EUA REPORT ON RANKINGS 2011

GLOBAL UNIVERSITY RANKINGS AND THEIR IMPACT

Andrejs Rauhvargers

EUA
European University Association

FUNDAÇÃO CALOUSTE GULBENKIAN

Robert Bosch Stiftung

Additional copies of this publication are available for 10 Euro per copy.

For ordering information, please contact publications@eua.be or write to:

European University Association asbl
Avenue de l'Yser 24
1040 Brussels, Belgium
Tel: +32-2 230 55 44
Fax: +32-2 230 57 51

A free electronic version of this report is available at www.eua.be.

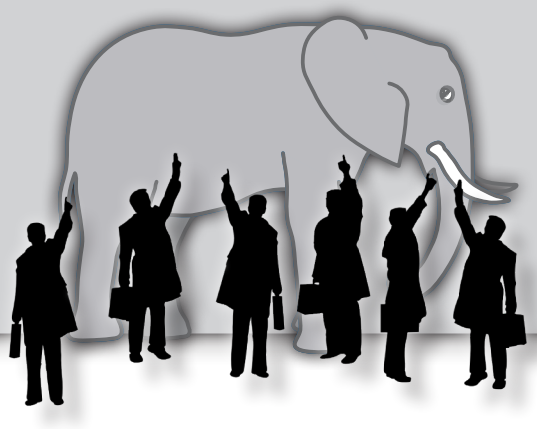
ISBN: 9789078997276

"Six Blind Men and an Elephant" image
Copyright © 2011 Berteig Consulting Inc. Used with permission.

EUA REPORT ON RANKINGS 2011

GLOBAL UNIVERSITY RANKINGS AND THEIR IMPACT

Andrejs Rauhvargers



The Blind Men and the Elephant

by John Godfrey Saxe (1816-1887)

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind.

The *First* approach'd the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant
Is very like a wall!"

The *Second,* feeling of the tusk,
Cried, -"Ho! what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"

The *Third* approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The *Fourth* reached out his eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he,
"Tis clear enough the Elephant
Is very like a tree!"

The *Fifth,* who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an Elephant
Is very like a fan!"

The *Sixth* no sooner had begun
About the beast to grope,
Then, seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a rope!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

MORAL.
So oft in theologic wars,
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen!

CONTENTS

EDITORIAL

EUA commissioned this report in response to the growth in international and national rankings, as a result of increasing questions from member institutions requesting information and advice on the nature of these rankings, because of the interest shown by national governments in ranking exercises, and finally in light of the European Commission's decision to develop a 'European ranking'.

An ad hoc working group of Council was first established in 2009 to consider how the association could best serve its members in this respect. This resulted in a decision of the EUA Council to launch a pilot project to publish, in 2011, the first in a series of EUA reviews of rankings.

This project was entrusted to an Editorial Board chaired by EUA President, Professor Jean-Marc Rapp, former Rector of the University of Lausanne, and including: Professor Jean-Pierre Finance, President of the Henri Poincaré University, Nancy 1 and EUA Board member; Professor Howard Newby, Vice-Chancellor of the University of Liverpool; Professor Oddershede, Vice-Chancellor of the University of Southern Denmark and President of the Danish Rectors' Conference. Professor Andrejs Rauhvargers, Secretary General of the Latvian Rectors' Conference accepted our invitation to carry out this analysis, and is the author of the Report.

We are honoured that the Gulbenkian Foundation and the Robert Bosch Foundation have, together, agreed to support this project over a two year period.

The report focuses on international rankings and also refers to a number of other ongoing projects seeking to measure university performance. It describes and analyses the methodologies used by the main international rankings using only publically available and freely accessible information. It is intended as a service to EUA members, often under pressure to appear in the rankings, or to improve their position in one way or another.

It is clear that despite their shortcomings, evident biases and flaws, rankings are here to stay. They 'enjoy a high level of acceptance among stakeholders and the wider public because of their simplicity and consumer type information' (AUBR Expert Group, 2009). For this reason it is important that universities are aware of the degree to which they are transparent, from a user's perspective, of the relationship between what it is stated is being measured and what is in fact being measured, how the scores are calculated and what they mean.

However, it is important to underline that international rankings in their present form only cover a very small percentage of the world's 17,000 universities, between 1% and 3% (200-500 universities), completely ignoring the rest. They are of direct relevance for only around half of EUA members, situated in a small percentage of those countries in which EUA has members, and strongly correlated with the wealth of those countries.

The report confirms that most international rankings focus predominantly on indicators related to the research function of universities. Attempts to measure the quality of teaching and learning generally involve the use of proxies, often with a very indirect link to the teaching process, and are rarely effective. The importance of links to external stakeholders and environments are largely ignored. Where existing data is used, it is often not used consistently, and reputational factors have in many cases disproportional importance. Taken together, this leads to an oversimplified picture of institutional mission, quality and performance, and one that lacks relevance for the large majority of institutions, especially at a time when diversification and individual institutional profiling are high on agendas across Europe.

On a more positive note, the arrival of global rankings over the last few years has focused considerable attention on higher education, and put the spotlight on universities that are increasingly being compared nationally and internationally. Rankings have certainly helped to foster greater accountability and increased pressure to improve management practices. They have encouraged the collection of more reliable data and in some countries have been used to argue for further investment in higher education. Although it is said they can be used to guide consumer choice, there is little convincing evidence that they do so (except in Asia or via the CHE Ranking in Germany).

It is our view that at present it would be difficult to argue that the benefits offered by the information they provide, given the lack of transparency that we have observed, are greater than the 'unwanted consequences of rankings'. For there is a danger that time invested by universities in collecting and using data and statistics in order to improve their performance in the rankings may detract from efforts to progress in other areas such as teaching and learning or community involvement.

Looking to the future, measures are being taken by the ranking providers to try to improve the methodologies they use, which we can only encourage, and which we will follow up in subsequent reports. The International Rankings Expert Group (IREG) has announced that it will conduct an audit of the various rankings. We hope that this exercise will include the use of independent experts as this would add considerably to its credibility. Among the open questions for future consideration is that of the 'democratisation' of rankings to allow more of the world's universities the opportunity to find their place, and what this would entail. With regard to some of the more recent European initiatives that seek to broaden the focus of rankings to cover the different missions of the university experience suggests that lack of internationally comparable data is a challenge. The debate will continue and EUA will take up these and other questions in future reports.

ACKNOWLEDGEMENTS

The considerable and detailed analysis for this report has been carried out by Professor Andrejs Rauhvargers. The Editorial Board would like to thank Professor Rauhvargers most sincerely for his commitment, for the enormous amount of time he has invested in researching, carefully describing and analysing the various rankings, ratings and classifications included in the review. It has been a challenging enterprise, not least given the initial decision made to take account only of publically available information on the various rankings included.

The Editorial Board would also like to thank all the EUA staff members who contributed to the preparation, editing and publication of this report.

Brussels, June 2011

Jean-Marc Rapp
Jean-Pierre Finance
Howard Newby
Jens Oddershede

ACRONYMS

A&HCI	Arts & Humanities Citation Index
AHELO	Assessment of Higher Education Learning Outcomes
ARWU	Academic Ranking of World Universities by Shanghai Jiao Tong University (in older publications referred to as SJTU)
AUBR	Working group on Assessment of University-Based Research
CPP	Number of citations per publication
ESI	Essential Science Indicators
FCSm	Mean fields citation score – bibliometric indicator
FTE	Full time equivalent
HEEACT	Taiwan Higher Education Accreditation and Evaluation Council
HEI	Higher Education Institution
ISI	Institute for Scientific Information (now Thomson Reuters)
IREG	International Ranking Expert Group
JCR	Journal Citation Reports
MNCS	Mean normalised citation score – bibliometric indicator
P	Number of publications (in referenced journals) – bibliometric indicator
QA	Quality assurance
SSCI	Social Sciences Citation Index
SCI	Science Citation Index
THE	Times Higher Education World University Ranking (in earlier publications referred to as THES – Times Higher Education Supplement Ranking)
THE-QS	Times Higher Education World University Ranking in cooperation with Quaquarelli-Symmonds until 2009
THE-TR	Times Higher Education World University Ranking in cooperation with Thomson Reuters after 2009
WG AUBR	*see AUBR*

GLOSSARY

Ranking

A ranking is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second. In mathematics, this is known as a weak order or total preorder of objects. It is not necessarily a total order of objects because two different objects can have the same ranking. The rankings themselves are totally ordered. For example, materials are totally preordered by hardness, while degrees of hardness are totally ordered[1].

Rating

1. A position on a scale
2. An evaluation of status, especially of financial status
3. A number, letter, or other mark that refers to the ability of something

A sports rating system is a system that analyses the results of sports competitions to provide objective ratings for each team or player. Rankings are then derived by sorting each team's ratings and assigning an ordinal rank to each team starting with the highest rated team earning the #1 rank.

Bursary

a monetary grant to a needy student

Scholarship

a grant-in-aid to a student (as by a college or foundation)

Quality assurance[2]

refers to a programme for the systematic monitoring and evaluation of the various aspects of a project, service, or facility to ensure that standards of quality are being met.

It is important to realise also that quality is determined by the programme sponsor. QA cannot absolutely guarantee the production of quality products, unfortunately, but makes this more likely.

Two key principles characterise QA: 'fit for purpose' (the product should be suitable for the intended purpose) and 'right first time' (mistakes should be eliminated). QA includes regulation of the quality of raw materials, assemblies, products and components; services related to production; and management, production and inspection processes.

It is important to realise also that quality is determined by the intended users, clients or customers, not by society in general: it is not the same as 'expensive' or 'high quality'. Even goods with low prices can be considered quality items if they meet a market need. QA is more than just testing the quality of aspects of a product, service or facility, it analyses the quality to make sure it conforms to specific requirements and complies with established plans.

Benchmarking

Benchmarking is the process of comparing one's business processes and performance metrics to industry bests and/or best practice from other industries. It is the process of identifying 'best practice' in relation to both products and the processes by which those products are created and delivered. The search for 'best practice' can take place both inside a particular industry, and also in other industries. Measurement of the quality of a firm's policies, products, programmes, strategies, etc., and their comparison with standard measurements, or similar measurements of the best-in-class firms. The objectives of this exercise are: (1) to determine what and where improvements are called for, (2) to understand how other firms achieve their high performance levels, and (3) to use this information to improve the firm's performance[3].

[1] http://en.wikipedia.org/wiki/Ranking
[2] An 'industrial' definition of quality assurance was deliberatley chosen in the rankings context.
[3] http://www.businessdictionary.com/definition/benchmarking.html

EXECUTIVE SUMMARY

In recent years university rankings have become increasingly important worldwide; the number of global rankings has grown during the period of this review and it is predicted that it will keep growing[4]. Rankings went global in 2003 when Shanghai Jiao Tong University published the results of the first global university ranking. The importance of rankings seems to have grown exponentially.

Rankings have always been controversial. Society may like to have a league table that allows one to see who is the 'best in the world' (and who is not). Politicians like to be presented with information in a business-like manner – and rankings do just that. In times of significant financial constraints, policy makers in different countries are increasingly interested in comparisons of the performance of various higher education institutions (HEIs) according to objective indicators. However, the results of any ranking, but especially global league tables, depend strongly on the choice of indicators and weights assigned to them. In addition, it is difficult, if not impossible, to measure and quantify quality itself, and therefore rankings use various proxies – some of which are rather distant from the actual quality of teaching or research.

Purpose and principles of this review

The purpose of this European University Association (EUA) review is to inform universities about the methodologies behind the most popular global rankings, and about their potential impact.

The EUA review of global university rankings has been produced following some agreed principles:

- It addresses the most popular global university rankings as well as some other attempts to measure performance (rankings, ratings, classifications) that are relevant for European universities. However, the review does not cover national rankings.
- The review is not aimed at ranking the rankings themselves, but at providing universities with an analysis of the methodologies behind the global rankings.
- Only publicly available and freely accessible information on each ranking, rather than surveys or interviews with the ranking providers, was used in the survey. Such an approach was used in an attempt to demonstrate how transparent each ranking is from a user's perspective.
- Efforts were made to discover what is said to be measured, what is actually measured, how the scores for individual indicators and, where appropriate, the final scores are calculated, and what the results actually mean.

We believe that this ranking methodology analysis will provide useful information to the universities in times when rankings increasingly influence higher education policies and public opinion about them.

Selection of rankings covered in the review

According to the above principles, the following international university rankings were selected for the EUA review:

- Shanghai Academic Ranking of World Universities (ARWU) – Shanghai Ranking Consultancy, China

[4] S. Marginson, interview in The Australian, 12 January 2011.

- Times Higher Education World University Ranking, Times Higher Education:
  – in cooperation with Quacquarelli Symonds until 2009
  – in cooperation with Thomson Reuters as of 2010

- World's Best Universities Ranking – US News & World Report in cooperation with Quacquarelli Symonds, US

- Global Universities Ranking – Reitor (Рейтор), Russia

- EU University-Based Research Assessment – AUBR Working Group, European Commission

- Leiden Ranking – Leiden University, The Netherlands

- Performance Rankings of Scientific Papers for World Universities – Higher Education Accreditation and Evaluation Council, Taiwan

- CHE University Ranking – Centre for Higher Education Development/*die Zeit*, Germany

- CHE Excellence Ranking – Centre for Higher Education Development/*die Zeit*, Germany

- U-Map classification – CHEPS

- U-Multirank ranking – EU funded project

- Assessment of Higher Education Learning Outcomes (AHELO) – Organisation for Economic Cooperation and Development (OECD)

- Webometrics Ranking of World Universities – Cybermetrics lab, Centro de Ciencias Humanas y Sociales, Spain.

Summary of observations and findings

General findings

1. Trends in recent years demonstrate that the **number of global university rankings is likely to keep growing**, although they will become more specialised.

2. Policy makers and society at large often see global university rankings as tools for university "transparency", although it might be difficult to argue the reverse – i.e. that, were there no rankings, universities would be "non-transparent".

3. The landscape of existing global university rankings is diverse covering:

   - University rankings whose main purpose is to produce league tables of top universities only – the Shanghai Academic Ranking of World Universities (ARWU) ranking, mainly based on research indicators; the Times Higher Education (THE) ranking initially in cooperation with Quacquarelli Symands (THE-QS), since 2010 THE in cooperation with Thomson Reuters (THE-TR); and using a different set of indicators; the Russian Reitor ranking, and others.

   - University rankings concerning research performance only – with or without league tables – the Leiden Ranking with no composite score, the Taiwan Higher Education Accreditation Evaluation Council university ranking (HEEACT) with a league table based on a composite score, and the EU Assessment of University-Based Research (AUBR) which is a research assessment methodology targeted at transparency for various purposes, rather than a ranking.

   - University rankings and classifications using a number of indicators with no intention of producing composite scores or league tables – the original German Centre of Higher Education Development (CHE) university ranking was designed to help potential students choose a university according to their requirements, the EU U-Map classification to allow them to find and compare universities with similar profiles, and the EU U-Multirank ranking to compare the performance of universities in various aspects of their activities.

   - OECD is in the initial phases of its Assessment of Higher Education Learning Outcomes (AHELO), which is intended to benchmark universities according to the actual learning outcomes demonstrated by students.

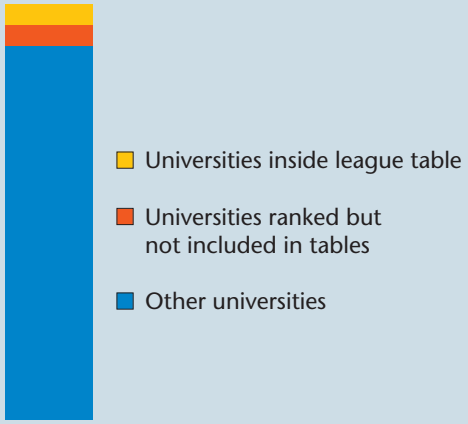
   - Rankings of universities only according to their visibility on the web – Webometrics.

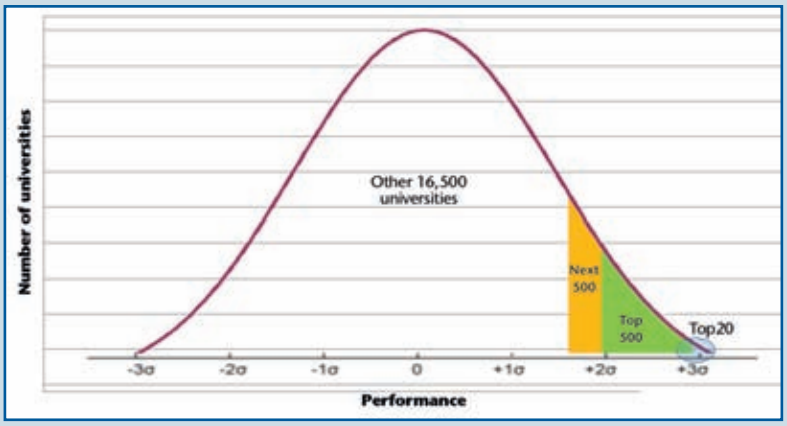
4. Despite their many shortcomings, biases and flaws '**rankings enjoy a high level of acceptance among stakeholders and the wider public because of their simplicity and consumer-type information**' (AUBR Expert Group 2009). Thus, university rankings are not going to disappear; indeed, the number of rankings is expected to increase although they will become more specialised (Marginson, 2011).

Ranking elite universities, shaking all

5. The most popular global league tables (ARWU, THE-QS and THE-Thomson Reuters, US News and World Report Ranking (USNWR), HEEACT, Reitor and others) concern the world's top universities only. First of all, the league tables include roughly 1% to 3% of universities (200-500 universities) out of approximately 17,000 universities in the world. Secondly, it is important to note that the rankings producing global league tables use methodologies that simply **cannot produce stable results for more than 700-1200 universities** in global league tables and just around 300 universities in subject area rankings).

Proportion of universities considered by existing global rankings vs. the total number of universities in the world





6. Due to the elitist approach applied in the methodologies of the global league tables, more than **16,000 of the world's universities will never obtain any rank** in those rankings. Jamil Salmi's (2010) rhetorical question "How many universities can be among the top 500?" and his answer "five hundred" is unequivocal.

7. One problem or 'unwanted consequence', as rankers sometimes call the negative impacts of rankings, is that both society and policy makers are tempted to judge all higher education in the world by the standards that rankings use to detect the top *research* universities, rather than applying one of the core principles of quality assurance – the 'fitness for purpose' principle. In other words, not only research universities deserve consideration, but also universities that are regionally important or those targeted at widening access to higher education with a view to involving a wider cohort of young people. Thus, one 'unwanted consequence' of global league tables is that **HEIs with other missions than that of being top research universities may have to re-justify their** profile at a time when mission differentiation is at the top of higher education agendas across Europe.

8. Descriptions of the methodologies of most global league tables are simplified and rarely allow a reader to follow the actual calculation of the scores of individual indicators and the composite final score. In order to be able to follow the calculations requires accessing more specialised sources, where the information is more sophisticated and would be of little help to an interested curious user who simply wants to understand where the numbers come from.

Combining indicators into final score – simply a calculation?

9. It should be noted that various indicators have different dimensions and denominators e.g. publication count, staff numbers, citations per academic etc. Before combining the scores of all individual indicators into an overall composite score, indicator scores are therefore treated mathematically in order to make them dimensionless. This means that **the published indicator scores are usually not the indicator values themselves but something else**, in most cases the proportion between the indicator value of the university in question and the university that has the greatest indicator value[5]. The league tables usually do not indicate in the column headings that the number provided is not the indicator value itself but rather a result of further mathematical operations. As a result, the scores in the league tables can lead to misunderstandings.

[5] For instance, if the indicator 'publications per academic' for university X has a value 57, it does not mean that each academic of university X publishes 57 publications per year. Instead, it means that the publication productivity of university X is 57% of the university which has the greatest producivity.

10. 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**. In other words, the composite score reflects the ranking provider's concept of quality. The above considerations demonstrate why rankings producing league tables cannot, in principle, be 'objective'.

11. Indicators used by rankings may use absolute values (count of publications, citations, students, staff members, etc.) or relative values (publications per staff member, citations per publication, funding per student, etc.). This simple aspect should be taken into account when analysing ranking results. 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. Examples of rankings predominantly using absolute numbers are, for instance, ARWU and Reitor. 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.

12. Current high positions of the top universities cannot be taken for granted. On the contrary, highly ranked universities have to make great additional efforts in order to maintain their current high positions, because their rivals evolve as well (CHERPA, 2010).

How far can we trust the indicators?

13. Overall, global university **rankings reflect university research performance far more accurately than teaching**. The bibliometric indicators, which are used for measuring research performance in most rankings, also have their biases and flaws, but they still are direct measurements.

14. Existing **indicators on teaching are all proxies** and their link to the quality of teaching is indirect at best. One extreme is measuring the quality of education by the number of Nobel Prize winners among the university's graduates (ARWU) – this indicator can be considered as being linked to the quality of education, but in a very special and rather indirect way. Judging teaching quality using staff/student ratios alone without examining teaching/learning itself (THE-QS) is another extreme. Moreover, it has been proven that staff/student ratios can be easily manipulated. Indicators such as teachers' salaries or time to degree do not actually measure quality. The time-to-degree indicator addresses an important issue in some countries, but is hardly seen as a valid indicator in others. It is not clear whether a much shorter time to degree or high graduation rates are signs of high quality or rather of low requirements. Those indicators may also open the door to manipulation.

15. In the CHE University Ranking, the indicators on teaching are selected with a view to helping potential students choose appropriate HEIs. The information used is largely taken from student satisfaction surveys – which work well for the purpose, especially as the CHE ranking does not produce league tables.

16. As regards bibliometric indicators used for measuring research performance, the natural sciences and medicine bias has been apparent since the first ARWU ranking was published in 2003. Natural sciences and medicine are favoured by all rankings based on bibliometric indicators – the ISI 21 'broad subject areas' are mainly sub-areas of natural sciences and medicine, while social sciences are underrepresented and humanities are simply ignored. At the same time, various areas have different publication and citation cultures. There are more publications and more citations per publication in natural sciences and especially in medicine, in particular because the main citations databases – WoS and Scopus – have little coverage of books.

    Attempts have been made to compensate for the field bias. Field normalisation is carried out through dividing the number of citations of an article by the expected average number of citations in the same field and year.

    Two citation impact indicators have been developed: the field normalised citation number (CPP/FCSm), the "crown indicator" of the Leiden Ranking and, more recently, the mean normalised citation number (MNCS). The calculation of those two indicators from citation data is described in detail in the main text of the survey. From a calculation point of view, the main difference lies in the sequence of mathematical operations. The CPP/FCSm indicator is calculated by first summing up values of citation counts per article and, separately, the average number of citations (in the same field in the same year), and then dividing the first sum by the second. The MNCS indicator, however, is calculated in reverse sequence, by dividing the citation number of each article by its particular average citation number, and then summing up all the results.

    It can be demonstrated that the CPP/FCSm indicator is naturally biased towards older publications – because older publications have, by definition, accumulated more citations. In addition, summing up citation numbers of articles in all possible fields before division somewhat blurs the outcome. In the case of the MNCS indicator,

there is no problem with the older publications and also the meaning of the outcome seems clearer. However, a new problem emerges: the newest publications have accumulated few citations and the world average citation numbers are therefore not reliable, making the result of the indicator unstable. For this reason, a modified MNCS2 indicator was proposed in 2010 which leaves out the most recent publications (of the last year).

Although one issue is thus removed, a new one is created. And, after all, these attempts to improve the methods of calculation in no way tackle the main problem, which is that citations of books or compendia of articles are still not considered.

17. Regarding the journal citation impact factor itself, it is important to note that, especially in social sciences and humanities, expert rankings do not correlate very well with impact factors (AUBR, 2010). In the above fields and in engineering, other sources, such as books and proceedings, are important as well. A warning on this issue can even be found on the Thomson-Reuter's website, which states that 'the impact factor should not be used without careful attention to the many phenomena that influence citation rates'[6].

18. **Peer review bias.** The term 'peer review' itself is ambiguous as it is used to denote quite different processes in quality assurance (QA) and rankings. In QA of both research and teaching, the term 'peer review' is used for assessment by (usually visiting) peers, which involves rigorous procedures. By contrast, in rankings, 'peer review' exercises are usually no more than reputation surveys. In the THE-QS ranking, even if a large number of academics have been approached, only some 5% actually answered. Secondly, at least in the case of the THE-QS-based ranking, the 'peers' are not in fact nominating the universities *they* consider excellent – they are restricted to pre-prepared lists, from which many universities and even whole countries have been omitted. Thirdly, there is evidence that the opinion of 'peers' can be influenced by the reputation that an institution has already built up (AUBR, 2010).

19. **Language bias and regional bias.** It has been noted since the publication of the first world rankings that global rankings favour universities from English-language nations because non-English language work is both published and cited less. A recent study by the Leiden Ranking team has shown that the citation impact of publications of French and German universities in French or German, respectively, was smaller than the citation impact of publications of the same universities published in English (van Raan et al., 2010).

Improving quality or improving ranking positions?

20. In an attempt to improve their positions in the rankings, universities are strongly **tempted to improve their performance specifically** in those areas which are measured by ranking indicators.

21. There have been cases where, rather than improving performance, data have been manipulated, for instance:

- merging universities just to get onto league tables
- number of applications to university
- standardised test scores of applicants
- number of academic staff
- student/staff ratio (it has been demonstrated that using different definitions of staff and students, the ratio could be anything between 6:1 to 39:1) (Baty, 2009)
- faculty salary
- reputation survey by students (by directly telling students to lie)
- even bibliometric indicators may be flawed due to the manipulation of data (AUBR 2010: 13).

How can rankings be improved?

22. Ranking providers are trying to **improve the methodology** they use. However, the improvements are often technical rather than conceptual. For instance, it is important to use field normalised data, and, in this sense, the new mean normalised MNCS2 indicator does indeed improve the mathematics compared to the previous 'crown indicator' CPP/FCSm[7]. However, this is of no help to humanities, which remain ignored by nearly all of the bibliometric indicators used in global league tables. Improving the calculation methods is not enough; rankings should make efforts to cover all research areas on an equal basis.

23. A number of university rankings claim that they **help students** to make their choices. Rankings do have the potential to help students choose the appropriate university in their home country or abroad. However, few of the existing league tables are currently able to do so.

[6] http://thomsonreuters.com/products_services/science/free/essays/impact_factor/
[7] See description of Leiden Ranking.

One of the few examples is the CHE Ranking. Generally speaking, to serve potential students, most of the rankings first need to choose appropriate indicators and provide substantially more explanations on what the scores of the indicators actually mean.

24. Now that rankings attract a great deal of attention from the general public, politicians included, there is a **demand for more 'democratic' rankings**. So far, the global league tables indicate a few hundred universities which are the 'best' in the world. In so doing, they have created problems for the thousands of 'normal' universities which simply do their job, such as training specialists for the labour market and conducting fundamental or applied research. The current rankings disease seems to have created a need to be ranked, because 'if you are not in the tables – you don't exist'. It should be possible to change the rankings substantially to allow more HEIs to be 'in'. This is especially important for those institutions that have been created with a special function, for instance to serve the region in which they are located or to provide higher education to adult students or those working alongside their studies. Including more universities could be seen as a way of recognising the important contribution of those well-functioning institutions that suffer from the 'unwanted consequences' of rankings.

25. Nationally, rankings foster the acquisition and publication of **reliable data** on higher education. In an international context, rankings encourage the adoption of agreed definitions for those aspects on which data is collected. The results of global rankings can lead to both national debates and a focused analysis of the crucial factors involved which, in turn, can bring about (hopefully) positive policy adjustment.

26. Most global league tables also publish lists concerning the '**performance' of countries**. These comparisons are made by counting each country's universities in the list of top universities, usually assigning a different number of points depending on whether the university appears in the Top 100, Top 100-200 or following top hundreds. The leading countries in the published lists then are the USA, the UK, Germany and France. However, if the published lists are 'normalised' by dividing the number of top universities by the number of inhabitants, new leaders appear, such as Switzerland, Sweden, Finland and Denmark (Salmi, 2010).

Recent developments in international rankings

27. Existing rankings, except possibly multi-indicator rankings, **cannot provide a diagnosis of the whole higher education system**, as they usually concern the top research universities only. Also, current global rankings can provide little useful information on issues such as the quality of teaching and learning, accessibility, regional involvement, involvement in lifelong learning, cost efficiency and others, simply because the indicators used do not cover such issues. The EU University-Based Research Assessments (AUBR), U-Map, U-Multirank and AHELO have been launched as attempts to develop international transparency tools for all HEIs, although they all have their limitations.

28. The AUBR working group carried out an analysis of the strong and weak elements of various research indicators, and their suitability for various assessment purposes, working out a methodology for the **assessment of university-based research**. The conclusions of the AUBR working group on various indicators are useful when analysing the global university rankings.

29. U-Map has been developed to classify all European HEIs **regardless of the institution type, focus etc.** and it reflects the variety of missions and profiles of European higher education institutions, without providing a final score. U-Map uses indicators that characterise the focus and intensity of various aspects of the work of HEIs rather than performance, impact or quality. U-Map indicators cover teaching level and subject focus, student body, research intensity, knowledge exchange, international orientation and regional involvement. U-Map has two visualisation tools: one to find higher education institutions which fit the characteristics set by the user, and one which allows the detailed comparison of up to three selected HEIs.

    Lack of internationally comparable data is a challenge to U-Map. Common definitions have yet to be developed to enable comparison of data from different countries. Until comparable data from across the EU has been collected, U-Map will have to rely on national and institutional data, and will therefore be more appropriate for comparing institutions within a country rather than internationally. U-Map is still being tested and data are being pre-filled by HEIs from volunteering countries.

30. U-Multirank is planned as a multidimensional ranking including all aspects of an HEI's work – education, research, knowledge exchange and regional involvement. No final score of a HEI will be calculated, but, until now, it is unclear how third parties will be prevented from turning the ranking results into a league table (Boulton, 2010). U-Multirank will utilise data from Thomson Reuters for its

bibliometric indicators. The other indicators will be based on self-reported data by HEIs on students, teachers, and research (except publications/citations). A large variety of data will be taken from student satisfaction surveys.

31. The issues related to internationally incompatible data may concern U-Multirank even more than U-Map. But for U-Multirank, as a ranking, there are further pitfalls. Using self-reported data may be safe in the case of U-Map as a classification tool. In the case of U-Multirank, even though it does not combine indicator scores into a final score, there are risks that self-reported data can be manipulated to score better in particular indicators. Relying on student satisfaction surveys is clearly preferable to drawing on reputation surveys answered by staff (which are not used in U-Multirank). However, the student satisfaction surveys may be less reliable in international comparisons, as the students are grading HEIs numerically, although grading traditions in different countries may differ. U-Multirank is still in its pilot phase, but seems to be moving towards a methodology of ratings rather than rankings. Data collection appears to be one of the most difficult issues.

32. 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. In these initial phases of the project, the developers have yet to find answers to a number of questions, including whether it is possible to develop instruments to capture learning outcomes that are perceived as valid in diverse national and institutional contexts.

Is self-regulation sufficient?

33. In autumn 2010, the International Rankings Expert Group (IREG) announced that it would start a rankings audit exercise. The audit will be carried out using 20 criteria based upon the Berlin Principles. Responsibility for the ranking audit lies with the Executive Committee of the IREG Observatory, which will also nominate the members of each audit team. The IREG Executive Committee has a mixed composition of ranking providers and experts who have followed developments in rankings.

    The IREG ranking audit procedure is modelled on higher education quality assurance procedures: a self-evaluation report is produced on the audited ranking based on a questionnaire; then the audit team performs an on-site visit; after the visit, the audit team compiles an audit report to be approved by the IREG Observatory Executive Committee. Positively audited rankings will be awarded an "IREG approved" label. IREG is made up of two categories of specialists: those who research rankings but do not produce them, and the ranking providers themselves.

    Meeting the requirements of the Berlin Principles is not easy and most, if not all, existing rankings would have to make changes in order to genuinely comply with them. It is of course too early to assess the IREG ranking audit. However, in order to be credible, the envisaged audit exercise would need to include independent, external experts bringing critical voices to the process.

I. INTRODUCTION

Purpose of the report

The purpose of this EUA report is to inform universities about the methodologies and the potential impact of the existing most popular global rankings, as well as other international efforts under development whose purpose is to rank, rate or classify university performance.

It was decided to concentrate only on international rankings so that the report is useful for all EUA members. National rankings have somewhat different features from the international ones. One difference is that national rankings are more often established and run by national governments, although national rankings can also be run by media or independent agencies. National rankings may be created to inform domestic students about the best study possibilities in their home country. However, especially if rankings are prepared by governments, the aim may be to use ranking results for higher education policy making, including allocation of funding.

Another difference is the selection of indicators: some types of indicators, such as various financial indicators, may work well in a national ranking but not so well in an international one due to a lack of internationally comparable data.

This review has been compiled following some agreed principles:

- The review should address the most popular global university rankings as well as major international developments involving attempts to measure performance (rankings, ratings, classifications) relevant for European universities. As indicated above, the review does not concern national rankings.
- This review is not aimed at judging or ranking the rankings themselves. Instead, it aims to provide universities with an analysis of the methodologies of a selected set of global rankings, with a view to helping readers understand what is actually measured, how the ranking results are calculated and how the results should be interpreted.
- Only publicly available and freely accessible information on each ranking was used to describe the methodology of each ranking and this was a deliberate choice. As a result, no fact finding was done through surveys or interviews with the ranking providers. Most rankings publish simplified descriptions of their actual methodology.
- This approach required deeper investigations of the ranking websites and publications of the ranking providers and this allowed us to see how transparent each ranking is.
- Efforts were made to determine what is said to be measured, what was actually measured, how the scores for individual indicators and, where appropriate the final score, are calculated, and what the results actually mean.
- We believe that the ranking methodology analysis will provide useful information to universities at a time when rankings are increasingly influencing higher education policies and public opinion.

Brief summary of the history of rankings

The first nationwide university ranking was published in the United States in 1983 by US News and World Report. However, classifications and specialised university rankings with a narrower focus had already been compiled in the US since 1870, see table 1 (Salmi & Saroyan, 2007).

Table 1. Chronology of ranking activities in the United States, 1870-1982

1870-1890	The Commission of the US Bureau of Education begins publishing an annual report of statistical data, classifying institutions.
1910	The Association of American Universities urges the US Bureau of Education to reinstate classifications.
1910-1933	James Catelli, one of America's first psychologists, professor at the University of Pennsylvania and then Columbia, publishes 'American Men of Science' in which he ranks institutions on the basis of the number of eminent scientists associated with an institution either as a student or a faculty member, and factors in the ratio of scientists at a given institution to the total number of faculty.
1925	Raymond Hughes, president of Miami University and later chair of the American Council on Education and its Committee on Graduate Instruction publishes 'A Study of the Graduate Schools of America' in which he uses reputational ranking of 26 disciplines in 36 institutions.
1957	Chesley Manly of the Chicago Tribune publishes six different rankings: ten best universities, co-educational colleges, men's colleges, women's colleges, law schools and engineering schools.
1959	Hayward Keniston of the University of Pennsylvania publishes reputational ranking of 15 universities in a range of disciplines.
1966	Allan Cartter of the American Council of Education publishes 'An Assessment of Quality in Graduate Education' which ranks 106 institutions.
1973-1975	Blau and Margulies conduct reputation ranking of professional schools.
1982	The US National Academy of Science commissions an assessment of research and doctoral programmes in the United States.
1982	Rankings begin to be extended to undergraduate education (e.g. Fiske Guide to Colleges, 1982; US News and World Report, 1983; etc.).

Source: table reproduced from Salmi & Saroyan, 2007

The era of global rankings started with the publication of the first results of the Shanghai Jiao Tong University ranking called 'Academic Ranking of World Universities' (hereafter referred to as ARWU) in 2003. Publication of the results of the first round of ARWU in 2003 "stirred the fire" (van Raan, 2005). The results shocked the world and particularly Europe, as, in the ARWU ranking, US and UK universities strongly dominated the Top 20 and Top 100 lists. The creation of the Times Higher Education Supplement World University Ranking in 2004 (later Times Higher Education Ranking, hereafter referred to as THE) was, in a way, a European answer to ARWU. Since then, more global university rankings have emerged.

Implications of the rankings in brief

In recent years, university rankings have gained in importance around the world. As long as only national university rankings existed, they were popular and important in some countries, while other university systems did not pay much attention to them. Rankings went truly international in 2003 when Shanghai Jiao Tong University published the results of the first global university ranking. The importance of rankings seems, since then, to have grown exponentially.

Rankings have always been controversial. Those who compile and publish rankings usually claim that rankings are 'objective' and that the position of a university in a ranking table corresponds to its 'quality'. Critics argue that the result of ranking depends strongly on the choice of indicators and weights assigned to the indicators and that, because it is difficult (if not impossible) to measure quality itself, rankings use various more or less distant proxies and claim that these measurements represent quality or excellence itself. As stated by Marginson and van der Wende, "A better approach to rankings begins from the recognition that all rankings are partial in coverage and contain biases, and that all rankings are purpose driven" (Marginson & van der Wende, 2007).

At the same time, society, including politicians, often like to see universities arranged in a neat league table according to the results attained through indicators, and truly believe that, in much the same way as for sports teams, each university in the table is 'better' than the one below and 'not as good' as the one above. For several years now, rankings have been widely discussed: in the media, among politicians, within the academic community and in society at large. Growth of interest in the results of rankings has changed the context in which universities function: for a university to be seen as 'successful' it has now become necessary to improve performance specifically in those aspects that are measured in rankings.

National or regional higher education policy makers increasingly consult ranking results when deciding on the allocation of resources or the structure of higher education systems.

University rankings themselves are intensely debated. While supporters of global university rankings argue that rankings improve transparency and allow students to make informed choices, critics say rankings do not address the various important functions of higher education, that the indicators used in rankings measure distant proxies rather than quality itself, and that rankings have serious biases (see e.g. van Raan, 2005). The strongest of the biases is that of favouring research in natural sciences and medicine, under-representing engineering and social sciences, and completely or almost ignoring the humanities. Favouring English language publications is often mentioned as a further bias (ibid.).

Whether one likes global university rankings or not, they are here to stay. Indeed, in an interview in *The Australian* on 12 January 2010, S. Marginson said "we are likely to see more university rankings, not less. But the good news is they will become increasingly specialised" (Hare, 2011).

It would be naïve to imagine that the media will ever give up a tool such as the global university rankings, which attract thousands of readers when the new results are published and which allows suspense to be maintained over an entire year, by publishing tiny snippets of information about minimal changes in ranking methodologies. The general public as well as politicians will always like easily readable tables, which clearly state which universities are the very best in the world.

As regards universities, they are often either flattered or ashamed depending on their current position in the league table or the change of position since the previous year. There are forces both inside and outside the university encouraging it to make every effort to improve its position in the rankings or simply be included in the league tables at all costs. As S. Marginson puts it, "Rankings are the meta-performance indicator, with a special power of their own. Rankings are hypnotic and become an end in themselves without regard to exactly what they measure, whether they are solidly grounded or whether their use has constructive effects. The desire for rank ordering overrules all else" (Marginson 2007).

At the same time the benefits of rankings are often stated, especially by ranking providers themselves. For example, rankings can inform a student's choice of institution or promote a culture of transparency.

Rankings strengthen competition among and often bring about policy change in universities, which strive to improve their standing in the league tables. They provide simple and easily readable information and are therefore beginning to be used as a basis for funding allocations to universities, as well as for developing national or regional higher education policies.

At the same time, a great variety of national and international rankings is being created for various purposes, using different indicators and often choosing various proxies instead of parameters that are difficult or impossible to measure.

University rankings have been much criticised, especially after the global rankings appeared. As regards the indicators used for rankings, there have been criticisms of the flaws and biases, such as favouring research but poorly reflecting teaching; completely ignoring institutional diversity and thus favouring research universities; completely or partly ignoring publications in book form; and, finally, field, language, geographical and other biases.

Other criticisms are more basic and usually concern the transparency of the rankings themselves. For instance it is not always easy to determine the identity of the ranking provider, the aims of the particular ranking, or the target groups: is the ranking targeted at students seeking the 'best' university, at the general public or at higher education policy makers? Or has the ranking perhaps been created to help universities improve their performance in particular areas?

Other complaints concern the transparency of ranking methodologies: the meaning of indicators has not been explained or are described so generally that it is impossible to understand:

- what is actually being measured: is it a straight measurement or is it a proxy that is measured, and if so, what is the actual relationship between the measurement result and the category that it is claimed to measure (for instance, measuring the staff/student ratio and calling it 'educational quality')?

- how is the indicator value calculated from the raw data: for example, is it clear that the number 26 as the 'publications' indicator does not mean that staff publish 26 papers per year, but that the publication intensity in the university in question is 26% of the publication intensity of the 'best' university in this indicator?

- how is the final score of the university calculated from indicator results: the weights of individual indicators are not necessarily publicised.

Berlin Principles on the Ranking of Higher Education Institutions

To improve the situation, it became clear that some common principles were needed that those who produce rankings could follow. A set of guidelines was drawn up by the International Ranking Expert Group (IREG)[8] in 2006. They are called the 'Berlin Principles on Ranking of Higher Education Institutions' (IREG, 2006), hereafter referred to as the 'Berlin Principles'. The main Berlin Principles are as follows (from IREG, 2006):

- **With regard to purposes and goals**, rankings should: be clear about their purpose and their target groups, recognise the diversity of institutions and take the different missions and goals of institutions into account, provide clarity about their information sources. They should specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked.

- **With regard to design and weighting of indicators**, rankings should: be transparent regarding the methodology; choose indicators according to their relevance and validity; measure outcomes in preference to inputs whenever possible; and make the weights assigned to different indicators (if used) prominent and limit changes to them.

- **With regard to collection and processing of data**, rankings should: pay due attention to ethical standards and good practice recommendations; use audited and verifiable data whenever possible; include data that are collected with proper procedures for scientific data collection; and apply measures of quality assurance to ranking processes themselves.

- **With regard to presentation of ranking results**, rankings should: provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed; be compiled in a way that eliminates or reduces errors in original data; and be organised and published in a way that errors and faults can be corrected.

There is no doubt that the Berlin Principles are a good guide, and ranking providers often claim that they comply with them. Reality, however, looks somewhat different. For instance, a group of researchers at Minnesota University quantified the Berlin Principles and rank the rankers themselves according to their congruence with the best practices described in the Berlin Principles (Stoltz et al., 2010). Using scores from 1 (no congruence), 2 (poor congruence), 3 (fair congruence), 4 (good congruence) to 5 (excellent congruence), the Minnesota University group of researchers ranked 25 European university rankings. 13 out of 25 rankings in this exercise failed to meet at least score 3 (fair congruence), among them the Times Higher Education (THE-QS) global ranking, which scored only 2.25. The researchers also found that the rankings basically fell short in their methodologies, showing closer congruence with the Berlin Principles in relation to transparency and customer friendliness.

But do league tables provide the information students want? It is one of the principles of 'good ranking' as embodied in the Berlin Principles that rankings should be geared towards their target group, and it can be safely said that, at least until recently, this aspect was given too little attention. It was simply assumed that whatever indicators were available must be relevant, and that this would apply to all groups of readers of rankings (King, Locke et al. 2008).

In autumn 2010, the IREG announced that it would start a rankings audit exercise. The audit is to be carried out using

[8] IREG was established in 2004 by UNESCO CEPES and Institute for Higher Education Policy in Washington.

20 criteria based on the Berlin Principles. Responsibility for the ranking audit lies with the Executive Committee of the IREG Observatory, which will also nominate the members of each audit team. The IREG Executive Committee has a mixed composition of ranking providers and experts who have followed developments in rankings.

The IREG ranking audit procedure is modelled on higher education quality assurance procedures: a self-evaluation report is produced on the audited ranking based on a questionnaire, and the audit team then performs an on-site visit. After the visit, the audit team compiles an audit report to be approved by the IREG Observatory Executive Committee. Positively audited rankings will be awarded an "IREG approved" label.

It is not easy to meet the requirements of the Berlin Principles. And since it has been demonstrated that existing rankings more often than not fail to comply with some of them, the involvement of independent experts in the audit procedure would greatly enhance its credibility.

II. METHODOLOGIES OF THE MOST POPULAR GLOBAL RANKINGS

There are various ways in which rankings can be grouped: according to their purpose, the parameters measured, the presentation of the results or intended impact. For the purposes of this report, the grouping of the international rankings covered by the report is as follows:

1. Academic rankings with the main purpose of producing university league tables
- 1.1 Academic Ranking of World Universities (ARWU) – Shanghai Ranking Consultancy[9]
- 1.2 THE World University Ranking – Times Higher Education
  - 1.2.1 in cooperation with Quacquarelli Symonds (until 2009)
  - 1.2.2 in cooperation with Thomson Reuters
- 1.3 World's Best Universities Ranking – US News & World Report in cooperation with Quacquarelli Symonds
- 1.4 Global Universities Ranking – Reitor (Peŭmop)

2. Rankings concentrating on research performance only (with or without league tables)
- 2.1 Leiden Ranking – Leiden University
- 2.2 Performance Rankings of Scientific Papers for World Universities – Taiwan Higher Education Accreditation and Evaluation Council
- 2.3 Assessment of University-Based Research – European Commission

3. Multirankings – university rankings and classifications using a number of indicators without the intention of producing league tables
- 3.1 CHE University Ranking – Centre for Higher Education Development/*die Zeit*
  - 3.1.1 CHE University Ranking
  - 3.1.2 CHE Excellence Ranking
  - 3.1.3 Other CHE Rankings
  - 3.2 U-Map classification – CHEPS
- 3.3 European Multidimensional University Ranking System (U-Multirank) – EU funded project

4. Web rankings
- 4.1 Webometrics Ranking of World Universities

5. Benchmarking based on learning outcomes
- 5.1 Assessment of Higher Education Learning Outcomes Project (AHELO) – OECD

For each ranking, the following methodological issues are analysed, in as far as they are relevant:

- Indicators used
- What is actually being measured, and which proxies are used
- Calculation of indicator values and overall scores where appropriate
- Change in methodology over time and its impact
- Strengths, weaknesses and "peculiarities" of the particular ranking
- Other products offered by the ranking – additional analyses produced (by subject, comparisons of countries, information for students, individual rankings)
- Elements of particular note regarding the particular ranking.

It should be noted that the more recent rankings, particularly U-Map, U-Multirank and AHELO, were still in their development stage at the time this report was prepared. Attempts have been made to analyse these developments and provide information that may be of interest to EUA. These developments will be covered in more detail in the next report in 2012.

[9] Initially Shanghai Jiao Tong University.

1. International rankings producing league tables

1.1 The Academic Ranking of World Universities (ARWU) – Shanghai Ranking Consultancy

The story of ARWU actually begins in 1998 when Shanghai Jiao Tong University (SJTU) was selected by the Chinese government to be among the first group of nine universities in the '985 Project'. This project was set up in response to a statement by Jiang Zemin (the then President of the People's Republic of China) that China must have a certain number of top, world-class universities (Liu, 2009). From 1999 to 2001, a SJTU team worked on a project to benchmark top Chinese universities with US research universities 'in order to find out the gap between Chinese universities and world-class universities' (Liu & Cheng, 2005). According to Liu (Liu, 2009), after the team submitted its report to the Chinese Ministry of Education and it was published, Chinese and foreign reactions to the report recommended making it into a real ranking of world universities. This ranking was first published in 2003 and has been updated annually ever since. SJTU supported the publication with ARWU until 2009 when an independent consultancy was established.

A list of criteria and indicators used in ARWU is provided in Table 2.

Table 2. List of the 21 broad subject areas as defined by ISI[12]

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

Which universities are considered for the ARWU ranking?

Universities that have Nobel laureates, Fields medallists, Highly Cited Researchers, or papers published in Nature or Science are included (Liu, 2009) in the ranking. In addition, universities with a significant number of papers indexed by Science Citation Index-Expanded (SCIE) and Social Science Citation Index (SSCI) are also included; see the published ARWU methodology of the 2010 ranking (ARWU 2010).

On this basis, ARWU in no way seeks to compare all the universities in the world – its criteria are targeted at the world's top *research* universities only. ARWU picks up only around 1000 of the 17,000 universities in the world, of which the first 500 are ranked in the league table.

Areas covered, indicators used and proxies

ARWU ranks universities according to their success in four areas; see Table 2 below for the subject areas.

- *Quality of education*[10]
- *Quality of faculty*[11]
- *Research output*
- *Per capita performance of the university*

What is measured and what proxies are used to estimate the strength of a university in each of those areas?

The proxy to measure *Quality of education* is the number of alumni winning Nobel prizes in the sciences or Fields medals in mathematics.

[10] See Table 2 to check what *Quality of education* means in ARWU.
[11] See Table 2 to check what *Quality of faculty* means in ARWU.
[12] Descriptions of the 21 categories can be found at: http://www.isihighlycited.com/isi_copy/Comm_newse04.htm

Quality of faculty is estimated on the basis of two proxies:

- number of staff winning Nobel Prizes and Fields Medals, and
- number of staff included in the lists of most highly cited researchers in 21 *broad subject areas* (see Table 2).

The list of areas alone demonstrates that the areas of natural sciences, medicine and engineering dominate the citation indicator, leaving social sciences far behind. According to Liu (2009), arts and humanities are not ranked because of the technical difficulties involved in finding internationally comparable indicators with reliable data. In the case of cross-disciplinary subjects, this has also proven challenging, due to their interdisciplinary character.

Research output is measured using:

- Number of papers published in *Nature and Science* over the last five years and
- Number of papers indexed in the Thomson Reuters Science Citation Index-Expanded and Social Science Citation Index in the year preceding the year of the ranking compilation (note that only publications in the form of 'Articles' and 'Proceedings Papers' are considered).

Per Capita Performance is not measured separately but calculated from the values of the indicators described above, and using numbers of academic staff drawn from national data. It should be noted that Per Capita Performance is the only ARWU indicator that takes into account the size of the institution. This means that ARWU reflects the overall strength of a university and that small but excellent institutions have less of a chance of figuring in ARWU.

Table 3. Criteria, indicators and weights used in the ARWU Ranking

	Indicator	Code	Weight
Quality of Education	Alumni of an institution winning Nobel Prizes and Fields Medals	Alumni	10%
Quality of Faculty	Staff of an institution winning Nobel Prizes and Fields Medals	Award	20%
	[Top 200] highly cited researchers in 21 broad subject categories	HiCi	20%
Research Output	Papers published in Nature and Science[13]	N&S	20%
	Papers indexed in Science Citation Index-expanded and Social Science Citation Index	PUB	20%
Per Capita Performance	Per capita academic performance of an institution	PCP	10%
Total			**100%**

Data sources

For all ARWU indicators, data is gathered from third parties, including the official site of the Nobel Prize, the International Mathematical Union for Fields medals and several Thomson Reuters websites for citation and publications. The numbers of academic staff for the *Per capita performance* indicator is gained from national sources. One could argue that this latter dataset is collected nationally from the universities themselves. Usher and Savino (Usher & Savino, 2006) therefore consider data on academic staff as 'university data'.

Calculating indicator values and transforming them into scores

It should be noted that an *indicator value* is distinct from the *score* of an indicator.

The indicator *value* is the result of the actual measurement, e.g. if University X has 352 publications in Nature or Science (N&S), the value of the indicator is 352.

Indicator scores in ARWU are calculated by dividing the actual indicator value by that of the university that holds the highest value and multipling by 100.

> *Example*. University X has 352 publications in Nature or Science, but university Y holds the best result - 398 publications in Nature and Science (N&S). The score of University X in indicator N&S will be $N\&S_X$ = 352/398*100=88.4.

This also means that the position in the league table does not tell readers what the indicator values are. Thus, if University Z

[13] For institutions specialised in humanities and social sciences, N&S is not considered, and the weight of N&S is relocated to other indicators.

has scored 9.8 in Per Capita Performance, this does not mean that a staff member of University Z has on average 9.8 publications, but rather that the success of University Z in indicator PCP is 9.8% of the university that scored best in this respect.

Alumni indicator. This value is calculated from the number of Nobel Prize or Fields medal laureates among alumni irrespective of the level of degree awarded to the prize winner in the university in question. Prize winners who graduated before 1991 gradually count for less: thus prize winners who graduated between 1981-1990 are counted as 0.9, those who graduated between 1971-1980 as 0.8 and so on, down to 0.1 for those who graduated between 1901-1910.

This indicator is intended to reflect 'Quality of teaching', which it does in a very specific and rather limited way.

Award indicator. This value is calculated from the number of Nobel Prize or Fields medal laureates among university staff. Prize winners who received the award before 2001 count for less: prize winners between 1991-2000 are counted as 0.9, those between 1981-1990 as 0.8 and so on down to 0.1 for prize winners between 1911-1920. If the prize winner is affiliated to several universities, the value is shared between the institutions; if the award has been shared among several winners, the value is divided between them proportionally according to the share of award.

HiCi indicator. This value is calculated in relation to the number of highly cited researchers in the 21 subject categories listed in Table 2. ISI identifies highly cited researchers[14] and compiles lists[15] of the top 250 most cited researchers in each of the 21 subject categories.

Nature and Science indicator. This indicator value is the number of papers published in Nature and Science over the last five years (ending with the year preceding the year of the ranking). For articles with several authors, 1 is assigned for the corresponding author, 0.5 for the first author (or second author affiliation, if first author affiliation is the same as corresponding author affiliation), 0.25 for the next author, and 0.1 for any further authors). Only publications of the types 'Article' or 'Proceedings Paper' are taken into account.

Publications indicator. The indicator value is the number of papers indexed in the Science Citation Index-Expanded and Social Science Citation Index in the previous year. Only publications of the types 'Article' and 'Proceedings Paper' are taken into account. A paper indexed in the Social Science Citation Index is counted as two.

Per capita performance indicator. The indicator score is calculated as the weighted scores of the above five indicators divided by the number of full-time equivalent academic staff combined. However, if the number of academic staff for institutions in a country cannot be obtained, the weighted score of the above five indicators is used.

Weights of indicators and areas for the calculation of the final score. The weight for each indicator is shown in Table 3. While the number of alumni winning Nobel prizes in itself is quite a one-sided proxy, it is the only indicator in the Quality of Education area, so the overall input of Quality of Education in the total score is a mere 10%.

The combined weight of the *Quality of Staff* area is rather high at 40%. In order to achieve high scores in this area, a university's staff members must mainly be active in the fields of science or medicine.

The combined weight of *Research output* is also 40%. Half of this score comes from publications in *Nature and Science,* which again gives pre-eminence to those publishing in the hard sciences or medicine. There is an attempt to take account of universities specialising in the humanities or social sciences. In those cases, the *Nature* and *Science* indicator is excluded and the score is shared equally between other indicators. However, while this measure helps specialised schools of social sciences or humanities, for multidisciplinary universities, the N&S score still ignores success in the humanities and overlooks the social sciences.

The *Per Capita Performance* area has a lower weight in the final score – just 10%. However, it is noteworthy that this is the only indicator that takes into account the size of the institution.

Changes in the methodology over time

Comparing the ARWU methodology used in 2004, 2005, 2006, 2007, 2008 and 2009, as available on the ARWU website, no changes of methodology could be identified. It should be noted, however, that in their 2005 article Liu and Cheng use the wording *Size of Institution* instead of *Per Capita indicator.* This suggests that the methodology described in the early years may have been revised later on and that there may be some differences, although these are probably minor.

Presentation of the ranking and additional analysis produced

The ARWU league table is available on the internet and is presented as lists of Top 100 universities, Top 101-200, Top 201-300, Top 301-400 and Top 401-500.

Statistics. The same ARWU data are also presented in the form of statistics of world regions and countries, showing the number of universities of a particular world region or country which are among the Top 20, Top 100, 200, 300, 400 and 500 universities.

Field rankings are produced in the natural sciences and mathematics, engineering, life sciences, medicine and social sciences. The five indicators used in the field rankings are

[14] See http://isihighlycited.com/isi_copy/howweidentify.htm
[15] Lists of highly cited researchers by subject categories are available at: http://hcr3.isiknowledge.com/

similar to those of the ARWU university ranking, but they have different weights and are applied field-specifically[16] (see 'ARWU subject rankings'). The *Per capita* indicator is not used, but is replaced with the Top indicator in the field rankings. Another indicator – *Funding* – is used only for the engineering field.

The *Alumni* indicator has a weight of 15% and its value is the number of alumni who are winners of Nobel Prizes and Fields medals (for sciences and mathematics) in their respective field since 1951. This indicator is not applicable to the Engineering field.

The *Award* indicator has a weight of 10% and its value is the number of alumni who are winners of Nobel Prizes and Field medals (for sciences and mathematics) in their particular field since 1961. This indicator is also not applicable to the Engineering field.

The *Funding* indicator was introduced for the Engineering field only; its weight is 25% (the combined weight of the Alumni and Award indicators used for other fields). The funding indicator value is the total expenditure for engineering-related research.

The *Highly cited researchers* indicator has a weight of 25%. The value of this indicator is the number of researchers listed in the ISI Highly cited researchers' lists. The 21 broad subject categories are assigned to the field rankings in the following way: the Sciences and Mathematics field ranking covers the categories of Mathematics, Physics, Chemistry, Geosciences and Space Sciences; the Engineering field ranking covers Engineering, Materials Science and Computer Science; the Life Sciences field ranking covers Biology and Biochemistry, Molecular Biology and Genetics, Microbiology, Immunology, Neuroscience, Agricultural Sciences, Plant and Animal Science and Ecology/Environment; the Medicine field ranking covers Clinical Medicine, Pharmacology and, partly, Social Sciences General; the Social Sciences field ranking covers, partly, Social Sciences General and Economics/Business. The subject category Psychology/Psychiatry is not covered by the ARWU field rankings.

The *Publications* indicator has a weight of 25%. Its value in different field rankings is the number of papers indexed in the Science Citation Index-Expanded in the appropriate field. Again, only publications of the types 'Article' and 'Proceedings Paper' are taken into account.

The *Top* indicator has a weight of 15%. It indicates the percentage of papers published in the top 20% of journals in each field, according to their impact factors in Thomson Reuters Journal Citation Reports. The threshold was defined as 10% of the average number of papers by the top three institutions in each broad subject field. If the threshold of a particular field is less than 100, then 100 is used. If the number of papers of an institution does not meet the minimum threshold, the *Top* indicator is not calculated for the institution and its weight is relocated to the other indicators.

ARWU subject rankings. In addition to field rankings that cover several subjects, ARWU also offers subject rankings for the following selected subjects: Mathematics, Physics, Chemistry, Computer Science and Economics & Business.

The use of indicators is similar to those of the field rankings, but data is collected for each particular subject. It should be noted that in the case of Computer Science, Turing Awards in Computer Science are used for the *Alumni* and *Award* indicators.

Potential new products. According to the ARWU providers, they are exploring the possibility of providing rankings based on different types of universities with different functions, disciplinary characteristics, history, size and budget (Liu, 2009). Besides ARWU, the Shanghai Ranking Consultancy will also provide various global comparisons and in-depth analyses on research universities[17].

1.2 THE World University Ranking – Times Higher Education

The Times Higher Education World University Rankings was first published in 2004. In a way, it was an 'answer' to the Shanghai ARWU ranking that was first published in 2003.

The then Times Higher Education Supplement, which later became the independent Times Higher Education Ranking, used Quacquarelli-Symonds (QS) as its data collection and processing engine between 2004 and 2009. In 2009 the Times Higher Education (hereafter THE) announced that it was ceasing cooperation with QS and that a new cooperation was being established with Thomson Reuters (Baty, 2009). THE has since announced its new methodology for the 2010 World Universities Ranking. In the meantime, QS has started a new partnership with US News and World Report to set up yet another global ranking. The THE methodology for the period 2004-2009 and the methodology proposed for the 2010 ranking will therefore be described separately.

[16] The table of Field ranking indicators can be seen at: http://www.arwu.org/ARWUFieldMethodology2009.jsp
[17] See http://www.arwu.org/aboutARWU.jsp

1.2.1 THE-QS World University Rankings between 2004 and 2009

The stated purpose of the THE World University Rankings is "to recognise universities as the multi-faceted organisations that they are, to provide a global comparison of their success against the notional mission of remaining or becoming world-class" (THE-QS, 2009).

Research quality, teaching quality, graduate employability and international outlook are listed as the four pillars of the world class university and consequently represent the areas covered by the THE Ranking (ibid.).

The approach used by the THE World University Ranking strongly differs from that of ARWU. While ARWU concentrates on research outputs, in the THE Ranking, a substantial share of the final score (initially 50%, later 40%) comes from a 'peer review' of universities, which is actually an internet reputation survey of academics. A further 10% of the final score is the result of a survey among employers.

Which universities are considered for the ranking?

According to the statement above, the ranking considers those universities that are or are becoming world-class universities. Another statement claims that the "THE QS World University Rankings™ were conceived to present a multi-faceted view of the relative strengths of the world's leading universities", thereby confirming that the ranking considers the world's elite universities only (THE-QS, 2009)[18]. It is therefore hardly surprising that the methodology singles out only around "600 universities altogether and 300 in each of five broad faculty areas" (ibid.).

Areas covered, indicators used and proxies

The four areas covered by the indicators are as follows: two indicators – *Peer review* and *Citations per Faculty* are used to characterise research. The *Employer review* is used to characterise graduate employability. The only proxy used to judge the quality of teaching is the much criticised *Faculty/Student Ratio*. Two proxies are used to characterise the International outlook of universities: the proportion of *international staff* and the proportion of *international students*.

Peer review indicator. 'Peer review' in this case is not the expert visit to a university, which is what is understood by 'peer review' in quality assurance procedures. Here, *peer review* is an internet survey in which peers are asked to select up to 30 universities from a pre-selected list.

This indicator has the highest weight in the THE-QS rankings – 50% in the first ranking in 2004, and 40% from 2005-2009. According to the stated methodology: "the results are based on the responses to a survey distributed worldwide both to previous respondents and subscribers to two key databases: The World Scientific (www.worldscientific.com), from which 180,000 email addresses are drawn, and the International Book Information Service (IBIS)"[19]. The overall score of the indicator is compiled from five sets of results in the following five subject areas: Arts & Humanities, Engineering & IT, Life Sciences & Biomedicine, Natural Sciences and Social Sciences (ibid.).

The respondents are asked to indicate their areas of competence and the geographic region with which they were familiar. Since 2007, respondents have been prevented from selecting their own institution and since 2008, they have been asked separately about universities within and outside their country, in order to reduce any response bias by country (THE-QS, 2009).

There are several important issues regarding this indicator. First, the scores are based on a rather small number of responses: 9386 in 2009 and 6534 in 2008; in actual fact, the 3000 or so answers from 2009 were simply added to those of 2008 (see Table 4). The number of answers is pitifully small compared to the 180,000 e-mail addresses used.

Secondly, the academics surveyed were asked "to select up to thirty universities from our international list that you regard as producing the best research in (name of appropriate subject)"[20]. What are those 'international lists'? After the QS 2010 methodology was published on the web (i.e. after the split with THE), we could not find any publicly available information regarding who compiles these lists for each subject area. More importantly, what are the criteria for leaving out a great number of universities or whole countries? For instance, the lists of universities pre-selected by THE-QS in the five subject areas usually contained universities from only 25/26 European countries out of the 48 countries in the European Higher Education Area. In addition, the academics ('peers') who responded to the survey could not select universities that did not appear on the list, let alone those from countries that have been left off the list altogether. The QS 2010 methodology description[21] may shed some light on these issues.

According to the 2010 explanation, the selection of universities for the ranking came from four sources:

- success in domestic rankings tracked by the QS Intelligence Unit

[18] Retrieved on 29 December 2010 from: http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/simple-overview

[19] Retrieved from: http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/academic-peer-review

[20] Retrieved 15 September 2010 from: http://research.qsnetwork.com/qs_surveysystem/index.php?viewonly&order=normal&partnerset=0

[21] Retrieved 29 December 2010 from: http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/institution-inclusion

- institutions suggested by respondents to the Academic and Employer Reputational Surveys
- geographical balancing (no further explanation of criteria or principles)
- direct case submission – institutions addressing QS directly asking for inclusion.

The above, however, still fails to explain why the scrollbars from which peers and employers select the 30 top institutions in each field differ, or who decides which universities become part of each individual scrollbar list.

Employer review indicator. The Employer review operates in a similar fashion to the Academic Peer Review in that it is based on a global online survey. Respondents are again sourced through three key channels: QS's extensive corporate database, a network of partners with whom QS cooperates in its events, and participating institutions who submit a list of professionals with whom they work (THE-QS, 2009)[22], thus creating a new bias. Respondents were again asked to identify up to 30 universities which produce the best first-degree graduates (ibid.).

The questions regarding the *Employer review* are the same as those regarding the *Peer review* indicator: who pre-selects the universities for the lists from which the Employers choose, and how are they chosen? The fact that the total number of worldwide responses in 2009 was only 2336 has implications. In particular, such a small sample of world employers might simply not be aware of excellent universities in smaller countries, and especially in those countries where neither English nor another world language are spoken.

Citations per faculty indicator. In the 2004-2006 rankings, THE-QS ranking used citation numbers from Essential Science Indicators (ESI) of Thomson Reuters Web of Science. However, as of 2007 they switched to Scopus (now owned by Elsevier) "for a number of reasons, but principally due to broader journal coverage leading to results for a larger number of institutions" (THE-QS, 2009, section on *Citations per faculty* indicator[23]). The staff number used is the total FTE, rather than separate numbers of teaching and research staff. This is mainly due to different definitions and hence incompatible data from different countries.

Faculty/student ratio. The THE-QS rankings from 2004-2009 used the staff-student ratio as the only proxy for *Teaching quality.* The weight of this indicator was 20%. The more recent THE ranking providers quote criticisms of the use of the staff-student ratio and criticise it themselves (Baty, 2010a). In any event, few academics will agree that the quality of their teaching can be judged by the number of students in the class.

Total student numbers are calculated by adding together the respective undergraduate and postgraduate student numbers supplied. Where this data is unavailable or incomplete, total student numbers are used (THE-QS, 2009, section on Staff faculty ratio)[24]. Faculty numbers used are total FTEs, although the ranking providers themselves confess that "it would be ideal to separate the notions of teaching and research and use the former for calculating this indicator" (ibid.). In addition, it has been demonstrated that, because definitions of staff categories vary, universities can easily manipulate the result of this indicator. According to Baty, Marney Scully, Executive Director of policy and analysis at the University of Toronto, has shown how student-to-staff ratios of anything from 6:1 to 39:1 can be generated on the basis of the same data, simply by playing with the definitions (Baty, 2010b). Wishing to boost their own positions in the rankings, universities may be tempted to present their data in the most advantageous way.

International staff and *International student* indicators have a 5% weight each. It may appear straightforward to calculate the values of those indicators by simply dividing the numbers of international staff or students by the total staff FTEs or full-time student numbers. However, the publicly available information on the THE-QS ranking does not contain definitions of international staff and students. For instance, the definition of an international student can be linked to the student's country of origin, country of residence and/or country of previous education. In addition, international students can be those studying for a full degree or those who are spending only a study period abroad. International faculty in turn may be those employed by the university, visiting staff whose stay is funded from extra-budgetary sources, permanent staff or staff staying for a fixed period, etc. Since the numbers of international students and faculty are part of the information obtained from the universities themselves, nuances of definition may also have an impact on the results.

[22] See http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/employer-review
[23] See http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/citations-faculty
[24] See http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/student-faculty

Table 4. Indicators and their weighting in THE-QS Ranking in 2005-2009

Indicator	Explanation	Weighting
Academic Peer Review	Composite score drawn from peer review survey (which is divided into five subject areas). 9,386 responses in 2009 (6,354 in 2008).	**40%**
Employer Review	Score based on responses to employer survey. 3,281 responses in 2009 (2,339 in 2008).	**10%**
Faculty Student Ratio	Score based on student faculty ratio	**20%**
Citations per Faculty	Score based on research performance factored against the size of the research body	**20%**
International Faculty	Score based on proportion of international faculty	**5%**
International Students	Score based on proportion of international students	**5%**

Source: QS Quacquarelli Symonds (www.topuniversities.com). With permission.

Data sources

Over the period of 2004-2009, the THE-QS ranking has used several data sources.

Individual academics/researchers have been surveyed for the heavily weighted *Peer review* indicator that is used to characterise the university's reputation with regard to research.

Employer organisations have been surveyed to determine the reputation of universities with regard to employable graduates for the *Employer review* indicator.

Third-party data have been used for the *Citations per faculty* indicator, taken from Essential Science Indicators (ESI) of the Thomson Reuters Web of Science in 2004-2006, and Scopus (Elsevier) as of 2007.

University information. THE-QS ranking has often used data obtained from the universities themselves, such as numbers of full-time students (undergraduate and graduate), number of faculty (FTEs) as well as numbers of international students and staff.

Calculating indicator values and transforming them into scores

Peer review. The indicator values of a university are first calculated for each of the five subject areas: Arts and Humanities, Engineering and IT, Life Sciences and Biomedicine, Natural Sciences, and Social Sciences.

In this process:

- The peers' answers regarding universities within and outside of their own country were first recombined "to reduce any response bias by country" (THE-QS, 2009[25]).
- Regional weightings were then applied to each of the five subject areas "to ensure equal representation from three 'super regions' of the Americas; Europe, the Middle East & Africa; and Asia Pacific" (ibid.). It was not possible, however, to obtain publicly available information on the values of those regional weightings and their application.

The results in the five areas are then summed up.

Employer review. As is the case with the *Peer review* indicator, the peers' answers regarding universities within and outside of their own country are recombined and regional weightings applied.

Citations per faculty. The indicator value is calculated by dividing the number of citations by the total number of staff (in FTEs), without dividing staff into teaching and research staff (see above).

Student/faculty ratio. Although often called the *Student/faculty ratio,* the value of the indicator is the faculty/student ratio, i.e. the reverse of what is said. It is calculated by dividing the number of students (combined undergraduate and graduate) by the total number of staff (in FTEs), again without dividing staff into teaching and research staff.

International staff and *International student* indicator values are calculated as international student or, respectively, staff as a proportion of the total student or staff numbers. Definitions of international staff member or student are not provided.

Normalisation of results

Before 2007, the final score of a university in the THE-QS Ranking was calculated by dividing the university's score by that of the university that held the highest value and

[25] Retrieved from: http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/academic-peer-review

multiplying the result by 100 (for more details of this approach see the description of the ARWU Ranking see p. 24).

From 2007-2009, standard deviations and mean values for all indicators were calculated before combining the scores of each indicator into one final score. Then, using statistical methods, it was verified whether the results of the various indicators were compatible with each other and allowed the data to be combined (the THE-QS, 2009, section on normalisation[26]).

The normalisation method used involved Z-scores. First, the natural logarithm was applied to the results of each indicator[27]. Then the mean value X and standard deviation of the results of each indicator was calculated.

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

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

Z-scores are dimensionless, which means that Z-scores from different indicators can be combined together into an overall score.

The final score is calculated by multiplying each indicator score by its weighting factor, summing the results together, rounding to one decimal place and then scaling to the top performing institution, resulting in a final score out of 100 (ibid.)

Changes in the methodology over time

There have been quite a number of changes in the THE-QS Ranking since 2004.

In 2005:

- the *Employer review* indicator was introduced
- the weight of the *Peer review* indicator was reduced from 50% to 40%.

In 2007:

- in the *Peer review* indicator, it was no longer possible to select one's own university as being among the 30 best performing universities
- the application of Z-scores for the calculation of scores in each indicator, as well as the combination of indicator scores into an overall score started (see above), i.e. a university's score was no longer calculated by dividing it by the top score and multiplying it by 100
- checking the compatibility of the results of various indicators was introduced
- the data source for citations was changed from ESI (Thomson Reuters) to Scopus (Elsevier).

In 2008:

- the *Peer review* indicator was adjusted, so that peers were asked to respond separately regarding universities within and outside their own country.

In addition, the definition of exactly what data are requested has evolved gradually over the years (THE-QS, 2009).

In 2009, after the THE-QS 2009 ranking was published, THE announced the end of its cooperation with QS as the data provider and signed a contract with Thomson Reuters (Baty, 2009).

It is self-evident that so many changes in the methodology used must have had an impact on results. It seems impossible to judge how the combination of those changes might have changed the positions of individual universities appearing in the league table.

Presentation of the ranking and additional analysis produced

The Times Higher Education itself publishes the World Top 200 Universities as well as the Top 50 Universities in five areas of expertise: Engineering & IT, Life Sciences & Biomedicine, Natural Sciences, Social Sciences and Arts & Humanities.

In 2008, the THE also produced a one-off *'System strength'* table (THE 9 October 2008) listing 40 top performing higher education systems. Country scores were calculated from four data sets:

- *System* aimed to look at the capacity of systems to produce world-class universities. It was calculated as the number of each country's universities in the top 600, divided by their average position
- *Access* was calculated by taking, for each country, the number of students (FTE) at the top 500 universities, and dividing this number by the square root of the country's population
- *Flagship* is the position of the top institution in each country
- *Economic* measure. To calculate this, five points were awarded to any university in the top 100, four to universities between 101 and 200, and three, two or one respectively for each university between 201 and 300, 301 and 400, and 401 and 500. The sum of points was then divided by the per capita GDP of the country concerned.

[26] http://www.topuniversities.com.dev.quaqs.com/worlduniversityrankings/methodology/normalisation

[27] The application of a logarithmic scale squeezes the differences between scores. Thus, the logarithm is applied to avoid the effect of what are really insignificant differences between universities appearing to be big.

As regards the THE-QS Rankings results on the QS website, Top 100, 101-200, 201-300, 301-400, 400-500 and 500+ lists are available. In the interactive rankings section, users can add more fields to the ranking, compare university scores with previous years, view the results of universities in the five subject areas or sort results according to scores for different indicators. However, the option of calculating the final score using different indicator weights, as suggested by the Berlin Principles, is not provided (Berlin Principles, 2006).

1.2.2 THE-Thomson Reuters methodology for the 2010 World University Ranking

On October 30 2009, Times Higher Education announced that it had signed an agreement with Thomson Reuters to provide the data for its annual World University Rankings (Baty, 2009). This was followed by substantial changes in the set of indicators used and the overall methodology for the 2010 rankings.

Areas covered, indicators used and proxies

In June 2010, the new methodology with a new set of indicators, which intended "to make the rankings more rigorous, balanced, sophisticated and transparent", was unveiled (Baty, 2010c). The description of the methodology was first published in June 2010. It roughly described the categories and indicators, but with few details. Further changes in the methodology followed. When the THE-Thomson Reuters Ranking and its methodology were published on 16 September 2010, the indicator categories, weights and some indicators themselves had substantially changed compared to what had been previously announced (compare Baty, 2010e and Baty, 2010c).

On 5 August 2010 (see Baty, 2010d), THE also announced that the new format of the THE Ranking would exclude several categories of universities: those that had not provided data, graduate schools, and those universities that had published fewer than 50 papers in 2008.

The THE-Thomson Reuters Ranking used 13 separate indicators to compile the league tables for 2010. These 13 indicators include *Research volume, income and reputation* (total weight 30%), *Research impact* (32.5%), *Economic activity and innovation* (2.5%), *International mix – staff and students* (5%), and *Teaching – the learning environment* (30%).

So far, the description of the methodology used is less elaborate compared to that of the THE-QS rankings and, indeed, does not allow one to follow the calculation of the final scores from the raw data. The weighting of indicator categories and the list of indicators of the 2010 ranking are provided in Table 5.

Table 5. THE-Thomson Reuters 2010 Ranking. Broad categories, indicators and weightings.

Weight (broad categ.)	Broad categories	Indicators	Weight
2.5%	**Economic activity/Innovation**		
		Research income from industry (per academic staff member)	2.5%
5%	**International mix – staff and students**		
		Ratio of international to domestic staff	3%
		Ratio of international to domestic students	2%
30%	**Teaching – the learning environment**		
		Reputation survey – teaching	15%
		PhDs awarded (scaled)	6%
		Undergraduates admitted per academic	4.5%
		PhD awards/bachelor awards	2.25%
		Income per academic	2.25%
30%	**Research – volume, income and reputation**		
		Reputation survey – research	19.5%
		Research income (scaled)	5.25%
		Papers per academic and research staff	4.5%
		Public research income/total research income	0.75%
32.5 %	**Citation – research influence**		
		Citation impact (normalised average citations per paper)	32.5%

Source – Baty, 2010e

The *Economic activity/innovation* category (weight 2.5%) is represented by a single indicator in 2010:

- *Research income from industry,* scaled against academic staff[28] numbers. It also has been said that, in the future, this category may include data on the papers co-authored with industrial partners, as well as employers' perceptions of graduates (Baty, 2010c).

The originally announced *Institutional diversity* category (suggested weight 10%) (Baty, 2010c) was transformed into *International mix* with a total weight of 5% shared between two indicators:

- *Ratio of international to domestic staff* – now 3% and
- *Ratio of international to domestic students* – now 2% (compared to 5% each, as indicated in June, 2010).

It had been previously announced (Baty, 2010c) that, in the future, this category might also include a measure of research papers co-authored with international partners.

The *Institutional indicators* category with a total weight of 25%, as announced in June 2010 (Baty, 2010c), was changed into *Teaching – the learning environment* (overall weight 30%) in September. It consists of five indicators:

1. *Reputation survey – teaching* with a weight of 15%. This is based on a survey of *experienced scholars*. Combined with a reputation survey on research (see below), 13,388 responses have been gathered. The ranking compilers explain little about the reputational surveys (one on research and one on teaching) on the THE website. Other publications state that "respondents rate reputation within their own area of expertise" and that the "survey was translated into multiple languages to overcome English language bias" (Pratt, 2010). However, nowhere has it been explained what the actual survey procedures are. It would be important to know which questions were asked of the experienced scholars, how many universities they were able to choose from, whether they could nominate the universities themselves or whether they had to choose from a pre-selected range, as in previous THE rankings.

2. *PhDs awarded* (6%) – the number of PhDs awarded by an institution was scaled against the number of academic staff.

3. *Undergraduates admitted per academic* indicator (4.5%) is actually the well-known and much criticised faculty/student ratio, which is, however, assigned a substantially lower weight compared to the 15% of THE-QS rankings.

4. *PhDs/bachelor awards* (2.25%). It is not specified whether only bachelor degrees or all undergraduate degrees are meant – this would make an important difference in some countries.

5. *Income per academic* (2.25%) measured as an institution's overall income per academic without further specification whether total, academic or research staff numbers are used. It is adjusted according to purchasing-power parity.

The *Citation – research influence* (32.5%) category is a stand-alone indicator, which is the most influential of the whole set. It is calculated as the number of citations for papers referenced in 12,000 journals indexed by Thomson Reuters' Web of Science database over a five-year period (2004-2008) (Baty, 2010e). The results are normalised – citations for each paper are compared with the average number of citations received by all papers published in the same field and year (ibid.) – but it is not specified which methodology of field-normalisation is used[29]. Institutions publishing less than 50 papers per year are left out; for institutions that produce few papers because the relative citation impact may be significantly influenced by one or two highly cited papers and therefore does not accurately reflect their typical performance.

Research – volume, income and reputation category (total weight 30%) uses data from Thomson Reuters' research paper databases and includes four indicators:

- *Reputation survey* (19.5%) (see comments on *Reputation survey* of teaching above)
- *University's research income* (5.25%) scaled against staff numbers and normalised for purchasing-power parity
- *Papers per academic and research staff* (4.5%) is the number of papers published in the academic journals indexed by Thomson Reuters per academic **and** research staff member
- *Research income from public sources versus total research funding* (0.75%). Data from universities are used.

If the 2010 THE-Thomson Reuters ranking methodology is considered from another angle, the weights of the indicators used can be combined differently:

- **Bibliometric indicators** (citations per paper and papers per staff) have the greatest share of the total weight with 37%
- **Reputation indicators** (of research and teaching combined) follow closely with a weight of 34.5%
- **Income indicators** (research income: from industry, overall research income and institutional income (all per academic) and public research income versus overall research income) feature with a total weight of 10.75%
- **Importance of PhD studies** (PhD versus undergraduate

[28] In other indicators 'acacdemic research staff' is mentioned the difference has not been explained.
[29] See discussion of CPP/FCSm, MNCS and MNCS2 indicators in the section of Leiden Ranking.

student ratio and PhD per staff) has a combined weight of 8.25%

- **Internationalisation indicators** (international staff and student ratio to domestic staff and students) have a combined weight of 5%

- **Student to staff ratio** have a combined weight of 4.5%.

Thus, though more than a third of the overall weight has been assigned to the bibliometric indicators, the reputation survey(s) also continue to constitute over one third of the total weight, past criticisms notwithstanding. The reputation surveys are not described in sufficient detail to judge whether their methodology has improved since 2009.

Next in importance are the income-related indicators. However, with an overall weight of 10%, they are not assigned too great an influence.

It is not explained why the internationalisation indicators have lower weights than originally planned, especially as internationalisation is described as one of the core features of "world class universities" (Salmi, 2009).

The weight of the heavily criticised student/staff ratio is relatively low this time.

A further conclusion could be that the THE-Thomson Reuters Ranking is heavily research oriented. If the research volume and impact indicators are combined, then this already produces a weight of 62.5%. However, if the indicator on research income from industry and those on PhDs (which also characterise research) are added, then the combined weight is 73.2%.

It should also be noted that all output indicators in this ranking are relative (per staff, per publication, etc.), and that the ranking score is therefore not size-dependent.

1.3 World's Best Universities Ranking – US News & World Report in cooperation with Quacquarelli Symonds (QS)

Early in 2010, the US News and World Report (USNRW) began cooperation with QS and, on 25 February 2010, posted its new 2009 ranking on the web. This was done with a report based on the same QS results as were posted on the 2009 THE-QS World Universities Ranking website and on the QS website itself. The difference between these three is that the USNRW-QS ranking publishes a list of the Top 400 universities while the THE publishes a Top 200 list and QS publishes a Top 500+ list. The description of the methodology for the USNWR 2009 World's Best Universities Ranking is the same as that given for the 2009 THE-QS Ranking, with minor changes in the text, mainly replacing 'THE' with 'U.S. News World's Best Universities' (compare the phrases "The academic peer review is the centrepiece of the U.S. News World's Best Universities rankings [...]" in 2010 with "The Academic Peer Review is the centrepiece of the Times Higher - QS World University Rankings [...]" on the QS website a year earlier).

Thus, it has been decided that as far as the methodology used by the new US News & World report/QS Worlds' Best Universities Ranking is concerned, it is sufficient to refer to the chapter on THE-QS 2009 Ranking.

1.4 Global Universities Ranking – Reitor (Peŭmop)

The Reitor Global Universities Ranking is carried out by a ranking agency[30], Reitor (also Reiter, in the original Russian language *Peŭmop*), located in Moscow. However, the 'ideology' of the ranking originates from both Reitor and Lomonosov Moscow State University (STRF, 2008; Doneckaja, 2009). The first and so far only Reitor Global University Ranking was compiled during 2008, and the results were posted in February 2009. Although, it has been stated that there is an intention to turn it into a periodic ranking, no further information has been supplied on this matter.

The stated purpose of the ranking is to cater for the Russian academic world, which has a growing interest in the international assessment of Russian universities as a means of situating them within the global system of higher education. It also responds to the need for an instrument for assessing the competitiveness of Russian professional higher education. The Reitor Global Universities Ranking has been included in this publication for two reasons. Firstly, there have been comments that the methodologies of the most popular world university rankings do not reflect the reality of universities in the Russian Federation; and that is probably why the Reitor Ranking was produced in the first place. Secondly, based on an initial reading the Reitor indicators (at least at those that were announced before the results of the ranking were published) looked interestingly different from those of the other world rankings.

[30] The agency is referred to as independent on its website. As several sources show that the 'ideology' of ranking is strongly influenced by Lomonosov Moscow State University, we refrain from using the word 'independent'.

Which universities are considered for the ranking?

According to the stated purpose, the ranking compilers invited the following groups of universities to participate in the ranking:

1. All the world universities participating in the four main global rankings (Shanghai Jiao Tong University (ARWU-500), The Times (QS-THES-200), Webometrics and National Taiwan University[31])

2. The top higher schools of the Russian Federation

3. Institutions from the countries from the former Soviet Union

In addition, any university willing to participate in the ranking was invited to fill in and submit the questionnaire.

Areas covered, indicators and proxies

The Reitor methodology, as described on the Methodology section of the website (see Reitor, Methodology), shows that the Reitor ranking compilers planned to use a number of indicators that many of the most popular global rankings do not usually use, for example:

- number of study programmes by level (bachelor, specialist)
- number of student winners of international student academic competitions
- number of staff publications other than articles in scientific journals, namely monographs, textbooks, manuals and others
- number of certificates on discoveries and patents for inventions obtained by the university and its research officers and scholars
- total value of the training and laboratory facilities of the universities in US dollars
- proportion of teaching staff having doctoral degrees
- number of staff who have been awarded honorary doctorates from foreign universities
- number of professors who are members of the international and national Academies of Sciences as of the last academic year
- characteristics of 'socially significant' activities of the graduates, e.g. being prominent in culture, business, politics, being government officials or executives of international organisations
- number of various kinds of publications by the faculty (articles, textbooks and manuals, monographs, etc.)
- more indicators on the internationalisation of universities in addition to the foreign student and staff proportions. In particular, indicators of short-term mobility of staff and students, measuring incoming and outgoing streams of both (Reitor 2008a).

In addition to quantitative indicators, using the results of a reputation survey was also foreseen. The survey was part of the overall survey form sent to the universities participating in the ranking. The universities participating in the ranking were thus also asked for their opinion on which 10 foreign universities are "leading by education and executive training quality" (Reitor, 2008b). The indicators and survey methodology is also described in an article by the leading Reitor Ranking compilers (Kruzhalinin & Artjushina, 2008), which states that several experts at the same university can vote for the 10 best foreign universities, but that their opinions must be consolidated into a single view expressed by the university. It is underlined that having to vote for foreign universities only "enables to exclude the «patriotic» component of the evaluation as well as it is possible to enlarge the university geography" (sic) (Reitor 2008a; Kruzhalinin & Artjushina, 2008).

What methodology has actually been used? Vladimir Moskovkin (2009) draws attention to the sharp discrepancies between the sections 'Methodology' and 'About the ranking' as well as to the fact that some important information on the indicators is not presented in the 'Methodology' pages, but rather described in the 'About the ranking' page. A more in-depth analysis of the description of the methodology led to a suprising finding: a major part of the indicators listed in the 'Methodology' section on the Reitor website as well as the planned expert survey have not actually been used. Bibliometric indicators have been used instead. Nevertheless, the 'Methodology' page of the Reitor website remains unchanged. A possible reason for this change of approach could be that, although a wide range of universities had been invited to participate and complete the ranking questionnaire, only 63 universities, including just nine universities from outside the CIS, actually responded. It can be confirmed that ranking compilers sent their thanks for completing the questionnaire to a mere 63 universities (Reitor, 2009b). However, a large part of the originally chosen indicators require data input from all universities, not just from a limited number. This might explain why the ranking providers suddenly switched to different indicators based on internationally available data.

While digging further into the Reitor website, two interesting sources were found, which clarified the methodology actually used: the following description of the methodology is based on Reitor's 'About the ranking' page (Reitor, 2009a) and Reitor's presentation of the ranking results at the Lomonosov Moscow State University (Reitor, 2009c), which are also

[31] The Performance Ranking of Scientific Papers for World Universities by the Taiwan Higher Education Accreditation and Evaluation Council is probably meant here.

quoted by Vladimir Moskovkin (e.g. Moskovkin, 2009). It should first be noted that some important decisions on the calculation of the ranking results were left to the discretion of the Reitor expert pool.

According to Reitor, the expert pool sets the:

- "Rating scales for every performance indicator
- Weight of every indicator in a block [of indicators] and
- Weight of the score of each block [in the total score]" (Reitor, 2008a).

While the weights of each block of indicators in the total score can be found on the 'About the ranking' page, the weights of the individual indicators within the block cannot be obtained via publicly available sources. Hence it is not possible to follow the calculation of the aggregated scores.

The methodology actually used. Table 6 (from Reitor 2009c) is a summary of the methodology actually used. It differs substantially from the indicator list set out on the Reitor website in the section entitled 'Methodology', as described above.

Table 6. Weighting of broad indicator categories and list of indicators that have actually been used

Blocks of indicators	Weight	Indicators
I. Educational activity	20%	1. Number of educational bachelor, specialist[32], master and doctoral programmes 2. Student/staff ratio
II. Research activity	20%	3. Number of certificates on discoveries and patents since 2001 4. Performance of the computer centre of the university 5. H-index of the university
III. Financial maintenance	15%	6. Total budget of the university per full-time student
IV. Professional competence of the faculty	20%	7. Number of staff winning world-level awards (Nobel Prizes, Fields medals and others such as Descartes prize; Abel prize, the Lomonosov medal, the Russian «Global Energy» award 8. Number of staff publications 9. Citations and references to staff publications
V. International activity	10%	10. International academic communities in which the university was involved in the last academic year 11. Proportion of foreign students in the previous year
VI. Internet audience	15%	12. Volume of web-products 13. Request popularity of the university 14. Page Rank of the main page of the university's site

*Source: information in the table is taken from a **presentation of the ranking results** at Lomonosov Moscow State University (Reitor, 2009c)*

Further details about the indicators:

The *Educational activity* of a university is characterised by four straightforward indicators:

- Number of programmes that the university offers in all three cycles, which count for three indicators
- Student/staff (FTE) ratio.

Research performance is measured by:

- Number of patents and registered discoveries since 2001, using data from Scopus
- Performance of the computer centre of the university as an indicator to show the university's research capacity. Data is taken from the Top 500 of supercomputers[33]
- H-index for all authors of a university, calculated from Scopus data.

Staff professional competence is judged by:

- Number of Nobel Prizes, Fields medals and others, such as the Descartes prize[34], Abel prize[35], the Lomonosov medal of the Russian Academy of Sciences, and the Russian «Global Energy» award since 2001

[32] 'Specialist programmes' in the Russian Federation and some other CIS countries are an equivalent of the pre-Bologna 5-6 year 'Diploma' programmes in continental Europe

[33] http://www.top500.or/

[34] The Descartes Prize is an annual award in science given by the European Union since 2000, named in honour of the French mathematician and philosopher, René Descartes. The research prize is awarded to teams of researchers who have "achieved outstanding scientific or technological results through collaborative research in any field of science, including the economic, social science and humanities."

[35] The Abel Prize, which was awarded for the first time in 2003, amounts to NOK 6 million (approximately EUR 750,000 or USD 1 million). It is an international prize for outstanding scientific work in the field of mathematics, including mathematical aspects of computer science, mathematical physics, probability, numerical analysis and scientific computing, statistics.

- Number of staff publications in referenced journals since 2001

- Citations and references to staff publications since 2001.

Financial maintenance, understood as financial capacity, has a weight of 15%. The indicator value is actually the total budget of the university in the preceding year per full-time student.

The *International activity* of a university has a weight of 10% and is evaluated using two indicators that are explained as follows:

- "Universities' activity in the international academic communities. Due to the [high] number of such communities of the world evaluation is carried out as expert appraisal of this activity" (Reitor, About the ranking)

- Total number of foreign students divided by the total number of students in the university.

Internet audience. Three indicators are used:

- The volume of web-products is included because "it characterises the development of Internet technologies in the universities in preparation and advancement of scientific and methodical publications" (Reitor, About the ranking[36])

- The 'request popularity' of the university is measured as the number of queries received by the university website in 2008

- Google Page Rank of the main homepage of the respective university's website, for autumn of 2008 was used.

Data sources

(according to Reitor 2008b)

Given that few universities – around 10% of those to whom the questionnaire was distributed – returned a completed survey, it is likely that the rest of the data have been taken from university websites or offices gathering statistical data on education. Searches on university websites is also mentioned as an additional data source by the ranking compilers themselves (Reitor, 2009a):

"The compilers of the rating have carried out the scaled work on search and editing of public data, using following informational sources:

- Official sites of universities

- Annual reports of universities

- National agencies specializing in gathering and processing educational statistics

- Various ratings which may also include estimated universities (for example, Top-500 rating of the world supercomputers)

- Records of the scientometric data base Scopus

- Google search system data".

Calculating indicator values and transforming them into scores

The description on the Reitor 'Methodology' page says that each expert evaluates each indicator for all universities. The total score for each particular indicator is calculated as the average of all expert evaluations. Should there be a spread of more than 15% across the expert evaluations, then an expert discussion takes place to determine the final value based on consensus (Reitor, 'Methodology'). It is difficult to judge whether this procedure has been followed or not.

Changes in the methodology over time

The ranking was first published in 2009 and there is no follow-up as yet. It is important to remember, however, that the methodology descriptions on the Reitor website are contradictory. Three different versions exist – a first given in the 'Methodology' section, a second contained in the 'About the ranking' page and a third provided in the presentation given at Lomonosov Moscow State University.

Presentation of the ranking and additional analysis produced

Besides the main league table, there is one ranking that includes countries by number of universities in the Top 500, and one showing the Russian Federation and Commonwealth of Independent States (CIS) universities.

[36] http://www.globaluniversitiesranking.org/index.php?option=com_content&view=article&id=96&Itemid=54

2. Rankings concentrating on research performance only (with or without producing league tables)

2.1 Leiden Ranking – Leiden University

The Leiden Ranking provider is the Centre for Science and Technology Studies (CWTS) at Leiden University, which has developed its own bibliographic indicators. The results of the Leiden Ranking were first published in 2008. On 2 August 2010, updated results were posted on the internet.

The stated purpose of the Leiden Ranking is as follows: "The Leiden ranking aims at comparison of research institutions with impact measures that take the differences in disciplines into account".

The Leiden Ranking does not present a composite overall score, but rather scores according to various indicators (four in 2008; five in 2010). Each of these indicators is called a 'ranking' and presented separately. However, while universities are ranked separately according to each indicator, their results in the other indicators are also shown.

Which universities are considered for the ranking?

The Leiden Ranking focuses on the universities with the largest number of Web of Science indexed publications per year. In 2008, roughly 1000 universities with the largest number of publications worldwide were covered, and tables of the Top 250 tables were published. In 2010, the tables presented the Top 100, Top 250 and Top 500 universities.

Areas covered and indicators used

The Leiden Ranking covers research only, with a special focus on scientific output. The five indicators used in the 2010 ranking are the following:

1. Number of publications (P) ('yellow' indicator according to CWTS). This indicator refers to the number of publications in journals covered by citation indexes (Web of Science, Scopus) over a certain period of time. The 2010 Leiden Ranking covers publications for the period 2003-2008 and citations for 2004-2009.

This indicator relates to the research output of the whole university and is therefore size-dependent.

2. Number of citations per publication (CPP) (the 'blue' indicator according to CWTE). The value of the CPP indicator is calculated leaving out self-citations. The results of the subtraction of the number of self-citations $\mathbf{C}_s$ from the total number of citations for each article $\mathbf{C}_i$ are summed up, and the sum is then divided by the total number of articles **P**:

$$CPP = \frac{1}{P}\sum_{i=1}^{P}(C_i - C_s)$$

Where:
CPP is the number of citations per paper
P is the number of publications
C_i is the total number of citations for the article
C_s is the number of self-citations for the article

The CPP makes it possible to judge the average scientific impact of the university. The disadvantage of the CPP as an indicator is that it does not take into account that older articles have usually accumulated more citations, and that citation rates vary between document types and subject areas (Rehn et. al, 2007).

In addition to this, the CPP indicator is a relative number and therefore does not demonstrate the overall strength of the university by itself.

3. Field-normalised citations per publication (CPP/FCSm) ('lighter green' indicator according to CWTE). The CPP/FCSm indicator was introduced in 1993, although the normalisation mechanism dates back to the 1980s (Waltman et. al, 2010). The Leiden Ranking providers themselves call this indicator 'the crown indicator' in the 2008 ranking, due to the fact that it is size-independent and field-normalised.

The value of this indicator is calculated by dividing the result of the previous ('blue') indicator CPP (see above) by the mean fields citation score (FCSm).

$$\frac{CPP}{FCSm} = \frac{\sum_{i=1}^{P}\frac{C_i}{P}}{\sum_{i=1}^{P}\frac{e_i}{P}} = \frac{\sum_{i=1}^{P}C_i}{\sum_{i=1}^{P}e_i} = \frac{C_1 + C_2 + C_3 \ldots}{e_1 + e_2 + e_3 \ldots}$$

Where:
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 in which publication *i* was published
P is the number of publications

The normalisation of the citation values is based on the sums of the citations and the field citation scores areas. However, using this approach, has some disadvantages (see Leydesdorf

& Opthof, 2010). It has already been noted that citation rates here are not normalised at the level of individual publications but at a higher aggregation level. Also, this method of calculation gives more weight to older publications, particularly reviews, and to those published in fields where citation levels are traditionally higher (Rehn et. al, 2007).

4. Mean-normalised citation score (MNCS2) – ('darker green' indicator according to CWTS). In 2010, CWTE developed a new indicator to replace the previous 'crown indicator'. The mean-normalised citation score (MNCS) had already been introduced in 1990 and is calculated as:

$$MNCS = \frac{1}{P}\sum_{i=1}^{P}\frac{C_i}{e_i} = \frac{1}{P} * \left(\frac{C_1}{e_1} + \frac{C_2}{e_2} + \frac{C_3}{e_3} + \cdots\right)$$

When comparing the calculation of CPP/FCSm (see the previous indicator described above) and MNCS, it emerges that, while the CPP/FCSm indicator normalises by calculating a ratio of averages, the MNCS indicator normalises by calculating an average of ratios (Waltman et. al., 2010). The difference between the two is that in CPP/FCSm, the citations (in the numerator) are first summed up, as are the expected numbers of citations of publication e_i in the denominator. After that, the first sum is divided by the second. Because more recent publications have fewer citations, they have little influence in the sums both in the numerator and the denominator. The impact of new publications is therefore very small.

In the case of the MNCS indicator, the number of citations of a publication is individually divided by the expected number of future citations. For new publications the number of citations is thus small, but so is the expected number of citations, which is why the impact of recent publications is not discriminated against when compared to older ones. However, the MNCS calculation entails a new disadvantage (Leydesdorf & Opthof, 2010), which is that for very recent publications, the expected further number of citations is difficult to predict, therefore making the results less stable.

Finally, the MNCS (or MNCS1) indicator was modified into the MNCS2 indicator, which differs from MNCS1 in that it leaves out recent publications, which are defined as those that have had less than one year to earn citations. The empirical analysis of Waltman et al. (Waltman et al., 2010) demonstrates that the MNCS2 indicator can indeed replace the former 'crown indicator' CPP/FCSm.

The CPP/FCSm and both the MNCS1 and MNCS2 indicators demonstrate the 'efficiency' of publications. For example, if for university X the value of one of those indicators is 1.57, it means that the publications of university X generate on average 1.57 times more citations than is typical for publications in that particular field for a given year. Or, if the value of one of those indicators for university Y is 0.4, it means that the 'efficiency' of its publications is lower and that therefore, in order to reach the same number of citations, university Y has to publish 2.5 times more papers than others.

5. The 'brute force' indicator P*CPP/FCSm – ('orange' indicator). To demonstrate the actual power of a research group or university in the world, the field-normalised citation number is multiplied by the total number of publications.

$$P * \frac{CPP}{FCSm} = P * \frac{\sum_{i=1}^{P} C_i}{\sum_{i=1}^{P} e_i}$$

Thus, the Leiden "brute force" indicator in a way represents the total number of publications, corrected so as to take into account their "efficiency".

Data sources

Only bibliometric data from the citation databases are used.

Calculating indicator values and transforming them into scores

The results of individual indicators are not combined into one 'final value', and no overall league table is produced.

Changes in the methodology over time

Compared with 2008, a new indicator 'mean normalised citation score' was added in 2010.

Presentation of the ranking and additional analysis produced

Top 100 and Top 250 (the latter added in 2010) tables are prepared both worldwide and for Europe; a worldwide Top 500 table was also prepared in 2010.

The Leiden Ranking team has been involved in the development of indicators and normalisation methods and has also computed the scores of some indicators for other rankings, including EU-supported ranking projects, such as the CHE Excellence Ranking and U-Multirank.

The Leiden Ranking team recently compared the citation impact as well as the ranking scores, based on the citation impact of English-language publications of the 69 most highly cited German and French universities, to the overall sum of their publications (i.e. also those in German and French, respectively). The results showed that the citation impact of publications in English is systematically higher than the impact of all their publications combined[37]. Correspondingly, the positions in the ranking of these universities, based on their English-only citation impact, are systematically better than the results based on the entirety of their publications (van Raan et al., 2010).

It should be added that this effect is most probably even more pronounced for languages other than French and German, for example Russian, Spanish, Chinese or Japanese.

[37] i.e. publications of French and German universities published in all languages.

2.2 Performance Rankings of Scientific Papers for World Universities – Taiwan Higher Education Accreditation and Evaluation Council Ranking

The Taiwan Higher Education Accreditation and Evaluation Council Ranking (hereafter referred to as HEEACT Ranking) is an annual world university ranking that has been produced since 2007. The HEEACT Ranking evaluates and ranks performance in terms of the publication of scientific papers for the top 500 universities worldwide, using data drawn from SCI and SSCI.

The ranking providers underline that they place emphasis on research performance, and that this is what distinguishes the HEEACT Ranking from the THE Ranking "that focuses on university ranking, and ARWU focusing on academic ranking" (Huang, 2009).

The ranking has eight indicators in three main categories: *Research productivity* (20% of the overall weight), *Research impact* (30%) and *Research excellence* (50%).

The HEEACT Ranking also takes account of university mergers and specialised university institutes or different campuses in a university system and therefore also includes publications by a given university's affiliated bodies, such as research centres and university hospitals.

Which universities are considered for the ranking?

The selection of universities is based on the number of journal articles and citations. To produce a Top 500 list, 700 institutions are first selected out of the 4000 research institutions listed in Essential Science Indicators (ESI). Institutions that are not universities are then removed, and the remaining institutions are compared with the THE, ARWU and US News and World Report ranking lists. This results in 725 universities.

Areas covered and indicators used

Four out of the eight indicators are calculated per staff (FTE), by which the ranking providers' attempt to mitigate size bias. The focus is on impact rather than performance. The impact indicators together constitute 80% of the total score of a university. Attempts are made to reflect recent successes of universities, as 55% of the total score stems from indicators using data from the past 1-2 years (Huang, 2009).

The **Research productivity** category has two indicators:

- *Number of articles* published in peer-reviewed academic journals in the past 11 years [per staff FTE][38]
- Number of articles published in the previous year [per staff FTE].

The overall weight of the productivity category is relatively low at 20%.

The **Research impact** section has three indicators and its overall weight is 30%:

- *Number of citations in the last 11 years* is the total number of citations of the articles of the university in question over the last 11 years, divided by the number of staff FTE
- *Number of citations in the last 2 years* is the total number of citations drawn from SCI and SSCI per staff FTE
- *Average number of citations in the last 11 years* is the total number of citations of a university over the last 11 years, divided by the total number of publications of the university over the last 11 years.

The **Research excellence** section has three indicators, which constitute 50% of the final score:

- *H-index of the last two years,* in which the value *h* is the number of articles published by a university in the last two years, which are cited no less than *h* times.

This indicator constitutes 20% of the total score.

- *Number of Highly Cited Papers* is the absolute number of papers of the university in question that belong to the 1% most cited papers in ESI published in the last eleven years.
- *Number of articles in high impact journals in the last year* is the absolute number of publications of the university in question published over the last year in one of the top 5% journals by impact factor.

[38] The indicator table on the HEEACT Ranking website is somewhat misleading, as the names of indicators do not show which indicators are per FTE and which are not.

Table 7. Areas and indicators of the Taiwan Higher Education Accreditation and Evaluation Council Ranking

Criteria, weight	Overall performance indicators	Weight
Research productivity (20%)	Number of articles in the last 11 years[39] [per staff FTE]	10%
	Number of articles in the previous year[40] [per staff FTE]	10%
Research impact (30%)	Number of citations in the last 11 years [per staff FTE]	10%
	Number of citations in the last two years[41] [per staff FTE]	10%
	Average number of citations [per publication] of the last 11 years	10%
Research excellence (50%)	H-index of the last two years	20%
	Number of highly cited papers in the last 11 years	15%
	Number of articles in high impact journals in the last year	15%

Data used to assess the performance of universities is drawn from ISI's ESI and Web of Science (WoS), which includes the Science Citation Index (SCI), the Social Sciences Citation Index (SSCI) and the Journal Citation Reports (JCR).

Although the Arts & Humanities Citation Index (A&HCI) is also available, the HEEACT Ranking does not include the database because "it may fail to objectively and accurately represent the research performance of arts and humanities researchers". The database mainly indexes English-language journals, while arts and humanities researchers often publish in their native languages and in various forms of publications (e.g. books). According to the ranking providers, focusing on data obtained from SCI and SSCI allows for fairer comparisons across universities globally.

The *Numbers of university faculty staff (FTE)* are obtained from QS, from university websites themselves or through the respective country's higher education administration.

The *Number of citations in the last 11 years and the Number of Highly Cited Papers* are taken from ESI, while the Number of citations in the last 2 years as well as the *H-index of the last 2 years* originate from SCI and SSCI at WoS, which have recent statistics. The *Number of articles in high-impact journals in the last year* uses data from JCR.

Calculation of the final scores

Indicator scores. As in several other rankings, the score of each indicator is calculated by dividing the indicator value of the university in question by that of the university holding the highest value. The result is multiplied by 100[42]. When calculated in this way, the scores of individual indicators are dimensionless relative values and can therefore be combined.

Overall score. The overall score is calculated by multiplying the score of each indicator by the weight of the indicator and summing them up.

Changes in the methodology over time

In the first HEEACT Ranking of 2007, there were altogether nine indicators, of which the *H-index* indicator was assigned 20%, while each of the eight other indicators weighed 10% of the final score. As of 2008, the indicator *Number of subject fields where the university demonstrates excellence* was removed, and its weight has been shared equally between the indicators *Number of highly cited papers* and *Number of articles in high impact journals*.

The score of the (now removed) indicator *Number of subject fields where the university demonstrates excellence* is the number of subject fields (out of the twenty-one ESI fields) in which the university in question is listed in ESI's Citation Rankings (only the top 1% of the most cited institutions are included in those lists). This indicator favoured smaller/specialised universities demonstrating excellence in specific subject fields and made them more visible. Its removal has, to a certain extent, weakened the position of such institutions. In fact, small/specialised institutions are generally overlooked by other indicators of the excellence section of the ranking, which sum up excellence in various subject fields.

[39] I.e. 1999-2009 for the 2010 ranking and accordingly for the rankings of previous years.
[40] I.e. 2009 for the 2010 ranking and accordingly for the rankings of previous years.
[41] I.e. 2008-2009 for the 2010 ranking and accordingly for the rankings of previous years.
[42] For clarification, refer to the example given in the chapter 'Calculating indicator values and transforming them into scores' on the ARWU ranking.

Presentation of the ranking and additional analysis produced

The HEEACT Ranking results are presented as a Top 500 list of universities listed by 1-100, 101-200, 201-300, 301-400 and 401-500 tables. Users can choose to see lists ranked by the total score or according to the score for each indicator. Ranking lists by continent and by country are also available.

HEEACT also publishes the results of rankings in the following six fields:

1. Agriculture and Environment Sciences
2. Engineering, Computing and Technology
3. Clinical Medicine
4. Life Sciences
5. Natural Sciences
6. Social Sciences.

The results of field rankings are presented as the Top 100 field lists, lists by continent and by country. Field rankings have the same indicators as the overall university rankings. However, compared to the overall university rankings, the Top 500 field rankings contain a different sample of universities. For each subject field, 400 universities are selected from the ESI database by both number of articles and number of citations in the past 11 years. Only institutions that provide undergraduate and graduate degrees in each field are included. The initial lists are supplemented through comparisons with the ARWU and THE ranking lists, and they finally include between 460 and just over 500 universities, depending on the specific field.

2.3 Assessment of University-Based Research – European Commission

In 2008, the Directorate General for Research of the European Commission established a Working Group on the Assessment of University-Based Research (hereinafter WG AUBR). The final report of the group (WG AUBR, 2010) was published at the beginning of 2010.

The terms of reference of the Working Group included the following objectives:

1. "Review the needs of various types of users of measurement of research quality at universities;

2. Review main methodologies for assessing/ranking research quality of universities, covering existing international assessments/ rankings and other methodologies being developed;

3. Propose as far as possible a consolidated multidimensional methodological approach addressing various users' needs, identifying data and indicators requirements (if necessary propose different approaches for different types of users)" (ibid.).

Thus, the task of the WG AUBR was not to create a ranking, but rather to analyse the research indicators used by existing university rankings and to suggest a methodology for a multidimensional assessment of university-based research.

Analysis of indicators used for research assessment

Among other things, the working group prepared a list of indicators that can be used for the assessment of university research and sought to identify the strengths and weaknesses of these various indicators (WG AUBR, 2010, pp. 44-48 and 69-82). Moreover, WG AUBR identified measures that should be taken prior to using each indicator. For instance, before applying the indicator *Research outputs per 'Research academic'*, agreement must be reached on the definition of a 'research academic'. Similarly, before applying the indicator *Research income per academic staff or FTE*, data needs to be adjusted to the scale and mission of the university in question (ibid., p. 45).

Some of the WG AUBR's considerations concerning the research indicators used, either directly or indirectly, in global university rankings are reproduced below:

- *Count of publications and other research outputs.* Different disciplines produce different types of research outputs. Also, this indicator puts emphasis on quantity of publications.

- *Number/percentage of publications in top-ranked, high impact journals.* Especially in social sciences and humanities, expert rankings do not correlate very well with impact factors. In these fields, and in engineering, other sources are important as well (books, proceedings).

- *Citations.* Citations are of limited value in disciplines not well covered by the citation indexes, especially certain parts of social sciences, humanities and engineering.

- *Number of prestigious national/international awards and prizes.* There are no agreed equivalences that apply internationally.

- *Visiting Research Appointments.* There are no agreed equivalences that apply internationally and facilitate comparison across disciplines.

- *Editorial and refereeing for prestigious national/international journals/publishers.* There are no agreed equivalences that apply internationally.

- *Commercialisation of intellectual property created through patents, licences or start ups.* Patents are a very poor indicator of commercialisation. They are sensitive to national context and to discipline.

- *Number of collaborations and partnerships.* It can be difficult to capture and verify the data due to lack of clarity.

- *Percentage of 'research active' staff per total academic staff.* There is no clear definition of 'research active'.

- *Level of funding attracted by researchers and universities from external sources.* Levels of external funding vary greatly across disciplines.

- *Research income per academic staff or FTE.* Comparability is dependent upon institutional mission, context and discipline.

- *Total R&D investment.* It is difficult to get valid, comparable institutional data.

- *Research infrastructure and facilities.* It is difficult to get valid, comparable data, favours older, well-endowed universities. (ibid., p. 43-47)

Multidimensional Research Assessment Matrix

An important result presented in the final report of the WG AUBR is the Multidimensional Research Assessment Matrix included as table 8 (reproduced from the WG AUBR, 2010).

The matrix allows for the identification of appropriate indicators from among five categories – *Research productivity, Quality and scholarly impact, Innovation and social benefit, Sustainability and scale* and *Research infrastructure* – depending on the purpose of the assessment: to allocate resources, to drive research mission differentiation, to increase regional/community engagement, to improve research performance, to assess value-for-money or cost-benefit of research, to encourage international co-operation, or to increase multidisciplinary research.

Table 8. Multidimensional Research Assessment Matrix

PURPOSE	RESEARCH PRODUCTIVITY	QUALITY AND SCHOLARLY IMPACT	INNOVATION AND SOCIAL BENEFIT	SUSTAINABILITY AND SCALE	RESEARCH INFRASTRUCTURE
Allocate Resources	• Research output/ bibliometric data	• Citation data • Peer review • Keynote, awards, etc.	• Research income		• 'Research active' as percentage of total academic staff • Libraries, equipment, etc.
Drive Research Mission Differentiation	• Research output/ bibliometric data • Output per research academic	• Peer review • Self-evaluation	• Ratio of research income vs. teaching income • External research income	• Ratio of undergraduate vs. master/PhD students	
Increase Regional/ Community Engagement	• Publications, policy reports, etc.	• End user reviews • Keynote, media awards, etc.	• Percentage of funding from end users • Patents, licences, spin-offs	• Number of collaborations and partnerships	
Improve Research Performance	• Research output/ bibliometric data	• Citation data • Number and percentage of publications in top ranked, high impact journals • Peer review			
Assess Value-For-Money or Cost-Benefit of Research	• Research output/ bibliometric data • Output per research academic	• Peer review and/or citation data • Commercialisation data • End user reviews	• Social, economic, cultural and environmental impact/benefit indicators • External research income • Employability of PhD graduates	• Number of collaborations and partnerships	
Encourage International Co-operation	• Research output/ bibliometric data with focus on European & international collaborations		• Percentage of research income from international sources	• Number of collaborations and partnerships	
Increase Multidisciplinary Research	• Research output/ bibliometric data with focus on interdisciplinary fields	• Peer review • Self-evaluation	• New research fields, interdisciplinary teaching programmes, etc.		• Research conducted by people from different disciplines

Source: WG AUBR, 2010

3. Multirankings

Multirankings are university rankings that use a greater number of indicators and usually do not produce a league table, but present instead results of individual indicators or groups of indicators.

3.1 CHE Ranking – Centre for Higher Education Development/ *die Zeit*

3.1.1 CHE University Ranking

The German Centre for Higher Education Development[43] (CHE) carries out a multi-indicator ranking which was first published in 1998. The purpose of the CHE University Ranking is to provide fair, informative and qualified information for young people who are choosing an HEI for their studies, for students already enrolled as well as for the HEIs themselves.

The results of the CHE University Ranking can be visualised in various ways (see below). It can be considered both as a rating and a ranking. In some of the visualisation options the universities are listed in alphabetical order. The results for several indicators are presented in a table that shows only whether the university belongs to the top, middle or bottom group for a particular indicator. In other visualisations, the universities are ranked by one indicator only. Once again the results show only to which of the three broad groups the university belongs. However, simultaneously the results for the other indicators are also displayed.

An 'overall score' is used only in *'My Ranking'* (see details page 48), where users may select up to 10 indicators that they consider important in order to choose an appropriate institution. The system then determines the universities that best fit the user's interests.

CHE's university rankings are highly interactive. By using the websites of *Die Zeit*[44] or DAAD[45], an individual user can choose between various indicators (or groups of indicators), thus creating a personalised ranking that retrieves the universities that best fit his or her chosen requirements. The CHE rankings are primarily designed to help students choose the university most suited to their needs. As indicated above, they can be accessed through the *Die Zeit* and DAAD websites.

Which universities are considered for the ranking?

The CHE University Ranking was initially designed for German higher education institutions, including universities, *Hochschulen* or HEIs more generally and *Fachhochschulen* or universities of applied sciences. The internationalisation of the CHE University Ranking started with pilot projects involving Austrian universities in 2004 and Swiss universities in 2005, although neither of these pilots was developed further. As of 2007, the CHE Rankings included HEIs from the Netherlands and the Flemish community of Belgium[46]. This resulted in the inclusion of practically all Dutch universities and other HEIs by 2009.

The internationalisation of the CHE Ranking continues through its extension to universities in countries in which German is used, either as the sole language of tuition or in combination with the national language, such as in Italy and Hungary.

Areas covered, indicators and proxies

One of the main aims of the CHE Ranking is to provide potential students with information underpinning their choice of HEI, taking account of their interest in a particular field of study. For this reason, the CHE Ranking covers all fields of study and is organised by field.

The description of the CHE University Ranking lists more than 100 indicators organised in nine indicator blocks. However, the actual number of indicators used in the various versions of the ranking described below usually does not exceed 30. If studies in the chosen field are offered in both universities and universities of applied sciences *(Fachhochschulen)*, then the user first has to choose between the two. The set of indicators from which the user can choose is different for these two types of institution.

CHE processes data and has accumulated results for a huge range of indicators. However, the indicators actually used in the different versions of the CHE University Ranking are mainly based upon students' assessment of a number of particularly relevant aspects, such as teaching support, the overall evaluation of teaching quality, the quality of libraries and IT, lecture and seminar rooms and university sports. It also covers aspects such as the university's research orientation and preparation for the labour market. Assessments by academics are used less often; they mainly concern the reputation of research and/or teaching in other universities. Some of the

[43] CHE was established in 1994 by the Bertelsmann Foundation and the German Rectors' Conference as a non-for-profit limited company.
[44] www.Zeit.de
[45] www.daad.de
[46] This was part of an EU-supported project.

indicators are based upon 'facts', i.e. statistical or bibliometric data.

A comparison of the indicators used in the CHE Rankings for 20 randomly chosen study fields demonstrated the following: currently, student opinion still makes up the lion's share, reaching 70% of the overall number of indicators used in some fields, but decreasing to just over 40% in those subject fields for which the methodology has recently been updated. The share of indicators based on 'facts' made up between 10% and 30%, while the combined weight of those based on staff assessment fell to between 5% and 10%. Where the opinions of graduates were included, these constituted up to 20% of the total indicators.

For all assessment-based indicators, respondents assign scores on a scale from 1 (very good) to 6 (very bad).

Changes in the methodology over time

The methodology is gradually being updated field by field. As a result, for some fields of study, the indicator sets are more up-to-date than for others.

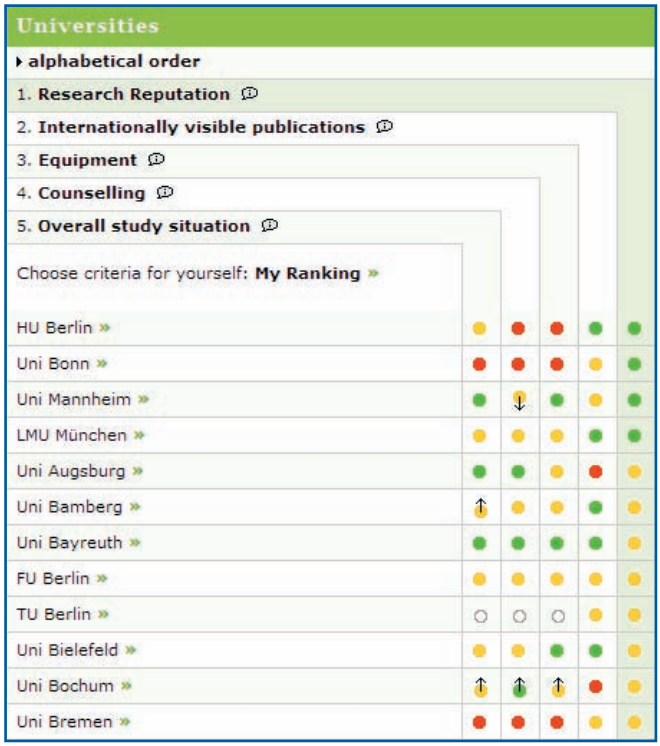
In the subject fields for which the methodology has recently been updated, the opinions of graduates are being introduced and will replace the opinions of students. There is also a tendency to increase the importance of statistical and bibliometric data compared to indicators based on reputation.

Presentation of the ranking and additional analysis produced

The various versions of the CHE University Rankings are: the *'Compact Ranking'* (also called *'Ranking Overview'*), the *'Quick Ranking'* (*Die Zeit* website), *'My Ranking'* and *'University Comparison'* (DAAD website).

The ***CHE 'Compact Ranking'*** (called CHE *'Ranking Overview'* on the DAAD website) allows for the comparison of study programmes in a particular field of study at different universities or *Fachhochschulen*. When the user has chosen between universities and *Fachhochschulen*, a list of institutions offering programmes in the chosen field appears. Initially, the list is in alphabetical order. It shows to which group – top, middle or bottom – the HEI belongs for each of the five indicators used (see example in Fig. 1). It is possible to rearrange the table with the final results according to any of the five indicators. However, the results never become a league table, since even within the top, middle or bottom groups, universities are listed alphabetically. The five indicators can differ depending on the field of studies: sciences, engineering, medicine, social sciences and humanities will have a somewhat different combination of indicators, based upon a selection by the ranking providers of the most appropriate indicators for comparing HEIs offering the particular type of programmes in question.

Figure 1. Example of the CHE *'Compact Ranking'* screen on the *Die Zeit* website



The principles underpinning the selection of the five indicators used for the *'Compact Ranking'* seem to be the following:

1. An indicator based on student assessment of the overall study situation is used for all programmes.

2. An indicator based on professors' opinions is also applied to all programmes. In most cases, this is the *Research reputation* indicator, but in the case of programmes offered by *Fachhochschulen,* it can also be the Professors' opinions on the quality of academic studies and teaching.

3. The *Teaching support* indicator, which is again based on students' assessment, is often used as a second indicator on studies and teaching.

4. A second research indicator is often used, regarding either *Citations per publication* (in sciences, engineering, medicine) or *Third-party funding of research per faculty member* (in other fields).

5. At least one indicator on study-related infrastructure is used. In most cases, this is the students' assessment of either libraries or laboratories. For some types of programmes, this indicator can be very specific, such as the number of hospital beds per 100 students in medicine, excursions in geosciences or practical placements in engineering.

Table 9. Examples of indicators typically selected for the five-indicator combinations of the CHE *'Compact Ranking'*.

Indicator	Meaning of indicator	Applied in
Overall study situation	Overall teaching/study situation assessed by students on a scale of 1 (very good) to 6 (very bad)	All fields
Teacher support	Student assessment of accessibility, consulting hours, counselling, discussion of homework etc. on a scale of 1 (very good) to 6 (very bad)	Often used as a second indicator on studies
Reputation for academic studies and teaching	Higher education institutions which professors recommend with regard to the quality of teaching.	Used for *Fachhochschulen*
Research reputation	Research reputation is drawn from opinions of professors on which tertiary institutions (except their own) are, in their opinion, the leading ones in research.	Used for universities
Citations per publication	Average number of citations per publication	Used as a second indicator on research e.g. in sciences, medicine, pharmacy
Research funding	Third-party funding from industry, foundations, etc. per faculty member (except third-party funded places)	Used as a second indicator on research in fields other than those above
Laboratories	Availability and state of the laboratory workplaces, etc. on a scale of 1 (very good) to 6 (very bad).	Used e.g. in sciences, engineering
Library	Availability of literature, user support, possibility of literature research and the availability of workstations, etc. on a scale of 1 (very good) to 6 (very bad)	Used alternatively to indicator on laboratories

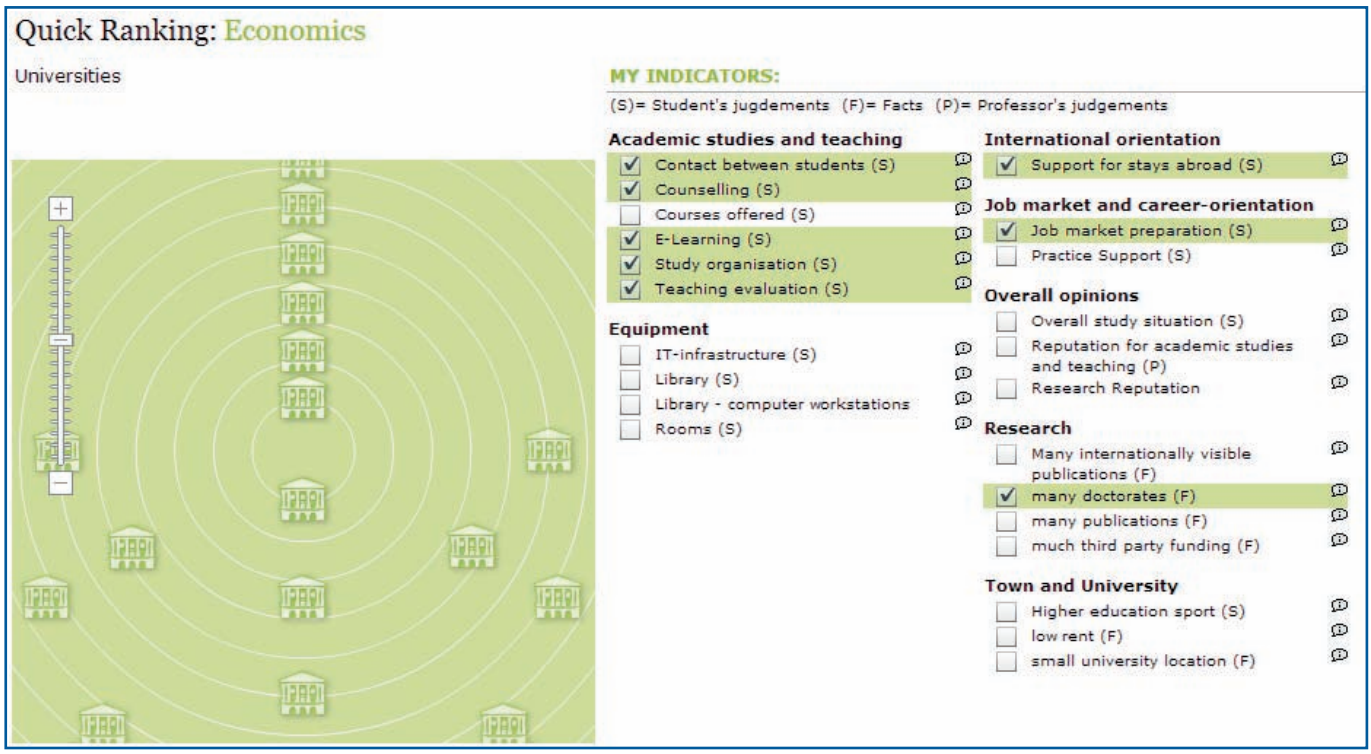
The **CHE *'Quick Ranking'* (DAAD website)** allows students to retrieve information quickly to find which universities best fit their particular wishes. Students first select the subject area and then choose between universities and *Fachhochschulen* (FH) or, where appropriate, between studies leading to a bachelor degree and a *Diplomstudium*. Then a list of indicators grouped under six or seven headings appears. There are usually several indicators under each heading. Users can choose the indicators they consider important (see Table 10).

Table 10. The CHE *'Quick Ranking'* – indicators offered for user's choice

Headings	Indicators	Comments
Academic studies and teaching		
	Contact between students and staff	Offered in all cases
	Counselling	Offered in all cases
	Courses offered	Offered in all cases
	E-Learning	Offered in all cases
	Study organisation	Offered in all cases
	Teaching evaluation	Offered in all cases
	Research orientation	Not FH, not all university studies
	More field-specific indicators possible	
Equipment		
	IT-infrastructure	
	Library	
	Rooms	
	Facilities like laboratories, sports facilities, hospital beds etc.	Where appropriate
International orientation		In most but not all programmes
	International orientation or support for stays abroad	Either one indicator or the other in most programmes
Job market and career orientations		
	Job market preparation	Offered in all cases
	Support in practical semester	Offered in all cases
	Several other indicators possible	In some programmes
Overall opinions		
	Overall study situation	Offered in all cases
	Research reputation	For university studies
	Reputation of academic studies and teaching	For FH studies
Research		
	Much third-party funding	Offered in all cases
	Many doctorates	Not for FH studies
	Many citations	Some university programmes
	Many publications	Some university programmes
Town and university		
	Higher education sports	Offered in all cases
	Low rent	Offered in all cases
	Small university location	Offered in all cases

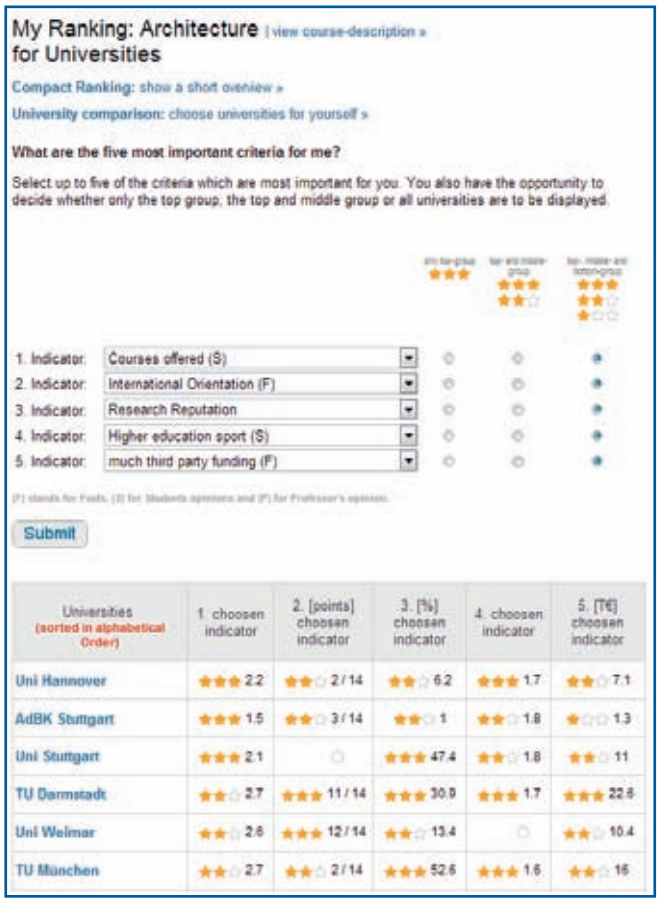
Upon selecting or deselecting indicators, the user is presented with a visualisation that demonstrates which HEI(s) are positioned closer to the centre of the 'target'.

Figure 2. Example of the CHE *'Quick Ranking'* screen on the *Die Zeit* website



The **CHE** ***'My Ranking'*** is a modification on the *'Quick Ranking'* available on both the DAAD and *Die Zeit* websites. The difference between *'Quick Ranking'* and *'My Ranking'* is that for the latter, users can only choose five indicators out of the list. As with *'Quick Ranking'*, the list of indicators offered varies depending on the subject and on whether universities or *Fachhochschulen* are chosen.

Figure 3. Example of the CHE *'My Ranking'* screen on the DAAD website



My Ranking: Architecture | view course-description »
for Universities

Compact Ranking: show a short overview »
University comparison: choose universities for yourself »

What are the five most important criteria for me?

Select up to five of the criteria which are most important for you. You also have the opportunity to decide whether only the top group, the top and middle group or all universities are to be displayed.

		top group	top and middle group	top, middle and bottom group
1. Indicator:	Courses offered (S)	○	○	●
2. Indicator:	International Orientation (F)	○	○	●
3. Indicator:	Research Reputation	○	○	●
4. Indicator:	Higher education sport (S)	○	○	●
5. Indicator:	much third party funding (F)	○	○	●

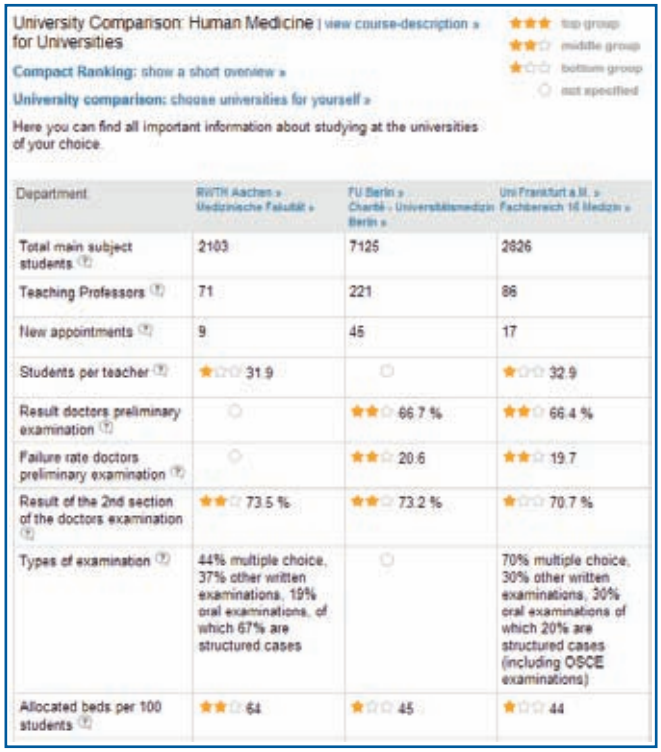
(F) stands for Facts, (S) for Students opinions and (P) for Professor's opinion.

Submit

Universities (sorted in alphabetical Order)	1. choosen indicator	2. [points] choosen indicator	3. [%] choosen indicator	4. choosen indicator	5. [T€] choosen indicator
Uni Hannover	★★★ 2.2	★★☆ 2 / 14	★★☆ 6.2	★★★ 1.7	★★☆ 7.1
AdBK Stuttgart	★★★ 1.5	★★☆ 3 / 14	★★☆ 1	★★☆ 1.8	★☆☆ 1.3
Uni Stuttgart	★★★ 2.1	○	★★★ 47.4	★★☆ 1.8	★★☆ 11
TU Darmstadt	★★☆ 2.7	★★★ 11 / 14	★★★ 30.9	★★★ 1.7	★★★ 22.6
Uni Weimar	★★☆ 2.6	★★★ 12 / 14	★★☆ 13.4	○	★★☆ 10.4
TU München	★★☆ 2.7	★★☆ 2 / 14	★★★ 52.6	★★★ 1.6	★★☆ 16

The **CHE '*Comparison of universities*'** enables the comparison of two to three universities with regard to a selected subject. It uses a similar set of indicators to the ones used in the *'Quick Ranking'* but without the town and university section (see Fig. 4).

Figure 4. Example of the 'University Comparison' screen on the DAAD website



University Comparison: Human Medicine | view course-description »
for Universities

★★★ top group
★★☆ middle group
★☆☆ bottom group
○ not specified

Compact Ranking: show a short overview »
University comparison: choose universities for yourself »

Here you can find all important information about studying at the universities of your choice.

Department	RWTH Aachen » Medizinische Fakultät »	TU Berlin » Charité - Universitätsmedizin Berlin »	Uni Frankfurt a.M. » Fachbereich 16 Medizin »
Total main subject students	2103	7125	2626
Teaching Professors	71	221	86
New appointments	9	45	17
Students per teacher	★☆☆ 31.9	○	★☆☆ 32.9
Result doctors preliminary examination	○	★★☆ 66.7 %	★★☆ 66.4 %
Failure rate doctors preliminary examination	○	★★☆ 20.6	★★☆ 19.7
Result of the 2nd section of the doctors examination	★★☆ 73.5 %	★★☆ 73.2 %	★☆☆ 70.7 %
Types of examination	44% multiple choice, 37% other written examinations, 19% oral examinations, of which 67% are structured cases	○	70% multiple choice, 30% other written examinations, 30% oral examinations of which 20% are structured cases (including OSCE examinations)
Allocated beds per 100 students	★★☆ 64	★☆☆ 45	★☆☆ 44

3.1.2 CHE Excellence Ranking

The CHE Excellence Ranking identifies universities, or rather the appropriate departments of those universities, which are excellent in biology, chemistry, mathematics, physics, political science, economics and psychology.

The CHE Excellence Ranking does not combine the results of individual rankings into a final score and does not use the results to produce a single league table. Hence, no weights are applied to the individual indicators.

Which universities are considered for the ranking and which fields are covered?

The CHE Excellence Ranking is a two-step exercise. In the first stage, universities which excel in the specific fields are pre-selected. This selection is based on the number of 'stars' awarded on the basis of pre-selection indicators (see below). The pre-selected departments are then analysed in depth.

CHE started its Excellence Ranking in 2007, then covering natural sciences only. In 2009, political science, economics and psychology were also included in the analysis. In 2010, CHE repeated the survey on natural sciences and combined the results with the 2009 results in political science, economics and psychology (Berghoff et al., 2010).

Pre-selection criteria

The pre-selection is based on the number of 'stars' that a university has received in the respective field. For economics, political science and psychology, the institution is pre-selected if it has two 'stars' altogether, with at least one of them in publications or citations. For the natural sciences and mathematics, the institution is pre-selected if it has two 'stars' in publication or citation indicators or three 'stars' altogether.

A 'star' is allocated to those institutions which account for at least 50% of the total achievement in the field. However, to be considered the university must have at least 3000 publications in the Web of Science, including publications from 1997-2007 for the natural sciences and mathematics, as well as publications in all other subjects for 1999-2006.

Indicators used for pre-selection

1. **Number of publications in the Web of Science** (applied to all fields). A 'star' is allocated to those institutions which belong to the group of institutions with a high number of publications, and, as a group, account for 50% of the total number of publications. In other words, this is the way in which a ranking of universities by number of publications is prepared. A threshold is drawn at 50% of the total number of publications, with those universities above this threshold receiving a 'star'.

2. **Citations** (applied to all fields). Field-normalised citations per publication (CPP/FCSm) or the 'crown indicator' of the Leiden Ranking are used (see details on this indicator in the chapter on the Leiden Ranking). This indicator is calculated as a ratio between the number of citations per publication[47] (CPP) and the average number of citations per publication in the same field in the same year (FCSm). This means that, if publications of a given university are cited with a frequency that is 'typical' for the field, the CPP/FCSm indicator will have the value 1. A value above 1 indicates that publications of the university in the particular field receive more citations than is 'usual'. In the CHE Excellence Ranking, a 'star' is allocated to universities which have a CPP/FCSm value of at least 1.1.

3. **Outstanding researchers** (applied to the natural sciences and mathematics). The university receives a 'star' in a particular field if any Nobel Prize winners, winners of Körber European Science awards or Fields medallists in mathematics are currently teaching at the institution. It is sufficient to have one such outstanding researcher.

4. **Number of projects in the Marie Curie programme** (applied to the natural sciences and mathematics). Based upon information drawn from the Cordis database, the following Marie Curie activities are taken into account:

   - IEF – Intra-European Fellowships for Career Development
   - IRG – International Reintegration Grants
   - ITN – Initial Training Networks
   - ERG – European Reintegration Grants
   - IAPP – Industry-Academia Partnerships and Pathways
   - IOF – International Outgoing Fellowships for Career Development
   - IIF – International Incoming Fellowships
   - IRSES – International Research Staff Exchange Scheme

   The approach to allocating 'stars' in this context is similar to the publication indicator. However in practice, to obtain a 'star', three projects are needed in biology, two in physics and chemistry and only one in mathematics.

5. **Student mobility** (applied to all fields). A 'star' is allocated to universities with the greatest mobility and which belong to the group that accounts for 80% of mobile students (50% in sciences) overall. Although this is not stated explicitly, the context appears to suggest that the reference group is that of incoming postgraduate students. The overall number of mobile students needed to obtain a 'star' varies from 35 in economics to 20 in physics.

[47] Excluding self-citations

6. **Teaching staff mobility** (applied to all fields). Staff members who have undertaken a short teaching period in the context of the Erasmus programme are counted for this indicator, with points being assigned to both the sending and the receiving institution. Universities with the highest mobility and which form part of the group that accounts for 80% of mobile teachers (50% in sciences) receive a 'star'. In practice, 3-4 mobile teachers are enough to receive a 'star'.

7. **Erasmus Mundus Master** (applied in all fields). A 'star' is allocated for a joint master programme in the Erasmus Mundus Programme. Sometimes a 'star' is awarded to only one department, rather than all partners, because, as noted by CHE, other partners may cover parts of the programme that do not belong to the academic fields covered by the Excellence Ranking. This is carefully checked.

8. **European Research Council grants** (applied to the natural sciences and mathematics). A 'star' is given to both the sending and receiving institution for each grant allocated in 2007 and 2008.

9. **Book citations** (applied to economics, political science and psychology). CHE states that, while the book citations indicator cannot provide an analysis that corresponds precisely to article citations, it is an important way to avoid discrimination against those fields in which book publications constitute the main way of publishing research results.

Although the indicators are not weighted, some of them are more important than others for the 'pre-selection' process or, in other words, for being included in the Excellence Ranking's final results. To be 'in', a university needs at least 3000 publications in the Web of Science, and it requires a certain number of 'stars' with regards to the publications or citations indicators.

So far, the CHE Excellence Ranking has identified and mapped excellence centres in the natural sciences, economics, political science and psychology in Europe. The reports were published in 2009.

In-depth analysis

The pre-selected universities are analysed further. Data is collected using institutional surveys and surveys of students in master and doctoral studies. Students answer questions regarding, for instance, the overall study situation, availability of advisors and quality of counselling and career centres, examinations, laboratories, library, training, study organisation, IT infrastructure, counselling, websites, rooms, social relations, conference attendance, research community, time taken to complete a PhD project, workrooms and workshops. The institutional survey includes various types of information on staff, students and the university.

Presentation of results. The results of the CHE Excellence Ranking are presented on the *Die Zeit* website. When selecting one of the academic fields covered by this ranking, the 'Excellence list' of those institutions that have been pre-selected in the chosen field appears. Lists are in alphabetical order. Users can choose two to three universities for comparison. On this basis more detailed information can be retrieved. This includes data gathered through the pre-selection procedures as well as through the student and institutional surveys. One can also access and compare information on the relevant study programmes in those universities.

As far as the choice of lists of research teams is concerned it is also possible to see the results in a different format. The user then has to choose a sub-area of the academic field (e.g. analytical chemistry), and a list of the research group appears. The user will then see the name of the group leader, the members of the group, their specific research interest as well as a short list of the team's most important publications.

Overall, the CHE Excellence Ranking is a good information tool allowing potential postgraduate students to determine which universities are excellent in their particular field(s) of interest.

3.1.3 Other CHE rankings

The CHE Research Ranking is a multi-indicator exercise that compares the success of German universities according to a set of indicators without compiling a league table.

The CHE Employability Rating is aimed at assessing bachelor programmes on the basis of how they promote skills and competencies that contribute to the professional capacity of the graduate. The first report was published in 2008.

3.2 U-Map classification of HEIs – CHEPS

The original aim of the U-Map tool was to design a European higher education classification tool that reflects the variety of missions and profiles of European higher education institutions (U-Map, 2010, p.11). The tool will focus on the differences between institutions (institutional diversity) in terms of their missions and profiles (horizontal diversity). The U-Map project has been funded by the European Union and is led by the Centre for Higher Education Policy Studies (CHEPS) of the University of Twente, the Netherlands.

U-Map is a multi-indicator tool that does not calculate an overall final score for a higher education institution, and hence

does not produce a league table. Efforts have been made to present the data in such a way that prevents other parties from using the data to produce a league table. The absolute values of indicators can only be seen when three HEIs are selected and compared indicator by indicator. It would indeed require enormous efforts to gather all the information needed to produce a league table from U-Map data – although it is not absolutely impossible. The classification tool is aimed at all European HEIs, without discrimination as to the type of or area covered by the institutions.

The U-Map project started in the second half of 2005 and its final report was published in January 2010. At the time of compiling this report, only pilot 'pre-fillings' had begun, involving HEIs in Norway and, since June 2010, the Netherlands and the Flemish community of Belgium.

Indicators and the categorisation of the results

Indicators are grouped into six 'profiles': *Education profile, Student profile, Research Involvement, Knowledge exchange ('3rd mission'), International orientation* and *Regional engagement* (see Table 11). Because the values of the indicators are visualised (see the section *Visualisation tools* below), the results for all but three indicators are categorised as *major, substantial, some* or *none*. In the visualisation tool *Profile viewer,* the four result categories are visualised by using four different sizes of the circle sector assigned to the particular indicator. While most indicators in the list are self-explanatory, some further comments are given in Table 11, mainly regarding the indicators that do not simply follow the *major-substantial-some-none* scheme (all those indicators belong to the *Education profile*).

Table 11. Profiles and indicators

Indicators	Calculation	Cut off points
Teaching and learning profile		
Degree level focus	% doctoral degrees in total number of degrees awarded % master degrees in total number of degrees awarded % bachelor degrees in total number of degrees awarded % sub-degrees in total number of degrees awarded	≥5% doctoral degrees – doctorate focus; ≥25% master degrees – master focus; ≥40% bachelor degrees – bachelor focus; ≥5% sub-degrees – sub-degree focus;
Range of subjects	Number of subject areas covered out of UNESCO/OECD 8 broad areas	>6 areas – comprehensive; 3 to 6 areas – broad; ≤3 areas – specialised
Orientation of degrees	share of qualifications awarded: • in regulated professions • other career-oriented • general formative	Focus[48]: >1/3 graduates in programmes leading to regulated professions – regulated professions focus; >1/3 graduates in other career-oriented programmes; >1/3 graduates in general formative programmes – general focus
Expenditure on teaching	% expenditure on teaching	>40% – major; 10% to 40% – substantial; 1% to 10% – some; <1% – none
Student profile		
Mature students	% of mature students	>20% – major; 10% to 20% – substantial; 5% to 10% – some; <5% – none
Part-time students	% of part-time students	>20% – major; 10% to 20% – substantial; 5% to 10% – some; <5% – none
Distance learning students	% of distance learners	>20% – major; 10% to 20% – substantial; 5%-10% – some; <5% – none
Size of student body	Total enrolment count	>30,000 – very large; 15,000 to 30,000 – large; 5,000 to 15,000 – medium; < 5,000 – small

[48] Mixed orientation is possible if >1/3 degrees is reached in in two or three categories.

Research involvement		
Peer reviewed publications[49]	Publications per academic staff	>2 – major; 1 to 2 – substantial; 0.1 to 0.75 – some
Doctorate production	Doctoral degrees awarded per academic staff	>1.5 – major; 0.75 to 1.5 – substantial; 0.1 to 0.75 – some
Expenditure on research	% expenditure on research	>40% – major; 10% to 40% – substantial; between 1-10% – some; <1% – none
Knowledge exchange		
Start-up firms	Start-up firms count per 1000 academics	>10 – major; 5 to 10 – substantial; 1 to 5 – some; <1– none
Patent applications filed	Patent applications count per 1000 academics	>10 – major; 5 to 10 – substantial; 1 to 5 – some; <1 – none
Cultural activities	Count of exhibitions, concerts, performances	>100 – major; 50 to 100 – substantial; 0 to 50 – some
Income from knowledge exchange activities	% income from licensing agreements, contracts, copyright and donations	>40% – major; 11% to 40% – substantial; 1% to 10% – some; <1% – none
International orientation		
Foreign degree-seeking students	% of degree-seeking foreign students in total student count	>7% – major; 2.5% to 7.5% – substantial; 0.5% to 2.5% –some; <0.5% – none
Incoming students in international exchange programmes	% incoming students from international exchange programmes in total enrolment	>2% – major; 1% to 2% – substantial; 0 to 1% – some; <0.5% – none
Students sent out in international exchange programmes	% of outgoing students within international programmes	>2% – major; 1% to 2% – substantial; 0 to 1% – some; <0.5% – none
International academic staff	Foreign staff as % of total staff headcount	>15% – major; 5% to 15% – substantial; 1% to 5% – some; <1% – none
The importance of international income sources	% of income from non-national sources excl. tuition fees	>10% – major; 5% to 10% – substantial; 1% to 5% – some; <1% – none
Regional engagement		
Graduates working in the region	% graduates working in the region	>10% – major; 5% to 10% – substantial; 1% to 5% – some; <1% – none
First year bachelor students from the region	% first year bachelor students from the region	>10% – major; 5% to 10% – substantial; 1% to 5% – some; <1% – none
Importance of local/regional income sources	% local/regional income sources	>10% – major; 5% to 10% – substantial; 1% to 5% – some; <1% – none

[49] Books and monographs are considered. It is unclear whether journals are comparable accross countries.

Source: information from (U-map, 2009a)

Degree level focus. Four data elements form the *Degree level focus* indicators. These four indicators are calculated simply as a percentage of doctoral, master, bachelor and sub-degrees[50] in the total number of degrees awarded. The HEI is labelled as having a focus on the doctorate level if the share of doctoral degrees in the total number of degrees awarded is 5%. A master focus requires 25% master degrees, a bachelor focus 40% bachelor degrees, and a sub-degree focus requires 5% sub-degrees. Given that the above percentages amount to only 75%, it is possible for an HEI to fulfil conditions for several, or even all conditions. Thus, an HEI can have a mixed focus. No explanation is provided (at least in the public part of the U-Map website) of how the lower percentages for those degrees are linked to the three smaller sizes of the appropriate segment in the 'sunburst chart'. For instance, if an HEI has 15% bachelor students, how will the *Degree level focus* indicator for bachelor students be visualised in terms of the size of its circle segment?

The indicator *Range of subjects* is contingent on how many of the eight UNESCO/OECD broad subject areas are covered by a particular HEI: universities with more than six areas are labelled as 'comprehensive', those with three to six as 'broad' and those with fewer than three as 'specialised'. However, an analysis of the actual number of broad study areas of those HEIs labelled as 'specialised' in the U-Map web shows that the situation on the ground is somewhat different. In some cases, the 'specialised' institutions actually offered programmes in all eight UNESCO/OECD areas. This seems to suggest that additional criteria are being used that have not been mentioned or described on the public website.

The indicator *Orientation of degrees/qualifications* distinguishes between programmes leading to licensed/regulated professions, those leading to other career-oriented programmes and general programmes. The appropriate orientation label is assigned if the share of the respective programmes reaches 1/3. Again, an HEI may therefore have two or theoretically even three orientations.

The indicator *Expenditure on teaching* is described but not actually used; at least it does not appear in the publicly available visualisations of the results.

There are several indicators such as *Number of start-up firms, Patent applications, Cultural activities, Peer-reviewed publications,* and *Doctorate production per academic*, which should be measured over a set period of time. While this is mentioned, no information is available regarding the period for which each of those indicators is measured.

Overall, it can be said – and this is considered as appropriate for a classification system – that the U-Map indicators are more closely linked to the focus and intensity, rather than to the quality of the various activities of HEIs. For instance, the publications count is used, but not citations or impact, while patent applications are counted, but not patents issued, etc. This latter point is not a criticism, but instead a caveat concerning the limits of a system that has been designed as a classification. It may also be one reason why a next phase, the U-Multirank, is being pursued.

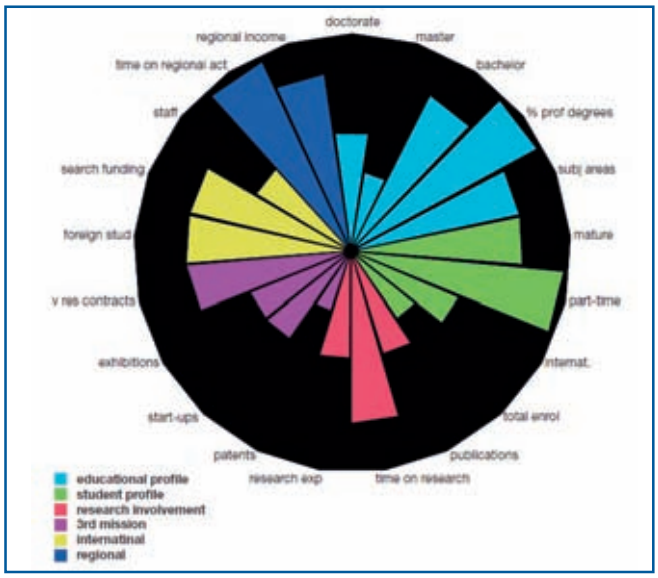
Visualisation of U-Map results

The results for an HEI are visualised using two online tools: *Profile finder* and *Profile viewer*.

With the *Profile finder,* it is possible to find HEIs according to a combination of criteria defined by the user. The *Profile finder* screen allows the user to choose a combination of indicators and categories for each indicator. The system retrieves those universities which fit the requirements set by the user. For instance, if the user chooses a *medium-size* university with a *mixed* focus, some cultural activities, *substantial* regional funding and *some* foreign staff, the system currently[51] retrieves 2 out of the 67 HEIs in the system. The user can then compare the HEIs retrieved by the *Profile finder* in greater detail by using the *Profile viewer* tool.

In the *Profile viewer,* each HEI is presented by means of a 'sunburst chart' (see Fig. 5). In this 'sunburst chart', each indicator is represented by a disc or circle sector. The sector may have four different sizes or lengths. With the exception of the three indicators described above, the four different sizes of the sectors are used to denote *major, substantial, some* and *none*, as described above. Although setting cut-off points for each indicator involves some subjective decision-making, it is possible to read the actual indicator values by pointing the cursor at the segment related to the particular indicator (see Fig. 6). *The Profile viewer* tool allows for the selection of up to three HEIs included in the database for indicator-by-indicator comparison.

Figure 5. The 'sunburst chart' of a HEI

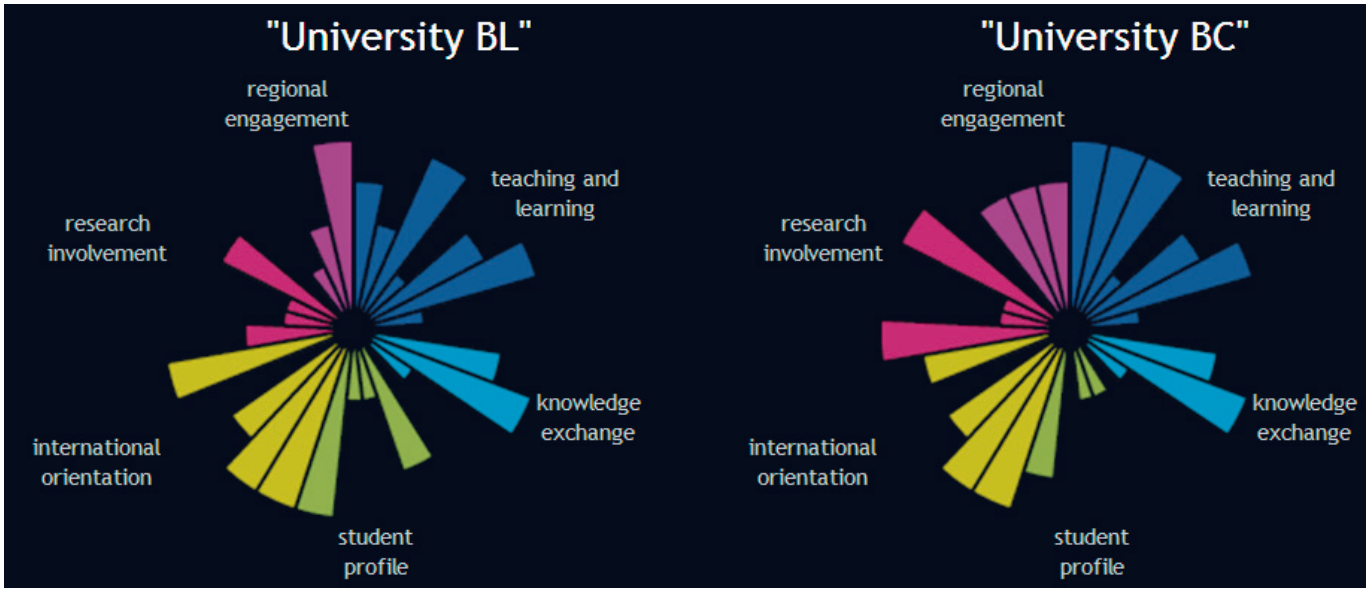


Source: U-map, 2010

[50] No explanation is provided in the report. However, the range of data collected allows us to suggest that the 'sub-degrees' are qualifications awarded after completion of short professional programmes.

[51] Tested on 1 Nov 2010

Figure 6. Comparing three HEIs in *Profile viewer*



Source: U-map.eu

Which universities are considered?

The classification tool is aimed at all European higher education institutions.

Data sources

Adequate data sources seem to be the main challenge for the use of U-Map as a tool for the international comparisons of HEIs. The U-Map team has explored possibilities to acquire comparable data on HEIs from different countries. Their conclusion is that a Europe-wide data system does not (yet) exist, and that it is indeed difficult to create such a system on the basis of (existing) national data sets (U-Map, 2010, p. 18). The European Commission and EUROSTAT have launched an initiative to support the development of a European higher education and research 'census', but this activitiy is still at an early stage. For this reason, U-Map is mainly using national data or data on individual HEIs. In its 'pre-filling' pilot phases, U-Map has used data from national higher education databases and from surveys completed by the HEIs themselves. The enormous differences in the ways countries define, collect, interpret and use data explains why there is lack of comparable data at the European level.

Thus it would appear that until these new European initiatives in data collection start to yield results, comparisons inside a single country will continue to work much better than at the European level.

3.3 European Multidimensional University Ranking System (U-Multirank) – EU funded project

The European Multidimensional University Ranking System (U-Multirank) is an EU-funded project aimed at creating a global ranking of universities that should, ideally, steer clear of the main drawbacks of current global university rankings. According to the objectives set by the EU (EU Commission, 2009), the ranking should be **multi-dimensional**, i.e. covering the various missions of institutions, such as education, research, innovation, internationalisation, community outreach and employability; **independent**, i.e. not be run by public authorities or universities; and **global**, i.e. cover institutions inside and outside of Europe. The project is being carried out by the CHERPA network led by the Centre for Higher Education Policy Studies at Twente University, the Netherlands, and the *Zentrum für Hochschulentwicklung* (CHE), Germany. The project is ongoing, and the only available source of information is the first interim report (CHERPA, 2009) on the design phase of the ranking. A thorough description of U-Multirank will thus be prepared after the project has been completed and the results published.

The U-Multirank project produces two rankings: a *Focused institutional ranking* and a *Field–based ranking*.

The *Focused institutional ranking* enables comparisons of institutions according to a single dimension of institutional activity, such as education, research, internationalisation

or knowledge transfer (CHERPA, 2010, p. 77). According to CHERPA (2010), the scores in these different *dimensions* will not be combined into an overall score. The question remains, however, whether it will be possible to present the U-Multirank results in such a way as to prevent third parties from combining the results and producing a league table. There is serious concern that this may be inevitable (see e.g. Boulton, 2010, Para. 28).

The *Field-based ranking* will be designed as a multi-dimensional ranking of a set of study programmes in a specific field or discipline, provided by institutions with a comparable profile. According to the U-Multirank website[52], achieving field-based rankings based on rational classification of institutions is a major aim of the project, since rankings of study programmes can only be meaningfully interpreted within the wider context provided by the multi-dimensional classification of the entire institution.

Which universities are considered for the ranking?

The U-Multirank will cover institutions inside and outside Europe, in particular those in the US, Asia and Australia (EU Commission, 2009).

Areas covered

The U-Multirank approach is multi-dimensional. It covers the various missions of institutions *(dimensions)*, as described above. The EU Commission emphasises that the existing rankings tend to focus on research in 'hard sciences' and ignore the performance of universities in areas like humanities and social sciences, and aspects such as teaching quality and community outreach. Therefore, the Commission states, the U-Multirank should attempt to cover all study fields and 'dimensions' properly (EU Commission, 2009).

Indicators and proxies

The U-Multirank initial list of indicators has been largely based on those developed by CHE for the CHE University Rankings for Germany. The indicators are divided into groups of *Enabling* (further divided into *Input* and *Process* ones) and *Performance* (further divided into *Output* and *Impact* ones) indicators.

Some indicators are to be used for both the *Focused institutional ranking* and the *Field-based ranking*, while others will be used for only one of these purposes. The lists of indicators had not been finalised when this survey was prepared and are therefore not presented.

Data sources

The interim report of the U-Multirank project (CHERPA, 2010) provides an excellent analysis of the advantages and disadvantages of the various available data sources. However, at the time of writing this report, it was not yet clear what measures will be taken in order to improve data collection for U-Multirank compared to existing rankings.

One can only agree with the U-Multirank team that the WoS and Scopus databases do not adequately reflect research in the arts and humanities, that books are not covered as well as journals and that the language bias remains. However, it remains to be seen how the U-Multirank project plans to avoid those pitfalls.

The interim report states that the bibliometric indicators will be based on Thomson Reuters and Scopus databases, and that patent databases will be used in addition. At the same time, self-reported university data will play a significant role in both the institutional and the field-based U-Multirank rankings (CHERPA, 2010, p.88). The self-reported data will relate to staff, students, resources and facilities, to research (except publications and citations), knowledge transfer (except patents) and regional engagement, as well as to teaching and learning. Student surveys will be the third type of data source used. U-Multirank will not draw on reputation surveys of academics, because the latter do not work well in international rankings (Federkeil, 2009).

Transforming indicator values into scores

There is no intention of calculating an overall score or to assign weights to individual indicators.

Changes in the methodology over time

The methodology is still being developed for the first phase of the ranking.

Presentation of the ranking and additional analysis produced

At the time of producing the current review, U-Multirank had not yet publicised exactly how its results will be presented. Information available indicates that both the *Focused institutional ranking* and the *Field-based ranking* will compare similar study programmes within groups of HEIs with a similar profile.

[52] http://u-multirank.eu/project/

4. Web Rankings

4.1 Webometrics Ranking of World Universities

The Webometrics Ranking of World Universities ('Webometrics') was launched in 2004. It is an initiative of the Cybermetrics Lab, a research group of the Centro de Ciencias Humanas y Sociales (CCHS), which is part of the National Research Council of Spain.

The stated aim of the project is "to convince academic and political communities of the importance of [...] web publication not only for dissemination of [...] academic knowledge but for measuring scientific activities, performance and impact, too" (http://www.webometrics.info).

The Webometrics Ranking measures the size and 'visibility' of university web pages (Aquillo et al., 2008). Size is characterised by the number of pages on the website of the university, as well as by the number of publications and of 'rich files' (*.pdf, .ppt, .doc* and *.ps*). The 'visibility' of the university is measured by the number of inward links to the university website.

The Webometrics team uses commercial search engines to collect data because "the websites can be trawled directly using specially designed robots that collect basic information through hypertextual navigation, or the statistics can be extracted from previously trawled databases obtained from commercial search engines. Despite coverage biases or other shortcomings, if a webpage is not indexed by them, then that page does not exist for any purpose" (Aquillo et al., 2009, p. 542). The Webometrics Ranking is updated every six months; data is collected in January and July and published one month later. Data collection is automatic, but the final positions of universities in the league table are calculated manually and comparisons with previous years are made.

Which universities are considered for the ranking?

Any university with an independent domain is considered, meaning that universities whose websites are within the domains of other institutions are not considered. More than 20,000 higher education institutions have been analysed and 12,000 included in the list; results for all other HEIs are not sufficiently meaningful. Where a university has several domains, they are all analysed, although only the highest ranked domain is included.

Areas covered, indicators and proxies

The Webometrics Ranking measures the volume, visibility and impact of university webpages with a special emphasis on scientific output.

The creators of Webometrics believe that a strong web presence provides information on a wide variety of factors that can clearly be correlated with the global quality of the university in question: "widespread availability of computer resources, global internet literacy, policies promoting democracy and freedom of speech, competition for international visibility or support of open access initiatives, etc." (Agillero et al., 2008, p. 235). Although more parameters are measured, such as the number of external links, the number of sub-domains and the number of visits to the website, just four indicators are included in the ranking. These relate to two criteria: *Size* and *Visibility.*

Size, i.e. the overall volume of information published is measured by three indicators:

1. Number of pages on the academic website of the university
2. Number of rich files *(.pdf, .ppt, .doc and .ps)* published
3. Number of published papers retrieved from Google Scholar.

Size indicators are a proxy for the intensity of the academic production of the university, albeit in a very particular way, combining research publications, presentations, teaching materials, raw data, drafts and other documents with relevance for research and teaching, as well as administrative information from the university (ibid.).

This requires an analysis of different file types in terms of their relevance for academic and scientific production. The four types of *rich* files have been selected as being most relevant.

The number of papers published and the number of citations of these papers is taken from Google Scholar. It should be noted that results from the Scholar database relate to papers, reports and other academic texts.

The *Visibility* of a university on the web is characterised by the number of external 'inlinks' established to their website, thus revealing the extent to which the university is 'interesting' to others.

Establishing a link to a university website can, to some extent, be compared to a citation. The link is established in order to refer to information on the website and to allow others to quickly retrieve the original information, or to read more in the original texts. Descriptions and weights of the four indicators are provided in Table 12.

Table 12. Indicators of the Webometric World University Ranking

Indicator	Definition	Description	Weight
Visibility (external links)	Total number of unique external links received (inward links) by a site can be only confidently obtained from Yahoo Search.	Results are log-normalised to 1 for the highest value and then combined to generate the rank.	50%
Size of university web	Number of pages recovered from: Google, Yahoo, Live Search and Exalead.	For each search engine, results are log-normalised to 1 for the highest value. Then for each domain, maximum and minimum results are excluded and every institution is assigned a rank according to the combined sum.	20%
Rich files	Number of Adobe Acrobat (**.pdf**), Adobe PostScript (**.ps**), Microsoft Word (**.doc**) and Microsoft PowerPoint (**.ppt**) files.	Numbers of these file types are extracted using Google. The results for each file type are log-normalised and then merged.	15%
Scholar	Number of papers and citations is extracted from Google Scholar.	These results from the Scholar database represent papers, reports and other academic items.	15%

Data sources

Data for the size of the university's website is taken from Google, Yahoo, Live Search and Exalead. The highest and lowest results are excluded.

Numbers of rich files are obtained using Google, as Google provides the technical possibility to retrieve the numbers of different kinds of rich files separately.

Commercial search engines are used because they already have well-designed and tested robots; they frequently update their databases and they have automatic tools that can be customised with powerful operators for data extraction (Agillero et al., 2008, p. 235). However, commercial search engines have their limitations and disadvantages, e.g. including inconsistent and rounded results, biases in geographical and linguistic coverage, or frequent and opaque changes in their working procedures (ibid.).

Calculating indicator values and transforming them into scores

The indicator value for a university is its position in the league table for that particular indicator. The final rank is calculated using the weights provided in table 12.

Changes in methodology over time

Webometrics has not changed its methodology, but the commercial search engines used make frequent and opaque changes in their working procedures (Agillero et al., 2008, p. 235).

Presentation of the ranking and additional analysis produced

The global league table is arranged according to overall rank. Each university's rank for each of the four indicators is also provided. Besides the main league table, rankings by continent and country are also provided.

Webometrics also carry out rankings of non-university research centres, business schools, hospitals and repositories.

5. Benchmarking based on learning outcomes

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

In 2008, OECD started discussions on international comparisons of quality in higher education provision. As a result, OECD launched an international study of what students in higher education know and can do upon graduation: the Assessment of Higher Education Learning Outcomes (hereafter referred to as AHELO).

The AHELO project aims to develop criteria that will make it possible to evaluate the quality and relevance of what students learn in institutions around the world (OECD, 2010). If successful, AHELO will therefore be able to provide information on those teaching strategies that are most effective in ensuring that envisaged learning outcomes are achieved. Ambitions are high: the developers of AHELO believe that the results will help HEIs reduce their drop-out rates, allow more students to complete their degrees successfully, and in this way foster equity, while also giving governments a tool to monitor the efficiency of HE systems (ibid.).

Three testing instruments are being developed: one for measuring generic skills and two for testing discipline-specific skills in economics and in engineering. A targeted population of students nearing the end of their first 3- or 4-year degrees will be tested (Lalancette, 2010).

The AHELO project has two phases[53]. The first phase, from January 2010 to June 2011 (OECD, 2010), includes the development of provisional assessment frameworks and testing instruments suitable for implementation in an international context. The second phase, from January 2011 to December 2012 (ibid.), will involve the implementation of the three assessment instruments in a small group of diverse higher education institutions. Contextual questionnaires for students, faculty and institutions will be developed in relation to each of the three assessments (generic skills, economics and engineering), in order to arrive at some initial assumptions about the relationship between context and learning outcomes. Implementation is scheduled to start in late 2011/early 2012 in about 10 higher education institutions per country for all three instruments (generic skills, economics and engineering). This will also include the implementation of the contextual surveys (ibid.).

However, the developers are still grappling with a number of important questions regarding the initial project phase in particular (Lalancette, 2010):

- Is it possible to develop instruments to capture learning outcomes perceived as valid in different national and institutional contexts?
- Do the test items perform as expected and do the test results meet pre-defined psychometric standards of validity and reliability?
- Is it possible to score higher-order types of items consistently across countries?
- Is it possible to capture information on teaching and learning contexts that contribute to explaining differences in student performance?

In addition, using learning outcomes for comparing the performance of teaching and learning processes can only be successful if the participating institutions and countries have actually adopted a learning outcomes-based approach in both the teaching process and in student assessment. While linking programmes and specific courses to learning outcomes is an accepted approach within the European Higher Education Area, recent progress reports of the Bologna Process demonstrate that it will take time before this practice becomes general (Vesterhijden et al., 2010, Rauhvargers et al., 2009: 25).

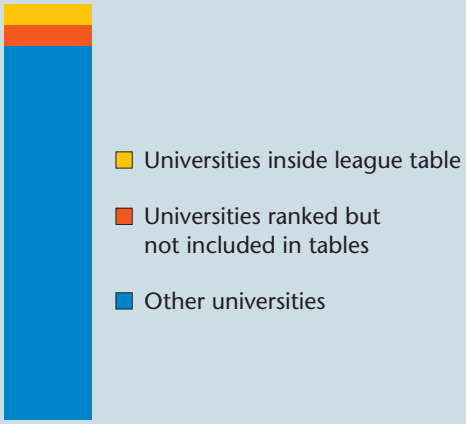
[53] The different publications and presentations of the project developers are somewhat contradictory as to the timing of the two phases. Information from the report published in December 2010 (OECD, 2010) is presented in this review, as it was the most recent at the time of writing.

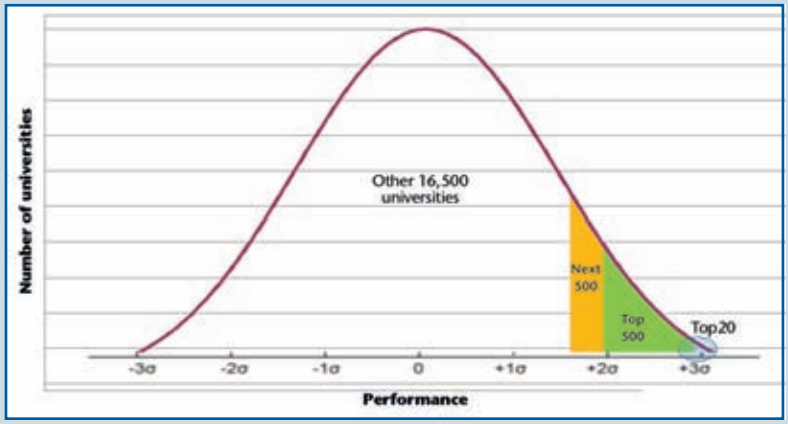
III. ANALYSIS OF RESULTS
(WHAT DOES IT MEAN FOR UNIVERSITIES)

Which universities are considered in world university rankings that produce league tables?

The most popular global university rankings (ARWU, THE-QS and THE-Thomson Reuters, USNWR, HEEACT, Reitor and others) consider the world's top research universities only. The criteria for these rankings are such that their application retrieves only around 1200 universities for ARWU, as well as some 600 universities overall and approximately 300 universities in each subject area for the THE-QS Ranking. In other words, these ranking methodologies are designed in such a way as to make the results meaningless for universities outside this limited group. This means that, although more universities might like to know how they rank – even if they are outside of the Top 100 or 500 – this is impossible: the methodologies of global rankings simply do not allow it.

Proportion of universities considered by existing global rankings vs. the total number of universities in the world





Thus, out of around 17,000 universities in the world, the global rankings consider only a tiny proportion. While these top class universities are very important, this also causes problems or 'unwanted consequences', as the ranking compilers phrase it: indeed, policy-makers and society in general increasingly tend to measure all higher education around the world by the standards of a limited number of top research universities. It it important to emphasise that merely being included in the global league tables already indicates that universities belong to the world's research elite. Countries, politicians and universities themselves regularly voice their ambition to see their university or groups of universities enter the Top 20, appear in the Top 100, or, indeed, simply enter the ranking lists in the future. Such statements are frequent, although they are seldom realistic. In Jamil Salmi's words, "how many universities can be among the Top 500? Five hundred" (Salmi, 2010).

Most global league tables also publish lists with the 'performance' of individual countries. These comparisons are made by counting the number of universities from each country that appear in global lists of top universities. Different numbers of points are usually assigned to universities, depending on whether they appear in the Top 100, the Top 101-200 and so on. According to this calculation, the leading countries in the published lists are the USA, the UK, Germany

and France. However, if the published lists are 'normalised' by correlating the number of top universities to the number of inhabitants, then new leaders, such as Switzerland, Sweden, Finland and Denmark appear (Salmi, 2010).

What rankings can and cannot offer. Rankings are hotly debated and strongly criticised. However, despite their many shortcomings, flaws and biases,"rankings enjoy a high level of acceptance among stakeholders and the wider public because of their simplicity and consumer-type information" (AUBR Expert Group, 2009). Those who say that university rankings are not going to disappear are probably right. Since the number of rankings is even expected to increase (Marginson, 2011), it is worth looking at both the positive and negative effects that rankings might have.

Among the positive implications of university rankings are the following. First, university rankings could help students choose the appropriate university in their home country or abroad. However, in order to serve this purpose, rankings would need to provide better explanations of what the indicator scores actually mean. The use of more 'democratic' indicators for selecting universities would also be helpful, as this would mean that the league tables would no longer be limited to the world's top research universities. As regards data collection at the national level, rankings encourage the collection and publication of reliable data on higher education. In an international context, rankings encourage the search for common definitions of those features for which data is collected. In addition, the results of global rankings can foster national debate and a focused analysis of the crucial factors underpinning success in rankings, which in turn may lead to positive policy change.

Nevertheless, rankings, at least those which produce global league tables, cannot provide a diagnosis of the whole higher education system, as they usually concern only the top research universities. In addition, current global rankings provide little useful information on issues such as the quality of teaching and learning, accessibility, regional involvement, involvement in lifelong learning, cost efficiency and other aspects, simply because the indicators used do not take account of these elements. In her recent book, Ellen Hazelkorn argues that "rankings also ignore the contribution, for example, of the creative/cultural industries to innovation or the way in which social innovation brings about fundamental change to the social economy via new forms of mutual action, new ways in which economies can be managed, new forms of consumption, and the organisation and financing of government" (Hazelkorn, 2011).

Rankings and the research mission of universities

The research performance of universities is far better covered in the rankings than teaching. The main groups of indicators and the ways in which these are used by the most popular global rankings is summarised as follows:

Number of publications, which is used in the following ways:

- ARWU includes two indicators with an overall weight of 40% of the total score: publications in *Nature* and *Science* and publications indexed in the Science Citation Index-expanded and the Social Science Citation Index
- HEEACT has four indicators with an overall weight of 50%: number of articles per staff in the past 11 years (10%), number of articles per staff in the previous year (10%), number of highly cited papers in the past 11 years, with publications belonging to the 1% most cited papers in ESI (15%), and articles published over the last year in one of the top 5% journals by impact (15%)
- THE-QS Ranking (up to 2009) does not use numbers of publications
- THE-TR (since 2010) includes one indicator with a weight of 4.5%: the number of publications per academic staff in academic journals indexed by Thomson Reuters
- Leiden Ranking includes one indicator (no overall score): the number of publications in journals covered by citation indexes
- Reitor has one indicator: the number of staff publications from the Scopus database (weight unknown, this indicator belongs to a three indicator 'block', which has an overall weight of 20%)
- CHE includes one indicator (no overall score): publications per academic staff
- U-Map has one indicator (no overall score): peer-reviewed publications[54] per academic staff
- U-Multirank will include the following indicators on the number of publications (no overall score): publications in international peer-reviewed scholarly journals in relation

[54] Books and monographs are included. It is unclear whether journals are comparable across countries (U-Map, 2010).

to staff FTE, number of top-cited publications, as well as indicators on several kinds of joint publications (within a country, regionally, and internationally).

The number of publications is thus being used in variety of ways: ARWU and the Leiden Ranking use it as a count of the actual number of publications, but it should be noted that, if the numbers are taken from the ESI or Scopus databases, they exclude publications in book form, thereby discriminating against those subject areas in which different publishing traditions exist[55]. The number of publications can also be used in the sense of publications per staff (THE-TR, CHE, U-Map), in order to suggest publication intensity, or as the number of publications with the highest impact, in order to indicate research excellence (HEEACT, U-Multirank).

Number of citations, which is also used in different ways:

- ARWU uses the number of citations indirectly, rather than directly (20% weight): referring to the number of staff belonging to the 200 top-cited academics in 21 broad subject fields

- HEEACT uses three indicators (30% weight overall): the number of citations per staff in the past 11 years (10%); the number of citations per staff in the past 2 years (10%); the average number of citations per publication in the past 11 years (10%)

- THE-QS (up to 2009) uses one indicator (20% weight): the number of citations per staff FTE using data from Scopus (Thomson Reuters in 2004-2006)

- THE-TR (in 2010) uses one indicator (32.4% weight): field-normalised[56] average citations per paper in the period 2004-2008

- Leiden uses four indicators (no overall score): 'simple' citations per paper; two normalised indicators: field-normalised citations per publication (CPP/FCSm) and mean-normalised citation score (MNCS2)[57]; and the 'brute force' indicator, which is calculated by multiplying the normalised number of citations per publication by the total number of publications

- Reitor uses one indicator: citations and references to staff publications (weight unknown, this indicator is one of three indicators in an indicator 'block' with an overall weight of 20%)

- CHE uses one indicator (no overall score): citations per publication

- U-Multirank will include one indicator on the field-normalised citation rate (no overall score).

As with the number of publications, the number of citations is thus also used in different ways. It can be used as the actual number of citations (Reitor[58]). In other rankings, the number of citations is used as citations per staff FTE (HEEACT, THE-QS), or citations per publication (THE-TR, CHE, U-Multirank), or both (HEEACT, Leiden Ranking). In several rankings, citation numbers are field-normalised (Leiden Ranking, THE-TR, U-Multirank).

H-index. The H-index of the university is used both in Reitor and HEEACT. In the latter, the H-index of the university is obtained by considering the publications of the university in the last two years (20% weight). In the Reitor Ranking, the weight of the H-index is a non-specified part of an indicator 'block' with an overall weight of 20%.

Research reputation surveys. The 'Academic peer review' of the THE-QS (40% weight) is rather a reputational survey than a peer review as understood in QA terminology. Research reputation is used in the THE-Thomson Reuters Ranking (19.5% weight). Apart from the two THE rankings, reputational surveys are also widely used in the CHE University Ranking.

In the THE-QS Ranking 'peers' were asked to nominate up to 30 universities they considered excellent in their field of research; they did so by choosing from pre-selected lists. Details on the research reputation survey used in the THE-Thomson Reuters Ranking are not available.

Research income, where used, can also be applied in several ways:

- THE-TR has three indicators: research income scaled against staff[59] (5.25% weight), research income from public sources vs. total research funding (0.75%) and research income from industry per academic staff (2.5%)

- CHE University Ranking, which produces no overall score, uses third-party funding of research per faculty member (other fields)

- U-Map uses the percentage of expenditure allocated for research (no overall score)

- U-Multirank uses external research income per staff FTE (no overall score).

Research income is thus scaled to staff FTE (overall research income for THE-TR, U-Multirank; third-party research income

[55] It should be noted here that ARWU counts one publication in the Social Sciences Citation Index as two, which partly compensates for the bias regarding social sciences.
[56] The methodology of field normalisation used is not specified.
[57] See section on the Leiden Ranking on the differences between the two indicators.
[58] The methodology description can be understood both as number of citations to publication (written by) staff members or as publications per staff member.
[59] Numbers normalised for purchasing-power parity.

for THE-TR, CHE). These amounts may be corrected according to purchasing power parity (THE-TR). The percentage of resources allocated for research as part of the total income (THE-TR) or expenditure (U-Map) can also be used.

Intensity of PhD production: The number of PhDs awarded in relation to academic staff numbers is an indicator used by THE-TR (6% weight), U-Map and U-Multirank. THE-TR also uses the ratio of PhDs awarded relative to bachelor degree awarded (2.25%).

Rankings and the teaching mission of the universities

At least some of the compilers of global rankings underline that the issue of teaching quality is important in rankings; this is, indeed the case. The following describes how teaching is presently incorporated into the various rankings presented in this report:

In the **ARWU Ranking**, the *Quality of education* is measured as the number of alumni of an institution that have been awarded a Nobel Prize. It is debatable whether this indicator reflects the teaching or, rather, the research performance of the university where the Nobel Prize winner has studied. Even if it does characterise teaching, it does so in a very limited and particular way, and it does not even apply to some of the universities ranked in the first 500.

In the **THE-QS Rankings**, the peer review element deals with research reputation only. Some idea of teaching quality could be obtained from the results of the employer review, which asked respondents to select the 30 best universities from their perspective. The survey was mainly addressed to large companies (most of them QS clients) and the number of responses, worldwide, at just 3281 in 2009 (and 2339 in 2008), was rather limited. Thus, the only 'measure of teaching quality' is the strongly criticised staff/student ratio, which is an input factor. It could give some idea about quality if applied within the same discipline, but, even so, more effectively in a national than in an international comparison.

In the **CHE University Ranking**, a number of indicators are related to teaching and learning: these include the overall teaching/study situation, accessibility, consulting hours, courses offered, study organisation, teaching evaluation, E-learning possibilities, availability and state of the laboratories and libraries, preparation for the job market and practical support (all of the above are assessed by students), as well as the reputation of teaching (assessed by professors). Since the information is based on the experiences of other students, it may help young people in their choice of university and a specific study programme. However, these indicators are all proxies, rather than actual measures of the quality of teaching and learning.

The **THE-Thomson Reuters 2010 Ranking** includes five indicators in the category *Teaching – the learning environment:* these are a reputational survey on teaching, the number of PhDs awarded per academic, the proportion of PhDs and bachelor degrees awarded, undergraduates admitted per academic (i.e. the notorious student/staff ratio), and income per academic. All five indicators may characterise the learning *environment*, but, again, all of these indicators are proxies and therefore assess teaching in an indirect way. The reputational survey provides feedback on universities in which the teaching is considered to be the 'best' in one or more fields, while those universities offering just 'good' teaching do not receive a score. The award of a high number of PhDs is a good sign, but it is not directly linked to the quality of teaching at bachelor or master level. The student/staff ratio and income per academic are even more distant proxies.

Unsurprisingly, the indicators used by **U-Map** are designed for classification, rather than ranking or quality assurance, which is why the teaching-relevant aspects focus on degree level, the range of subjects, the orientation of degrees, and expenditure on teaching. These indicators do not make it possible to judge teaching performance or quality.

U-Multirank will include several indicators on *Education*[60]. However, while these characterise the teaching and learning environment, none of them measure the quality of teaching and learning directly. Some, such as (bulk) expenditure on teaching, may furthermore be difficult to use in international comparisons. One can of course measure time-to-degree and graduation rates, but this is also problematic, because it could tempt institutions to improve graduation rates simply by lowering standards. Indicators such as the relative rate of graduate unemployment and relative graduate earnings may provide some indication of the quality of teaching, but the results on these indicators also depend on the situation of, and changes in the national economy. The proportion of academic staff with doctoral degrees would have to take account of different types of higher education institutions and national traditions to be relevant; while indicators such as

[60] See tables U-Multirank 1 and U-Multirank 2.

student satisfaction with computer facilities, libraries, rooms and work experience as part of study programmes are useful, they still belong to the *enabling*, rather than the *performance* criteria.

Rankings based on the web performance of universities are even farther from being a direct measurement of teaching and learning performance or quality.

In conclusion, it appears that practically all the indicators used to characterise teaching and learning actually provide information on the teaching and learning *environment*[61], rather than measuring the quality of teaching and learning itself.

An attempt to evaluate the quality of teaching and learning is being made by the AHELO project, which aims at measuring actual learning outcomes. But this initiative is still in its initial stages, and it is, therefore, too early to tell how successful it will be.

Biases and flaws

Natural sciences and medicine vs. social sciences bias. Numerous previous publications have demonstrated that medicine and sciences are much better represented in bibliometric indicators than engineering, social sciences and especially humanities, which are often ignored completely (see, for instance, Leeuwen et al., 2001; Moed, 2005; van Raan, 2005; WG AUBR, 2010; CHERPA, 2010). To a large extent, the bias is caused by the fact that bibliometric indicators primarily cover publications. At the same time, as the EU Working Group on University-Based Research Assessment has underlined, "while natural and life scientists write books, their primary outlet is peer-reviewed journal articles. Engineering scientists primarily publish in conference proceedings although they also publish in journals and design prototypes. Social scientists and humanists have a wide range of outputs of which books are important sources of communication, while the arts produce major art works, compositions and media productions" (AUBR, 2010, p. 26).

Providers of the ARWU Ranking have recently undertaken a study (Cheng, 2010) in which publication and citation traditions in several fields are examined. Cheng demonstrates that, according to US data, in biological sciences the number of publications per staff member is 7.62, while it is 6.04 in mathematics and 2.14 in social and behavioural sciences, demonstrating huge differences in publication behaviour between different fields. As regards citations, biologists can claim 7.82 citations per paper, engineers 2.95 and social and behavioural scientists a mere 2.56. According to Chinese data, funding per academic staff in science, engineering and medicine is 3.2 times higher than in social sciences. Data on these and several other indicators point to the need, at the very least, to field-normalise the data to make the measurements more objective.

Field-normalisation – solutions and new flaws. Attempts have been made to compensate for the field bias. These include the Leiden Ranking 'crown indicator' – CPP/FCSm – which is calculated by dividing the sum of citations in the publications of a university by the sum of the expected number of citations in the same fields in the same year as the publication. However, because the most recent publications have fewer citations, they have little influence in the sums both as numerator and denominator, and hence the impact of new publications is very small[62]. The same goes for those fields in which there are traditionally fewer citations: they also have less impact on the sums in the numerator and denominator. Thus, although some normalisation happens, it is biased towards older publications and towards fields with an intensive citation culture. In addition, adding up all citations and all expected citation rates blurs the picture.

Another attempt is the mean-normalised citation score (MNCS), which is calculated by first dividing the number of citations of each publication by the expected citation number in the same year and field. Now, no fields are discriminated against, but since new publications have fewer citations, the results are more unstable. To correct this flaw, in MNCS2 indicator publications of the last year are simply ignored, thus running the risk of creating a new bias. In addition, this still fails to help those who publish books or in conference proceedings as these are not counted whichever mathematical approach in chosen.

Impact factor – to be used with care. Regarding the citation impact factor, it is also important to remember that "especially in social sciences and humanities, expert rankings do not correlate very well with impact factors. In these fields and in engineering, other sources are important as well (books, proceedings)" (AUBR, 2010, p. 43). It is also expected that "the coverage of both [the WoS and Scopus] databases is likely to remain unsatisfactory in those fields where neither journals nor conference proceedings papers are used by researchers and scholars as their main vehicle for knowledge dissemination: the arts and humanities in particular" (CHERPA, 2010). A warning is also posted on the Thomson Reuters

[61] Term used in the 2010 THE-Thomson Reuters Ranking.
[62] See discussion of the mathematical expressions and references in the description of the Leiden Ranking.

website, originally published in 1994: "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).

In her recent book, Ellen Hazelkorn comes to the conclusion that "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 bias. To begin with, the term 'peer review' itself is biased because it is used to denote quite different processes in QA and in rankings. In QA of both research and teaching, the term 'peer review' is used to denote an assessment by peers, with rigorous procedures. Historically, competent peers read the publications of the unit being assessed, be it a university, a research institute, a programme or department (see, for instance, U-Multirank, 2010). More recently, it has usually involved analysing the self-assessment report and other documents provided by the unit in question, typically followed by a peer visit to that unit. Certain concerns related to the impact of rankings on the peer have been discussed recently in some detail (AUBR, 2010; U-Multirank, 2010). The AUBR Working Group has summarised them as follows: "[peer] evaluators may be influenced by competitive pressures, including possible implications for their own work or that of their colleagues. They may evaluate research in terms of what they know [...]. In this way, novel and challenging ideas can be marginalised because they challenge established ideas. Finally, peers […] may be influenced by a researcher's reputation rather than his or her actual contribution to knowledge" (AUBR, 2010, p. 40).

Moreover, in rankings, the term 'peer review' usually refers to a simple reputation survey, though respondents may, indeed, be renowned academics. The fact that 'peer review' denotes a reputation survey in rankings is problematic. Firstly, even if a large number of academics is approached, few actually answer. THE-QS received 9386 responses (of which 6354 dated from 2008) after contacting approximately 180,000 people. The Reitor Ranking had to eliminate the reputation indicator completely as there were only a handful of respondents. Secondly, the 'peers' cannot freely nominate the universities they consider excellent. They choose from pre-prepared lists which, at least in the case of the THE-QS, failed to include many universities and even whole countries. Thirdly, the opinion of 'peers' is influenced by the reputation already acquired by the institution. This means that a university which already ranks highly in one ranking is very likely to obtain a high reputation score in another one. It may also lead to an institution being automatically considered as excellent overall, because it is known to be excellent in one area (AUBR, 2010, p. 20). There have been anecdotal cases in which peers have stated that a university is excellent in a field in which it is not active at all.

Ellen Hazelkorn observes that rankings are based on a somewhat outdated view of how research is carried out (Mode 1), rather than capturing the new organisational developments (Mode 2). She notes: "Mode 1 research achieves accountability and quality control via peer-review process, while Mode 2 achieves accountability and quality control via social accountability and reflexivity. Whereas Mode 1 relies upon a traditional elite model of knowledge creation, the latter democratises knowledge production, application and exchange [...]." (Hazelkorn, 2011). She comes to the conclusion that "rankings [...] reassert the hierarchy of traditional knowledge production" (ibid.).

Language bias. Since the publication of the first world rankings, it has been noted that global rankings favour universities in countries in which English is spoken, because non-English language work is both published and cited less (Marginson & van der Wende, 2007). According to Altbach, American scientists mainly cite other Americans and tend to ignore scholarship from other countries, which may artificially boost the ranking of US universities (Altbach, 2006). In addition, the ISI (now WoS) database mainly contains journals published in English, which are selected taking account of the practice in the academic systems of the United States and the United Kingdom (ibid.). This may also strengthen the field bias, since English is not always necessarily the central medium of communication in the humanities, law and a number of other fields (ibid.).

Regional bias. Ever since the onset of global rankings, some world regions, notably North America and Europe – and particularly Western Europe – have been better represented in the league tables than others. Indeed, global rankings implicitly refer to the Anglo-Saxon model of research organisation (CHERPA, 2010, p. 24).

The regional bias can to a large extent be explained by the same reasons as the language bias. However, different world regions may also have different publication and citation traditions; even within a particular field, citation and publication traditions may differ regionally. Reputation surveys are particularly problematic: indeed, there is a danger that universities from certain countries or regions will only be selected in reputation surveys if these already rank highly in existing league tables. THE-QS therefore applies a regional weighting to each subject area with a view to ensuring equal representation for four 'super regions': the Americas; Europe; the Middle East and Africa; and Asia Pacific (THE-QS, 2009). The THE-Thomson Reuters 2010 Ranking applies the mechanism *Accounting for international factors* (Pratt, 2010). In both cases, however, the regional weights and factors or the criteria on which they are based are not available on the public websites.

The dangers of taking ranking positions too seriously

The existence of rankings no doubt encourages universities to improve their performance. The question remains, however, exactly which type of actions they lead to. Striving to improve their position in the rankings, universities are strongly tempted to improve performance in those specific areas that are measured by the indicators used to prepare rankings.

Highly ranked universities have to make great efforts to keep their positions because their rivals evolve as well (CHERPA, 2010). 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 was later also articulated in relation to biological systems (Valen, 1973), 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).

The following are some examples of distortive action that may be taken just to perform better in university rankings:

- More Nobel Prize winners among staff – Most universities have none, but why not try to hire one?

- More publications and more publications per FTE – this could be a positive development, but in practice it simply means more articles in Thomson Reuters – or Scopus-covered journals, rather than in books and other types of publication. If attempts are made to improve research performance solely to improve ranking scores, this may lead to the one-sided support of research in medicine and sciences at the expense of the social sciences and by completely excluding research in the humanities.

- More citations: more citations per FTE and more citations per paper - Although these elements are more difficult to manipulate if the ranking in question uses field-normalisation, reduced support for the humanities and social sciences and a greater focus on the 'hard' sciences and medicine might benefit ranking scores.

- For all indicators that involve FTE (research staff, teaching staff, all academic staff combined), performance may be 'improved' by playing with the definitions of categories of staff. In rankings that use the staff/student ratio as the only proxy to characterise the quality of teaching, making changes can lead to remarkable improvements.

- As several rankings that produce league tables use indicators which reflect the overall product of the university (overall number of Nobel Prize winners, articles, citations, etc.), mergers of universities can either improve the position in these rankings, or help institutions to enter the league tables in the first place.

- Indicators on the proportion of international staff and students depend on the definition of these categories. Therefore, the use of such indicators in global rankings is prone to manipulation, as long as there are no exact definitions, for example as to whether domestic students (or staff) with foreign citizenship can be counted as being 'international'.

There are a number of examples of actual manipulation in order to obtain higher scores in rankings. Several cases of the manipulation of different indicators in the USNWR Ranking appear in the U-Multirank interim report (CHERPA, 2010, p. 60):

- Some institutions have put considerable effort into encouraging students to apply, even though they had no chance of ever being accepted (Schreiterer, 2008), thus ensuring that the institutions receive a higher score on selectivity.

- In order to raise the average standardised test scores of applicants, some institutions have made the submission of test scores voluntary, in the knowledge that only applicants with a high score have an incentive to do so.

- As USNWR counts full-time faculty for its student/staff ratio in the autumn term, departments have encouraged their faculty to take academic leave in spring, not in autumn (Espeland & Sauder, 2007).

- In an attempt to demonstrate how flawed the student/staff ratio indicator is, Marney Scully from the University of Toronto has shown how student/staff ratios, using the same data, can vary from 6:1 to 39:1 (Baty, 2010b).

- Faculty salary is also taken into account in rankings, and there are examples of institutions increasing salaries for this reason.

- In a well-documented case in the UK in 2008, more than 100 students were told to give glowing reports of their university to improve its standing in the rankings.

- If a ranking includes a reputation survey referring to the reputation of other institutions, it is clearly in an HEI's interests to manipulate these results (CHERPA, 2010). For this reason, CHE is gradually eliminating academic surveys from its rankings.

- The Working Group AUBR discovered that even bibliometric indicators might be flawed due to the manipulation of data (AUBR, 2010, p. 13).

Improving ranking scores vs. fulfilling other important HEI tasks. Efforts by European universities to improve their positions in the rankings may keep them from concentrating on elements of their mission that do not directly influence the ranking scores. The authors of EUA's Trends 2010 report warn against the development of indicators and statistics for rankings, given the mixed effects this could have on higher education institutions and students. It could potentially weaken the focus on partnerships in the Bologna Process, on quality development and improvement and thus dilute the central philosophy underpinning Bologna (Sursock & Smidt, 2010, p. 27). In other words it can be argued that supporting the implementation of the Bologna Process is unlikely to improve ranking scores, and that a perceived need to improve ranking scores could lead to resources being allocated to actions perceived to be more conducive to this end.

Thus, paying too much attention to improving ranking scores can be detrimental to the fulfilment of other important tasks of higher education institutions.

IV. MAIN CONCLUSIONS

1. There is no doubt that the arrival on the scene of global classifications and rankings of universities has galvanised the world of higher education. Since the emergence of global rankings, universities have been unable to avoid national and international comparisons, and this has caused changes in the way universities function.

2. Rankings and particularly the global league tables have adopted methodologies which address the world's top research universities only. *De facto,* the methodologies give stable results for only 700-1000 universities, which is only a small portion of the approximately 17,000 universities in the world. The majority of the world's universities are left out of the equation. While such an approach may well serve the purpose of producing a list of top universities, the problem is that the flurry of activity surrounding these rankings, often initiated by the ranking providers themselves, affects the whole higher education community as it tends to result in all higher education institutions being judged according to criteria that are appropriate for the top research universities only.

3. Rankings so far cover only some university missions. Few rankings address the broad diversity of types and various missions of higher education institutions.

4. Rankings, it is claimed, make universities more 'transparent'. However, the methodologies of the existing rankings, and especially those of the most popular league tables, still lack transparency themselves. It is difficult, if not impossible, to follow the calculations made from raw data to indicator values and, from there, to the overall score, just by using publicly available information.

5. "There is no such thing as an objective indicator" (see AUBR, 2010). The lack of suitable indicators is most apparent when measuring university teaching performance, for which there are no suitable proxies. The situation is better when evaluating research performance. However, even the bibliometric indicators used to measure research performance have their biases and flaws. Ranking providers are making some effort to improve their methodologies, but the improvements usually concern 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. Proxies can be improved, but they are still proxies.

6. At present, it would be difficult to argue that the benefits offered by the information that rankings provide, as well as the increased 'transparency,' are greater than the negative effects of the so-called 'unwanted consequences' of rankings.

7. New attempts to develop classifications, rankings and ratings targeting all higher education institutions and their various missions, such as the AUBR EU Assessment of University-Based Research, U-Map, U-Multirank and AHELO, all aim to improve the situation. However, it is too early to tell how these new tools will work; they are still at various stages of development or pilot implementation, and all of them still face 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.

V. GUIDANCE TO INTERPRETING RANKING RESULTS

General guidance

1. Look at the purpose and target groups of the ranking. For some rankings, this information helps to understand the selection of indicators and other features of the ranking; for others, it might reveal that the ranking does not actually follow the stated aims.

2. Find out what is actually being measured and compare it with what is said in the descriptions of indicators or groups thereof. It will clarify whether it is the aspect itself that is being measured, or a distant proxy.

3. Check whether you can fully understand how the result of each individual indicator is calculated from the raw data. It will help you to understand what the resulting number actually means.

   - Indicators have different dimensions and therefore, to put them into a final single score, the results have to be recalculated so that they become dimensionless. To do that, the result of the university in question is often divided by the result of the university that has the best result in this indicator and then multiplied by 100. So the number that appears in the table is usually not the number of publications, citations, students, etc. (as announced in the heading), but rather the percentage of the best result.

   - Check whether the results are normalised against some aspect. Results of research indicators are field-normalised in some (but not all) rankings; that would mean that the actual results are divided by, for instance, the average number of publications per staff FTE in the particular field, by the number of citations per article in the particular field, or by some other element. It improves comparability, but it also means that the indicator score is not the actual measurement but rather a composite value involving multiplication or division by some factors – which someone has decided on. To be sure, one might like to know the values of those factors, but such details are not usually shown in the simplified methodology descriptions that are published together with the league tables.

     Field-normalisation is not the only normalisation process used. The THE-Thomson Reuters 2010 ranking, for instance, applies *Accounting for international factors* (Pratt, 2010). However, details of these regional factors are not available on the public website.

4. Pay attention to the weights of indicators and indicator groups – sometimes they differ widely. For instance, in the 2010 THE-Thomson Reuters Ranking, the weight of citations is 32.5%, while the weight of research income from industry is 2.5%. The citation indicator is thus 13 times greater than the impact of funding from industry.

5. Check how data are collected and what the data sources are. It will help you to understand the ranking better if you know the proportion of hard data that can be verified, of data that is based upon the opinions of peers or employers, of data based on the judgements of students or staff of the university, and of data from national statistics or self-reported data from universities.

What should be noted regarding each university ranking?

Academic Ranking of World Universities (ARWU) – Shanghai Ranking Consultancy

- ARWU is not a ranking for all universities. It only considers around 1200 of the world's top universities which qualify for the ranking.
- ARWU was fully fit for its original purpose, i.e. to compare Chinese universities with the top US research universities. When used as a global ranking, its results may have negative consequences. This is particularly the case when, as a comparison of the research capacity of the world's elite research universities, it is used to judge the overall performance of a range of other universities.
- The 21 *broad subject areas* covered by ARWU demonstrate the dominance of natural sciences, medicine and, to some extent, engineering, while humanities and interdisciplinary areas are deliberately left out.
- ARWU compares the total strength of the university: all but one indicator are based on absolute numbers (of Nobel prizes, highly cited researchers, numbers of articles, etc.). Although the last indicator is per capita, its impact is only 10% of the total score. In other words, the ARWU methodology favours large universities.
- It follows from the above that merging universities would improve the position in ARWU, while the splitting off of faculties, especially those of medicine, sciences or engineering would weaken an institution's position.
- Improving the quality of teaching or increasing regional involvement will not improve a university's scores in ARWU.
- The position in ARWU is not affected by success in the arts and humanities. It is to be hoped that universities will not decide to make savings in these areas and strengthen further the natural sciences and medicine, which have a strong effect on scores in ARWU.
- Publishing books does not affect the value of the publications indicator.
- Is ARWU useful for students to make their choices? For top-performing young people who would like to study natural sciences, medicine and engineering – yes. Whether it is also of use to talented young people who are looking for a university 'simply' to become a well-trained citizen and professional, is open to debate.
- ARWU favours universities that are very strong in the sciences, those in countries where English is spoken. English is generally the language of research, non-English language work is published and cited less. Also, as far as US universities are concerned, it appears that Americans tend to cite other Americans (Altbach, 2006).

THE-QS World University Ranking (2004-2009)

- THE Rankings address only the world's top universities.
- The areas covered by the THE methodology are their 'four pillars' of world-class universities, and this methodology generates just over 600 universities in total and about 300 universities in each subject area.
- The 'peer review' is limited to a mere opinion poll among advanced researchers, rather than involving an expert visit to universities, as is common in quality assurance reviews.
- Peers or employers choose top universities from preselected lists, from which many universities and whole countries have been left out.
- Half of the overall score comes from two surveys, i.e. peer and employer reviews, both of which have low response rates; other data, with the exception of the number of citations, are provided by the universities themselves.
- The only proxy used for teaching quality is the staff/student ratio.
- Regional weights are used to achieve balance between world regions.
- Data normalisation and Z-scores are used to calculate the final score.

THE-Thomson Reuters World University Ranking (2010)

- The Times Higher Education started its cooperation with Thomson Reuters in autumn 2009 after the publication of the last THE-QS Ranking.
- The THE-Thomson Reuters Ranking remains focused on the world's top universities. It specifically excludes several categories of universities, particularly graduate schools and those universities that have published fewer than 50 papers.
- The current methodology description does not make it possible to reconstruct the calculation of the scores.
- The THE 2010 Ranking refers to publications from 2004-2008, excluding those of 2009.
- Bibliometric indicators have the greatest share of the overall weight (37%).
- The results of the citations indicator are normalised; citations for each paper are compared with the average number of citations received by all papers published in the same field and year.
- Reputation surveys on research and teaching still have a huge impact; their combined weight is more than one third of the overall score (34.5%).
- Indicators in this ranking are relative (per staff, per publication, etc.). Therefore, the ranking score is not size-dependent.

World's Best Universities Ranking – US News and World Report in cooperation with QS

- Up until now, the methodology is the same as the 2009 THE-QS methodology.

Global Universities Ranking – Reitor

- The aim of the ranking is to determine the position of Russian universities, and to develop a methodology that is more suited to the specific features of Russian higher education.
- The universities that have been invited to participate are those which rank highly in ARWU, THE and HEEACT, in addition to those universities which are located within the borders of the former Soviet Union. However, any university can participate.
- The ranking has only been compiled once, in 2008. Intentions to make it a regular global ranking have been announced, but no activities have been detected so far.
- The number of Russian universities in the Top 500 in the Reitor Ranking is 69 compared to seven in the ARWU 2009 ranking and four in the THE-QS 2009 ranking.
- The organisation of the ranking lacks transparency:
  - The 'Methodology' and 'About the ranking' sections of the Reitor website contain contradictory information.
  - A number of important decisions have been left to the discretion of Reitor's expert pool, and those decisions are not publicly available.
  - The weights of individual indicators have not been published, only the overall weights for 'blocks' could be found.
  - Details on the calculation of indicator scores are not provided.
- The original indicator system described in the 'Methodology' section of the Reitor website has not in fact been used. Instead, a different set of indicators has been used, which includes more bibliographic indicators and indicators for which data can be gathered from university websites or national higher education statistics.

Leiden Ranking – Leiden University

- The Leiden Ranking measures research output only.
- Only bibliometric indicators are used.
- The Leiden Ranking has no overall score; universities are ranked separately by each indicator.
- The indicators demonstrate the total publication production (P) and citations per publication – as such (CPP indicator), field-normalised (CPP/FCSm indicator) and mean-normalised (MCSN2 indicator). An indicator that reflects the total power of a university in the world is also used (P*CPP/FCSm).
- Both 'crown indicators' fail to cover the most recent publications, CPP/FCSM because more recent articles have accumulated fewer citations than older ones, and MCSN2 because the past year's publications are left out deliberately.
- The Leiden Ranking does not take into account the diversity or variety of university missions.
- To enter the ranking, a university needs to have a high number of publications that are covered by the citation indexes.
- Two of the indicators (P) and the 'brute force indicator' (P*CPP/FCSm) are size-dependent, i.e. they favour large universities.

Performance Rankings of Scientific Papers for World Universities (HEEACT) – Taiwan Higher Education Accreditation and Evaluation Council

- HEEACT ranks universities exclusively by bibliometric indicators, concentrating on research productivity, impact and excellence. Thus, the HEEACT ranking serves its purpose to rank universities purely by research performance.
- HEEACT places emphasis on current research performance and this distinguishes it from THE or ARWU.
- HEEACT does not rank all universities in the world; it considers only around 700 top universities for its overall university ranking and around 500 top universities for each subject field.
- Unlike the Leiden Ranking, which considers both universities and other research institutions, HEEACT looks at university research only. For field rankings, the university in question must have undergraduate programmes in the respective field, in order to be included in HEEACT.
- HEEACT uses the SCI and SSCI citation indexes, but not the humanities citation index, thus excluding humanities.
- Like ARWU, HEEACT covers the ESI 21 *broad subject areas*, which once again results in the predominance of the natural sciences, medicine and, to some extent, engineering.
- HEEACT attempts to compensate for the size of a university (unlike, for instance, ARWU or Leiden): 50% of the indicators are calculated per staff FTE.
- In accordance with its stated purpose, HEEACT disregards teaching and learning and leaves aside other university missions and the diversity of higher education institutions.
- HEEACT attempts to avoid cases where several universities form part of an overall university system (e.g. university systems in US states). These are considered as one university. Similarly subordinated units of a university are taken into account in measuring that university's performance whenever appropriate.
- As positioning in HEEACT is not affected by success in the arts and humanities, there is some danger that universities, in order to improve their position in this ranking, may re-allocate resources to strengthen those fields that affect HEEACT scores.
- Publishing books does not affect the value of most indicators.
- Is HEEACT useful for students to make their choices? For young people who intend to become researchers in the natural sciences, medicine, engineering and, partly, social sciences – to some extent, yes. For others – no.

Assessment of University-Based Research–European Commission

- AUBR is not a university ranking; it is a methodology for the assessment of university-based research.
- The AUBR methodology envisages a multi-indicator approach without calculating an overall score and therefore does not seek to produce a league table.
- The AUBR Working Group has analysed the strengths and weaknesses of various research indicators.
- Suitable combinations of indicators depending on the purpose of the assessment are offered in the Multidimensional Research Assessment Matrix.
- The conclusions of the AUBR Working Group on various indicators are useful when analysing global university rankings.

CHE University Ranking

- CHE is a multi-indicator ranking of universities whose main purpose is to help students find an appropriate higher education institution, but also to provide information helpful to higher education institutions themselves.
- No final score for a university is calculated. Instead, universities are placed in a top, middle or bottom group according to achievements in particular aspects.
- CHE covers universities in German-speaking countries as well as those that teach in German.
- When searching for an appropriate HEI, an individual can choose between various indicators in the groups listed, producing a personalised ranking which retrieves the universities that best fit the defined requirements.
- Indicators cover teaching and research, but also issues such as university buildings, sports, campus size, etc.
- The selection list of indicators presented to the user is different for universities and universities of applied sciences *(Fachhochschulen)*.
- Most indicators currently used in the CHE University Ranking are based upon students' assessments of various aspects of universities, and, substantially less on assessments made by professors or based on certain 'facts', such as statistics or bibliometric data. There is a shift toward assessments by graduates and towards a greater use of statistical and bibliometric data.

U-Map classification – Centre for Higher Education Policy Studies (CHEPS) of the University of Twente, the Netherlands

- U-Map is a multi-indicator classification tool. No final score is calculated. With regard to all but three indicators, each aspect is categorised as *major, substantial, some* and *none*.
- U-Map has been developed to classify all European HEIs regardless of the institution type, focus, etc.
- As a classification rather than a ranking, U-Map uses indicators that characterise the focus and intensity of various aspects of the work of HEIs, rather than performance, impact or quality.
- U-Map has two visualisation tools: *Profile finder* for finding higher education institutions which fit the characteristics set by the user; and *Profile viewer*, which enables a more thorough comparison of up to three selected HEIs.
- Indicators cover teaching level and subject focus, student body, research intensity, knowledge exchange, international orientation and regional involvement.
- Lack of internationally comparable data is a challenge for U-Map. The European Commission and Eurostat have launched new initiatives aimed at collecting comparable data across Europe. However, until these plans are fully implemented, U-Map will have to rely on national and institutional data, and is thus better suited for comparing institutions on a national, rather than an international basis.
- Another precondition for the success of such a European classification is the availability of Europe-wide agreed definitions of terms such as academic staff, teaching staff, research staff and peer-reviewed publications (especially with regard to books and monographs).
- U-Map is a new tool, finalised at the beginning of 2010. It is still being tested and data are being pre-filled by HEIs from volunteer countries.
- It is evident that the developers have made every effort to prevent other parties from constructing a league table from U-Map indicator data. Indeed, it would require an enormous amount of work to gather all the information needed to produce a league table from U-Map information. However, it is not absolutely impossible.

European Multidimensional University Ranking System (U-Multirank) – EU funded project

- U-Multirank is a multidimensional ranking that is currently being developed with the support of the EU. It follows the development of U-Map, the European classification of HEIs.

- U-Multirank will cover all aspects of HEIs' work including, education, research, knowledge exchange and regional involvement.

- Two kinds of rankings are being developed:

    1. *Focused institutional rankings* that allow for comparisons of institutions along a single *dimension* of institutional activity, such as education, research, regional involvement, etc.

    2. *Field-based rankings* that allow for comparisons of study programmes in the same scientific/study field in a group of institutions with a similar profile.

- No final score will be calculated. However, it unclear so far how the results can be protected in such a way as to prevent other parties from turning the ranking results into a league table.

- The ranking will be multi-indicator. Some indicators will be used in both institutional and field-based ranking, others in one only.

- The indicators cover teaching level and subject focus, student body, research intensity, knowledge exchange, international orientation and regional involvement.

- Despite the criticisms and disadvantages outlined in its interim report, the main data sources for U-Multirank will be Thomson Reuters and Scopus for bibliographic data and self-exported data by HEIs on students, teachers and research (except publications and citations). A large variety of data will be taken from student satisfaction surveys. Teacher surveys will not be used.

- The ranking results are aimed at students, academics and various stakeholders in society.

- More analysis will be provided once work on U-Multirank is completed.

Webometrics Ranking of World Universities

- The original purpose of the Webometrics Ranking was to encourage the academic community to increase online publication. Webometrics aims to promote Open Access initiatives as well as electronic access to scientific publications and other academic material. The Webometrics Ranking considers all higher education institutions that have their own independent domain, while most global rankings only concentrate on research universities belonging to the scientific elite.

- Webometrics measures two main aspects: the *Size of the university web*, represented by three indicators, which include the number of pages on the web; the number of publications and the number of file types found that are considered as relevant for academic purposes (.pdf, .ppt, .doc and .ps); and the *Visibility* of the university on the web, represented by one indicator, which is the number of external inward links (i.e. links that others have established to the university's website).

- The *Size* indicator is used as a proxy for the intensity of academic production in the university, albeit in a particular way, by combining research publications, presentations, teaching materials, raw data, drafts and other documents relevant for research, as well as teaching and administrative documentation from the university.

- Establishing a link to the university website can be compared to a citation – the link is established to refer to some information on the website and to allow others to quickly retrieve the original information. The number of inward links to the university website is thus used as a proxy for citations.

- Although the proxies seem quite distant, Webometrics results correlate rather strongly with those of the other global rankings.

- 20,000 HEIs are under analysis and the Webometrics league table lists 12,000 universities.

- Webometrics uses data from four search engines: Google, Yahoo, Live Search and Exalead.

- The Webometrics ranking providers have made public all the possible biases relevant for their analysis (e.g. geographical and language biases caused by search engines).

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

- OECD's AHELO project is an attempt to compare HEIs internationally on the basis of actual learning outcomes.
- The current AHELO project was launched to research the possibilities of measuring and comparing the actual learning outcomes of higher education on an international basis.
- Its more far-reaching goals are not only to measure learning outcomes as such, but to provide information on effective teaching strategies to ensure that learning outcomes are achieved. The intention is that this should allow more students to complete their degrees successfully and foster equity in different HE systems. AHELO would also give governments a tool to monitor the efficiency of their HEIs.
- The assessment tool for general skills is based on the US Collegiate Learning Assessment, which requires students to use an integrated set of skills, including critical thinking, analytical reasoning, problem-solving and written communication.
- As regards the development of assessment tools for discipline-specific strands, the approach of the Tuning project is being used as a basis upon which to define the set of the expected/desired learning outcomes.
- Each of the project strands will be tested in several countries located in various parts of the world. Students of 10 HEIs in each participating country will be tested shortly before they complete their first (bachelor) degree programme.
- The project team is aware of the potential difficulties involved in the international use of assessment tools, and is therefore also gathering background information that might be of help in international comparisons.
- More analysis will be provided once the current AHELO project is completed.

REFERENCES

Altbach. (2006). The Dilemmas of Ranking. *International Higher Education,* No. 42, winter 2006.

Annapolis group (2007). Annapolis group statement on rankings and ratings, 19 June 2007. Retrieved on 12 Jan. 2011 from: http://www.collegenews.org/x7131.xml

Aquillo, I., Ortega, J., Fernandez, M. (2008). Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments. *Higher Education in Europe*, Vol. 33, No. 2/3, July–October 2008, pp. 233-244.

Aquillio, I. (2009). Measuring the institution's footprint in the web. *Library Hi Tech*, Vol. 27 No. 4, pp. 540-556.

Aguillo I., Bar-Ilan J., Levene, M., Ortega J. (2010). Comparing university rankings. *Scientometrics*, 85, pp. 243-256.

ARWU. (2009). Methodology. Retrieved on 12 Jan. 2011 from: http://www.arwu.org/ARWUMethodology2009.jsp

Assessment of University-Based Research Expert Group (AUBR). (2010). *Assessing Europe's University-Based Research*, K1-NA-24187-EN-N, European Commission, Brussels, pp. 151. Retrieved on 12 Jan. 2011 from: http://ec.europa.eu/research/era/docs/en/areas-of-actions-universities-assessing-europe-university-based-research-2010-en.pdf

Baty, P. (2009). New data partner for World University Rankings. *Times Higher Education*, 30 Oct. 2009. Retrieved on 1 Nov. 2010 from: http://www.timeshighereducation.co.uk/story.asp?storycode=408881

Baty, P. (2010a). THE University rankings, 25 March 2010. http://www.timeshighereducation.co.uk/story.asp?storycode=410974

Baty, P. (2010b). THE World University Rankings 27 May 2010. Retrieved on 11 Jan. 2011 from: http://www.timeshighereducation.co.uk/story.asp?storyCode=411739&sectioncode=26

Baty, P. (2010c). THE unveils broad, rigorous new rankings methodology, *Times Higher Education*, 3 June 2010. Retrieved on 12 Jan. 2011 from: http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=411907&c=1

Baty, P. (2010d). We have had to make some extraordinarily difficult decisions this week - on which institutions to exclude from our 2010 World University Rankings. THE World University Rankings 5 August 2010. Retrieved on 12 Jan. 2011 from: http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=412832&c=1

Baty, P. (2010e). Robust, transparent and sophisticated. Times Higher Education, 16 Sep 2020. Retrieved on 30 Dec 2010 from: http://www.timeshighereducation.co.uk/world-university-rankings/2010-2011/analysis-methodology.html

Berghoff, S., Federkeil, G., Giebisch, P., Hachmeister, C., Hennings, M., Roessler, I., and Ziegele, F. (2009). CHE Hochchulranking Vorgehensweise und Indikatoren, CHE Arbeitspapier 19 May 2009, pp. 90. Retrieved on 12 Jan. 2011, from: http://www.che.de/downloads/CHE_AP119_Methode_Hochschulranking_2009.pdf

Berghoff, S., Brandenburg, U., Carr-Boulay, D., Hachmeister, C., Hennings, M., Leichsenring, H., and Ziegele, F. (2010). Identifying the bast: The CHEE Excellenc Ranking 2010. *Centre for Higher Education*, Working paper, No. 137, pp. 49. Retrieved on 25 Feb 2011 from: http://www.che.de/downloads/CHE_AP137_ExcellenceRanking_2010.pdf

Boulton, G. (2010). University rankings: diversity, excellence and the European initiative. Advice paper, No. 3, *League of European Research Universities* (LERU). Retrieved 31 Dec. 2010 from: http://www.leru.org/files/publications/LERU_AP3_2010_Ranking.pdf

Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M., Vadillo-Muño, O. (2007). Comparative study of international academic rankings of universities, *Scientometrics*, Vol. 71, No. 3, pp. 349-365.

Cattau, D. (2010). 'Fight the college rankings madness', *blog Opinion*, Dallas News, 27 Aug 2010. Retrieved 30 Oct. 2010 from: http://www.dallasnews.com/sharedcontent/dws/dn/opinion/viewpoints/stories/DN-cattau_29edi.State.Edition1.13c4754.html

CHE (2010). Methodology. *Die Zei*t website retrieved Aug 17 2010 from: http://ranking.Zeit.de/che2010/en/ueberdasranking/methodik

Cheng, Y. (2010). 'To better Measure social sciences performance. A review of existing ranking indicators'. Presentation at *IREG 2010 conference*, Berlin, 6-8 Oct. 2010. Retrieved 30 Oct. 2010 from: http://www.ireg-observatory.org/index.php?option=com_content&task=view&id=92&Itemid=86

CHERPA. (2010). *U-Multirank: Design Phase of the Project 'Design and Testing the Feasibility of a Multi-dimensional Global University Ranking'*, CHERPA Network, pp. 217. Retrieved 12 Jan. 2011 from: http://www.u-multirank.eu/UMR_IR_0110.pdf

Doneckaja, S. (2009). Российский подход к ранжированию ведущих университетов мира. ЭКО, No. 8, pp. 137-150. Retrieved on 12 Aug. 2010 from: http://www.sibupk.su/stat/publicat_nir/smk_4.pdf

EU Commission. (2009). Commission launches feasibility study to develop a multi-dimensional university ranking, *EU commission press release IP/09/859*, Brussels, 2 June 2009. Retrieved 12 Jan. 2011 from: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/08/1942&format=HTML&aged=0&language=EN&guiLanguage=en

Garfield, E., (1994). The Thomson Reuters impact factor, published in ISI Current Contents print editions on 20 Jun 1994. Retrieved 11 Jan. 2011 from: http://thomsonreuters.com/products_services/science/free/essays/impact_factor/

Federkeil, G., Ying, C. (2010). IREG Ranking audit - purpose, criteria and procedure. Presentation at IREG-5 Conference, Berlin, 7 October 2010.

Florian, R.V. (2007). Irreproducibility of the results of the Shanghai academic ranking of world universities. Scientometrics, Vol. 72, No. 1, pp. 25-32.

Hare, J. Coming up, body to rank rankers. The Australian, 12 Jan 2011. Retrieved 12 Jan 2011 from: http://www.theaustralian.com.au/higher-education/coming-up-body-to-rank-rankers/story-e6frgcjx-1225985832449

Hazelkorn, E. (2011). Rankings and the Reshaping of Higher Education. The Battle for World-Class Excellence. Plagrave.

Hicks, D., Tomizawa, H., Saitoh, Y. and Kobayashi, S. (2004). Bibliometric techniques in the evaluation of federally funded research in the United States. Research evaluation, 13, pp. 78-96.

Heylighen, F. (1993): "Red Queen Principle", in: F. Heylighen, C. Joslyn and V. Turchin (editors): *Principia Cybernetica Web* (Principia Cybernetica, Brussels).

Huang, M. (2010). Performance Ranking of Scientific Papers for World Universities. Retrieved 12 Jan 2011 from: http://ranking.heeact.edu.tw/en-us/2010/Page/Background

IREG. (2006). Berlin Principles on Ranking of Higher Education Institutions adopted at IREG's 2nd meeting in Berlin, 18-20 May, 2006.

IREG. (2010). IREG-Ranking Audit: Purpose, Criteria and Procedure, Draft version November 2010. Retrieved 10 Jan. 2011 from: http://www.ireg-observatory.org/index.php?option=com_content&task=view&id=110&Itemid=101

King, R., Locke, W., et al. (2008). Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England. Hefce issues paper, 14 April 2008. Retrieved from: http://www.hefce.ac.uk/pubs/hefce/2008/08_14/

Kruzhalinin, V., and Artjushina, I. (2008). Rankings and quality of education. Retrieved 5 Aug. 2010 from: http://reitor.ru/ru/observatory/global/index.php?id19=474

Lalancette D. (2010). Assessment of Higher Education Learning Outcomes:a ground-breaking initiative to assess quality in higher education on an international scale. Presentation at IREG-5 Conference, Berlin, 7 Oct. 2010. Retrieved 17 Dec. 2010 from: http://www.ireg-observatory.org/prezentacje/2010/S2_2_Lalancette.pdf

Leiden ranking. (2008). http://www.cwts.nl/ranking/LeidenRankingWebSite.html

Leydesdorff, L., and Opthof, T. (2010). Normalisation, CWTS indicators, and the Leiden Rankings: Differences in citation behavior at the level of fields. Retrieved on 17 Dec. 2010 from: http://www.leydesdorff.net/reply2cwts/reply2cwts.pdf

Liu, N.C., and Cheng, Y. (2005). The Academic Ranking of World Universities. Higher Education in Europe, Vol. 30, No. 2, pp. 127-135.

Liu, N. C. (2009). The Story of Academic Rankings. International Higher Education, No. 54, 2-3. winter 2009.

Marginson, S. (2007). Rankings: Marketing Mana or Menace? The Big Picture. 16th Annual New Zealand International Education Conference, 8-10 August 2007, Christchurch, New Zealand. Retrieved 11 Jan 2011 from: http://www.cshe.unimelb.edu.au/people/staff_pages/Marginson/KeyMar8-10Aug07.pdf

Marginson, S., and Wan der Wende, M. (2007). To Rank or To Be Ranked: The Impact of Global Rankings in Higher Education. Journal of Studies in International Education, Vol. 11, No. 3/4, Fall/Winter 2007, pp. 306-329. Retrieved 10 Jan. 2011 from: http://www.cshe.unimelb.edu.au/people/staff_pages/Marginson/Marginson&VanDeWendeJSIEFallWinter07.pdf

Marmolejo, F. (2010). The madness of rankings. Chronicle of Higher education, 22 Sept. 2010. Retrieved 12 Jan. 2011 from: http://chronicle.com/blogPost/The-Madness-of-Rankings-/27100/

Moskovkin, V. (2009). From shifting places… What is the use of global ranking for Russian universities? Poisk, newspaper of scientific society, 28 April 2009. Retrieved from: http://www.poisknews.ru/2009/04/28/ot_peremeny_mest.html

Moed, H. (2005). Citation Analysis in Research Evaluation. Dordrecht, Springer, pp. 353.

OECD. (2008). Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools. ISBN: 9789264050228.

OECD. (2010). AHELO (assessment of higher education learning outcomes). Project update - December 2010. Retrieved 17 Dec. 2010 from: http://www.oecd.org/dataoecd/8/26/45919053.pdf

Olds, K. (2010). Governing world university rankers: an agenda for much needed reform, 21 Sept. 2010 at globalhighered. Retrieved from: http://globalhighered.wordpress.com/2010/09/21/governing-world-university-rankers-an-agenda-for-much-needed-reform/

Pratt, S. (2010). Times Higher Education World University Rankings Global Institutional Profiles Project. Presentation at IREG-5 Conference, Berlin, 7 October 2010. Retrieved 12 Jan. 2011 from: www.ireg-observatory.org/prezentacje/2010/S4_1_Pratt.pdf

Rauhvargers, et al. (2009). Bologna Stocktaking report 2009, pp. 25, Retrieved 28 Dec. 2010 from: http://www.ond.vlaanderen.be/hogeronderwijs/bologna/conference/documents/Stocktaking_report_2009_FINAL.pdf

Rehn, C., Kronman, U., and Wadskog, D. (2007). Bibliometric indicators – definitions and usage at Karolinska Institutet. http://ki.se/content/1/c6/01/77/43/Bibliometric%20indicators%20-%20definitions_1.0.pdf

U-Multirank. (2010). Interim Progress Report, CHERPA-Network, January 2010, internet: www.u-multirank.eu/UMR_IR_0110.pdf

Reitor. (2008a). Methodology. Retrieved 31 Dec. 2010 from: http://www.globaluniversitiesranking.org/index.php?option=com_content&view=article&id=68&Itemid=128

Reitor. (2008b). Questionnaire. Retrieved 31 Dec. 2010 from: http://www.globaluniversitiesranking.org/images/banners/Questionaire_1.doc

Reitor. (2009a). About the ranking. Retrieved on 9 Aug. 2010 from: http://www.globaluniversitiesranking.org/index.php?view=article&id=96%3A2009-02-11-13-58-17&tmp

Reitor. (2009b). From the ranking compilers (list of universities that actually answered). Retrieved 31 Dec. 2010 from: http://www.globaluniversitiesranking.org/index.php?option=com_content&view=article&id=95&Itemid=1

Reitor. (2008c). Presentation at Moscow University. Retrieved on 6 Aug. 2010 from: http://www.reitor.ru/common/img/uploaded/files/msu_present_19-20.11.pdf

Sadlak, J. and Liu, N.C., Eds. (2007). The world-class university and ranking: Aiming beyond status. Bucharest UNESCO-CEPES, pp. 378. ISBN: 978-92-9069184-0.

Saisana, M., and D'Hombres, B. (2008). Higher Education Rankings: Robustness Issues and Critical Assessment. How much confidence can we have in Higher Education Rankings? Luxembourg, Office for Official Publications of the European Communities. Retrieved 12 Jan. 2011 from: http://globalhighered.files.wordpress.com/2008/11/eur23487.pdf

Science and technology in the Russian Federation. (2008). Мировой рейтинг вузов: Российская версия, 11 Apr. 2008. Retrieved 12 Aug. 2010 from: http://strf.ru/organisation.aspx?CatalogId=221&d_no=14080

Salmi, J., and Saroyan, A. (2007). League tables as policy instruments: Uses and misuses. Higher. Education Management and Policy, Vol. 19, No. 2, pp. 31-68.

Salmi, J. (2009). The challenge of World class universities, World bank, Washington, pp. 136. Retrieved 12 Jan. 2011 from: http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079956815/547670-1237305262556/WCU.pdf

Salmi, J. (2010). If ranking is the disease, is benchmarking the cure? Presentation at IREG 2010 conference, Berlin, 7-8 October 2010.

Samarasekera, I. (2007). Rising Up Against Rankings. Inside Higher Education, April 2. Retrieved 12 Jan 2011 from: http://www.insidehighered.com/views/2007/04/02/samarasekera

Stolz, I., Hendel, D., and Horn, A. (2010). Ranking of rankings: benchmarking twenty-five higher education ranking systems in Europe. Higher Education, Online, pp. 22. ISSN 0018-1560. Springer, Netherlands.

Sursock, A., Smidt, H., et al. (2010). Trends 2010: A decade of change in European Higher Education. European University Association, Brussels, pp. 123.

THE (2008). World University Rankings 2008. Page retrieved 29 Dec. 2010 from: http://www.timeshighereducation.co.uk/hybrid.asp?typeCode=416&pubCode=1&navcode=137

THE (2008b). World University Rankings 2008: Strong measures. Retrieved 29 Dec. 2010 from: http://www.timeshighereducation.co.uk/story.asp?storycode=407247

THE (2009). World University Rankings 2009. Retrieved 29 Dec 2010 from: http://www.timeshighereducation.co.uk/hybrid.asp?typeCode=431&pubCode=1&navcode=148

THE-QS (2009). Methodology: purpose and approach. Retrieved 28 Dec. 2010 from: http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/purpose-approach

THE (2009b). Rankings 09: Talking points. Retrieved 29 Dec. 2010 from: http://www.timeshighereducation.co.uk/story.asp?storycode=408562

Tuning (2004). Tuning project methodology. Tuning Project, 2004. Retrieved 8 Jan. 2011 from: http://tuning.unideusto.org/tuningeu/index.php?option=content&task=view&id=172&Itemid=205

U-Map. (2009a). Overview of indicators and data-elements, by dimension. Retrieved 22 Sept. 2010 from: http://www.u-map.eu/U-Map%20dimensions%20and%20indicators%20overview.pdf

U-Map. (2009b). Overview of dimensions and indicators. Retrieved 12 Jan. 2011 from: http://www.u-map.eu/U-Map%20dimensions%20and%20indicators%20overview.pdf

van Vught, F., Kaiser, F., File, J., Gaethgens, C., Peter, R., and Westerheijden, D. (2010). The European Classification of Higher Education Institutions, CHEPS, Ensschede, pp. 46. Retrieved 22 Sep 2010 from: http://www.u-map.org/U-MAP_report.pdf

van Vught, F., Bartelse, J., Huisman, J., and van der Wende, M. (2005). Institutional profiles, Towards a typology of higher education institutions in Europe. CHEPS, Enschede. Retrieved 22 Sept. 2010 from: http://doc.utwente.nl/53776/1/engreport05institutionalprofiles.pdf

van Vught, F., Kaiser, F., Bohmert, D., File, J., and van der Wende, M. (2008). Mapping diversity; Developing a European classification of higher education institutions. Enschede, CHEPS.

van Vught, F., Ed. (2009). Mapping the Higher Education Landscape: Towards a European Classification of Higher Education. Dordrecht etc., Springer., pp. 170.

Vestrehijden, D. et al. (2010). The first decade of working on the European Higher Education Area. The Bologna Process Independent Assessment, Vol. 1, pp. 39. Retrieved 28 Dec. 2010 from: http://www.ond.vlaanderen.be/hogeronderwijs/bologna/2010_conference/documents/IndependentAssessment_1_DetailedRept.pdf

Usher, A., and Savino, M. (2006). A World of Difference: A Global Survey of University League. Tables. Toronto, ON: Educational Policy Institute. Retrieved 12 Jan. 2011 from: http://www.educationalpolicy.org/pdf/world-of-difference-200602162.pdf

van Raan, A. (2005). Fatal Attraction: Ranking of Universities by Bibliometric Methods Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics Vol. 62, No. 1, pp. 133-143. Retrieved 10 Jan. 2011 from: http://www.cwts.nl/TvR/documents/AvR-RankingScientom.pdf

van Raan, A. (2010). Presentation at shool of Scientometrics, Humboldt university 16 June. Retrieved 12 Jan 2011 from: http://www.scientometrics-school.eu/images/esss1_vanRaan.pdf

van Raan et al. (2010). Anthony F.J. van Raan, Thed N. van Leeuwen, and Martijn S. Visser Van. Germany and France are wronged in citation-based rankings. Retrieved 11 Jan. 2010 from: http://www.cwts.nl/pdf/LanguageRanking22122010.pdf

Waltman, L., van Eck, N., van Leeuwen, T., Visser, M., and van Raan, A. (2010). Towards a new crown indicator: An empirical analysis. Centre for Science and Technology Studies, Leiden University, The Netherlands. Retrieved 3 Jan. 2011 from: Centre for Science and Technology Studies, Leiden University, The Netherlands http://arxiv.org/PS_cache/arxiv/pdf/1004/1004.1632v2.pdf

van Noorden, R. (2010). A profusion of measures. Vol. 465, 17 June 2010, pp 864-866. Retrieved 12 Jan 2011 from: http://planetsci.uwo.ca/PS%20Seminar/PS%20Seminar/Papers/A_Profusion_of_Measures.pdf

The European University Association (EUA) is the representative organisation of universities and national rectors' conferences in 47 European countries. EUA plays a crucial role in the Bologna Process and in influencing EU policies on higher education, research and innovation. Thanks to its interaction with a range of other European and international organisations EUA ensures that the independent voice of European universities is heard wherever decisions are being taken that will impact on their activities.

The Association provides a unique expertise in higher education and research as well as a forum for exchange of ideas and good practice among universities. The results of EUA's work are made available to members and stakeholders through conferences, seminars, website and publications.

EUA
European University Association

European University Association asbl
Avenue de l'Yser 24 · 1040 Brussels, Belgium · Tel: +32-2 230 55 44 · Fax: +32-2 230 57 51 · www.eua.be