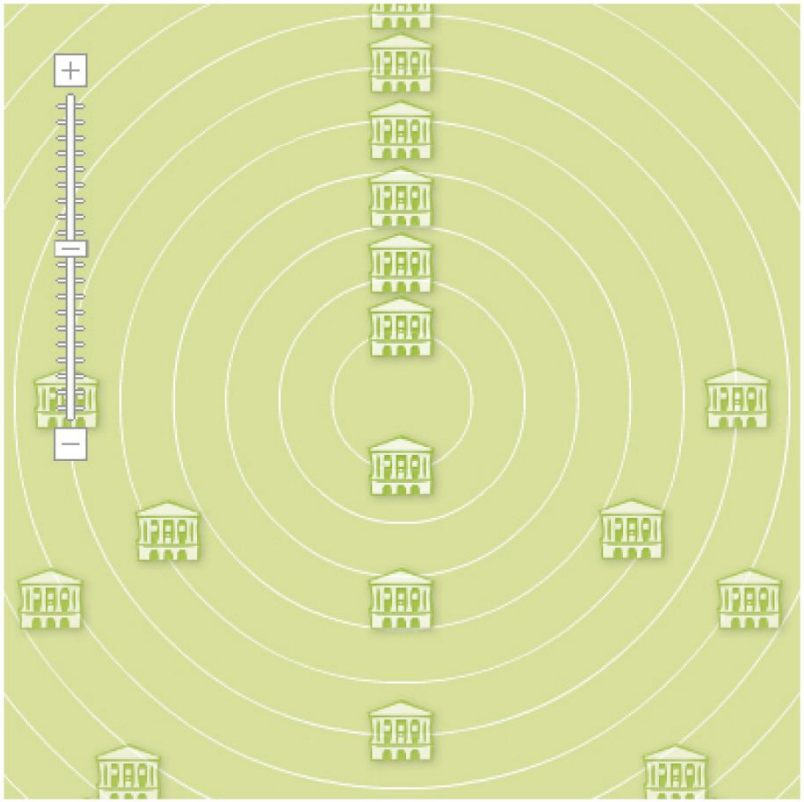
You're always responsible for those who tamed

(The tale of Antoine de Saint-Exupéry)

Informing student choices – CHE university rankings

Quick Ranking: Economics

Universities



MY INDICATORS:

(S)= Student's jugdements (F)= Facts (P)= Professor's judgements

Academic studies and teaching

- [x] Contact between students (S)
- [x] Counselling (S)
- [ ] Courses offered (S)
- [x] E-Learning (S)
- [x] Study organisation (S)
- [x] Teaching evaluation (S)

Equipment

- [ ] IT-infrastructure (S)
- [ ] Library (S)
- [ ] Library - computer workstations
- [ ] Rooms (S)

International orientation

- [x] Support for stays abroad (S)

Job market and career-orientation

- [x] Job market preparation (S)
- [ ] Practice Support (S)

Overall opinions

- [ ] Overall study situation (S)
- [ ] Reputation for academic studies and teaching (P)
- [ ] Research Reputation

Research

- [ ] Many internationally visible publications (F)
- [x] many doctorates (F)
- [ ] many publications (F)
- [ ] much third party funding (F)

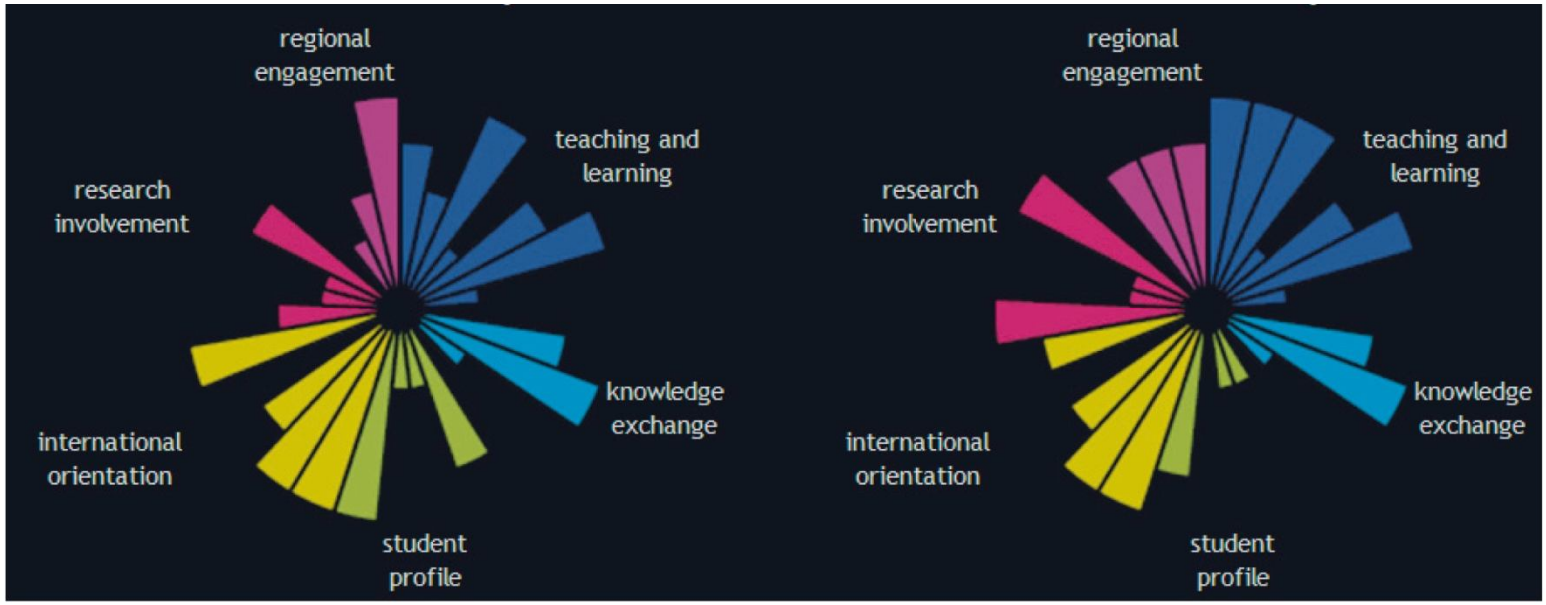
Town and University

- [ ] Higher education sport (S)
- [ ] low rent (F)
- [ ] small university location (F)

The new developments

- Assessment of university-based research (AUBR): Analysis of research indicators, and their suitability, working out a methodology research assessment.
- U-Map uses indicators that characterise the focus and intensity of various aspects in HEIs
- U-Map has two visualisation tools allowing to classify HEIs and to make detailed comparison of selected HEIs.

Source: U-map



The new developments: U-Multirank

- U-Multirank will be a multidimensional ranking including all aspects of an HEI's work – education, research, knowledge exchange and regional involvement.
- No composite score is produced.

Has to be seen in future:

- how well self-reported and student satisfaction data will work in international context,
- whether other parties will turn Multirank into a league table and what will be the consequences

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. *De facto*, the methodologies of global rankings give stable results for only 700-1000 universities.
   The result is that all HEIs are judged according to criteria that are appropriate for the top research universities only.
3. Rankings so far cover only some university missions.

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.

Main conclusions (contd.)

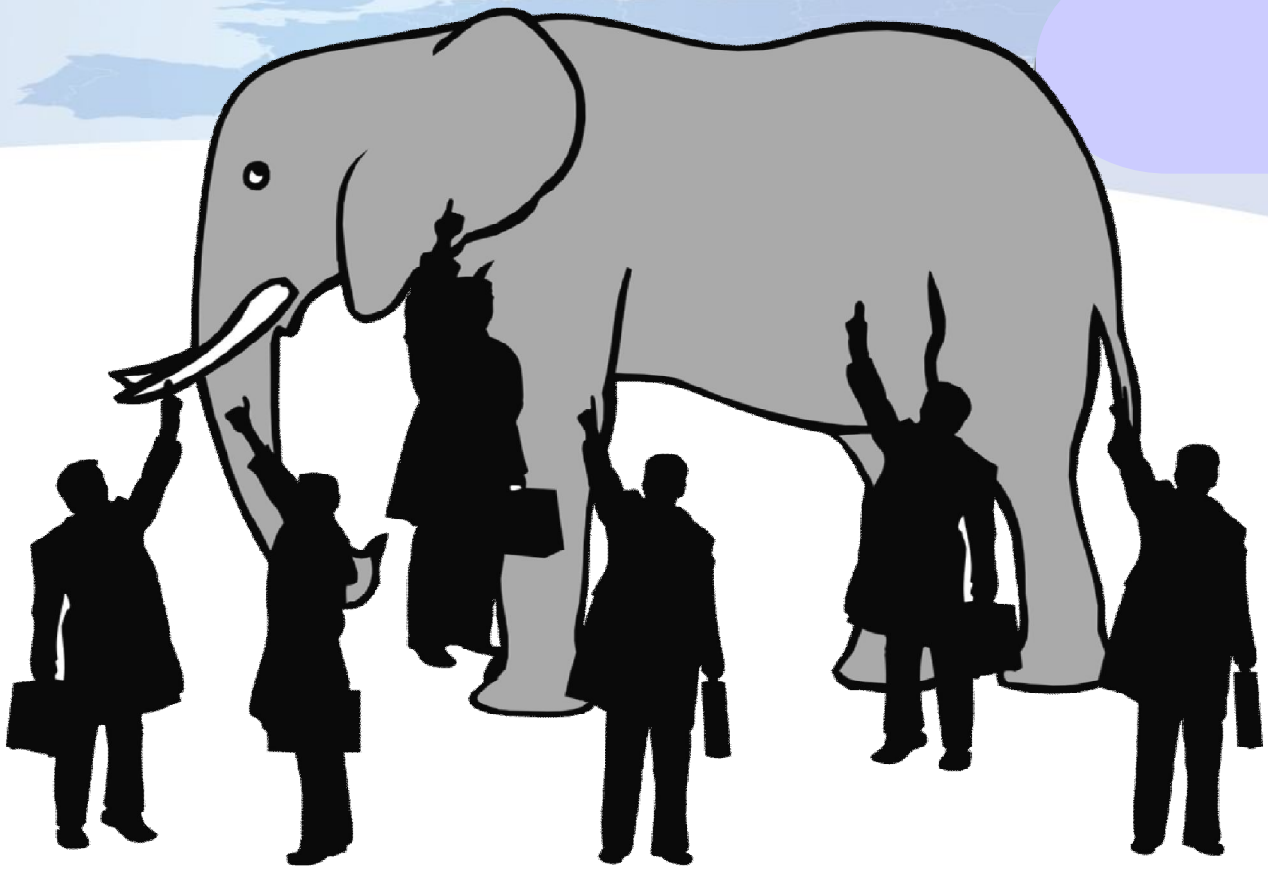
4. Rankings, it is claimed, make universities more 'transparent'. However, the methodologies, especially those of the most popular league tables, still lack transparency themselves.

5. The lack of suitable indicators is most apparent when measuring teaching performance. The situation is better when evaluating research.
   However, even the bibliometric indicators have their biases and flaws. Efforts are made to improve methodologies, usually addressing the calculation method, while the real problem is the use of inadequate proxies, or the omission of part of the information due to methodological constraints

Main conclusions (contd.)

6. At present, it would be difficult to argue that the benefits that rankings provide are greater than the negative effects of the so-called 'unwanted consequences' of rankings.
7. New attempts as the AUBR EU Research Assessment, U-Map, U-Multirank and AHELO, all aim to improve the situation. They are still at various stages of development or pilot implementation, and all of them still have to overcome difficult issues, particularly problems of data collection and the development of new proxies.
8. **Higher education policy decisions should not be based solely on rankings data.**

EUA
European University Association



Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

by John Godfrey Saxe (1816–1887)