

32nd Annual Conference of the Association for the Study of Higher Education
7-10 November, Louisville, Kentucky

Symposium, 10 November:

'Comparing colleges: the implications of classification, ranking and peer analysis
for research and practice'

Global university rankings [presentation version]

Simon Marginson *

[1] Introduction: The last four years have seen the emergence of two systems of global university rankings, conducted by the Shanghai Jiao Tong Institute of Higher Education and the Times Higher Education Supplement; and one regional system of comparison managed by the Centre for Higher Education Development (CHE) in Germany. These data are all issued on an annual basis. In addition, in 2006 *Newsweek* issued a composite ranking that combined part of each of the Jiao Tong and Times indicators. **[2]** These rankings generate much media coverage, throughout most of the world (with one notable exception), and have begun to exert direct effects in the marketing and development strategies of universities, and in some nations the policies and priorities of government. There could be no more clear-cut indication of the growing role of the global dimension in higher education. The exception is the United States where it is taken for granted by most people that American universities are superior to those from foreign countries, and the only rankings that matter are the national tables issued each year by *US News*.

[3] Shanghai Jiao Tong rankings: More than a decade ago the Chinese government decided that to be number 1, China had to catch

* Simon Marginson is a Professor of Higher Education in the Centre for the Study of Higher Education at the University of Melbourne, Australia. He is a Fellow of the Academy of Social Sciences Australia and the Society for Research into Higher Education (UK), and board member of *Higher Education*, *Higher Education Policy*, *Higher Education Quarterly*, *Journal of Education and Work* and *Thesis Eleven*.

e-mail address: s.marginson@unimelb.edu.au

fax: 613-83447576

telephone: 613-83448060

postal address: Centre for the Study of Higher Education, University of Melbourne, Victoria 3010 Australia

up to the US as a 'knowledge economy'. This meant not just commercial science, the course followed by Margaret Thatcher in the UK a decade earlier (Mode 2 science as Michael Gibbons puts it), but Mode 1, basic research. So China set out to measure the research performance of all research universities world-wide, so as to benchmark the performance of Chinese universities; to work out how much China needed to improve, and how much it would need to invest, and where. The Shanghai Jiao Tong University Institute of Higher Education was funded to prepare world rankings. The first rankings came out just four years ago, in 2003.

[4] The Jiao Tong group argues that the only data sufficiently reliable for ranking purposes are broadly available and internationally comparable data of measurable research performance. It is considered impossible to compare teaching and learning 'owing to the huge differences between universities and the large variety of countries, and because of the technical difficulties inherent in obtaining internationally comparable data'. Further, the Jiao Tong group states that it does not want to employ subjective measures of opinion or data sourced from universities themselves as are used in some national rankings. The SJUIHE group is involved in a continuing reflexive interchange with the users of the data. It annually tunes its rankings and invites open collaboration in that. This is a strength of the academically rigorous Jiao Tong approach.

[5] Being China, there is much respect for scholar-savants. So Nobel Prize winners make up 30 per cent of the Jiao Tong index. This is questionable, because the prize is submission based and open to politicking. **[6]** Most of the rest, 60 per cent, is comprised by academic publication and citation indicators. **[7]** The citation and publication data are from Thomson-ISI, one of two firms specializing in such information. There is no correction for different disciplinary patterns of publication and citation – books are not included, just articles. Medical researchers cite more than engineers so Jiao Tong favours universities with big medical faculties, though there is some correction for size: the last 10 per cent is a measure of the performance on the other 90 per cent divided by the number of academic staff. Other than that, within its frame, the collection is rigorous and has been improved each year. Shanghai Jiao Tong also

publishes discipline rankings, in five fields: engineering, physical sciences, life sciences, medical sciences and social sciences.

[8] American universities dominate the Jiao Tong listing, with 17 of the top 20 and 54 of the top 100. The English speaking countries together have 71 per cent of the top 100: USA 54, UK 11, Canada 4 and Australia 2. Jiao Tong is biased towards high science in general and as noted medicine in particular, so favouring English language countries. English is the sole global language of research. It also helps to have a large number of Thomson-ISI 'HiCi' researchers because they dominate much of the index. In late September this year there were 3835 HiCi researchers located in the USA, almost nine times the number in any other country. The UK had 456, Germany 256, Japan 253, Canada 182, Switzerland 112, Australia 107 altogether. In the U.S. Harvard had 160, Stanford 135 and Berkeley 82. There were 44 HiCi researchers at Cambridge UK. **[9]** However in future a greater plurality of research capacity will emerge. **[10]** China, Singapore and Korea have undertaken significant investments in basic research capacity and Chinese research universities are likely to become the second strongest system in the world **[11]**.

[12] Times Higher rankings: The Times rankings were born in 2004.

[13] Unlike the Jiao Tong rankings which focus solely on research, the Times Higher set out to produce a composite indicator which could be presented as a holistic measure of university standing, and as such relevant to the cross-border student market, much as *US News*. The Times criteria are qualities seen to shape global market position.

[14] Half of the *Times Higher* index is comprised by university reputation. As well as the 40 per cent comprised by a reputational survey of academics ('peer review'), another 10 per cent is determined by a reputational survey of 'global employers'. In addition there are two internationalisation indicators: the proportion of students who are international (5 per cent) and the proportion of staff (5 per cent). Another 20 per cent is determined by the student-staff ratio, a quantity measure used as a proxy for teaching 'quality'. The remaining 20 per cent is comprised by research citation performance.

The Times Higher market researchers seek part of the relevant data directly from institutions, and are subject to institutional lobbying.

[15] According to a conference paper delivered in February this year by Ben Sowter from QS Marketing, which does the market research for the Times, last year QS sent out 200,000 e-mail questionnaires to academic staff around the world. When the replies came back in, they totalled a bit more than 2000. After the invalid responses were weeded out, the response rate was 1 per cent. Remember that this survey comprises almost half the Times ranking. A further difficulty was that the pool of responses was heavily weighted in favour of academic 'peers' from nations where the *Times* is well-known, such as the UK, Australia, New Zealand and Malaysia. Despite this, the pool was not re-weighted for global evenness (which may not have been very practical given an overall response rate of 1 per cent).

The results of the Times ranking closely reflect the composition bias in the survey. The UK has 15 per cent of the GDP of the USA, but 50 per cent of the USA's number of universities in the Times top 100 in 2006. In 2006 the UK had two of the Times Higher top three and Cambridge UK almost closed the gap on Harvard. (Note however that the Harvard faculty are cited at three and a half times the rate of Cambridge UK faculty, according to the Jiao Tong). In 2007 the Times managed to place three UK universities in the top five and to move University College London up to 9th place. Though the USA had 54 research universities in the Jiao Tong top 100, the Times cut it down to 33 in 2006 though it rose again to 38 in 2007. Meanwhile the British 'achieved' an amazing 19 in the top 100 in 2007.

[16] Arguably, the Times Higher rankings also distort the global standing of the Australian universities. In the Times indicators the Australian universities have a relatively poor citation rate and moderate staffing ratios. But they perform very well in the student internationalization indicator, which rewards volume. They also perform well in the peer survey of academic staff. In 2006 the Australian National University and Melbourne, were both ranked by the 'academic peers' as equal to Yale and ahead of Princeton, Caltech, Chicago, Columbia, Cornell, Duke, Texas Austin, Johns Hopkins, UCLA, Michigan, Illinois, LSE, University College London,

Birmingham, Manchester, Ecole Normale Superieur and the rest of Western Europe.

[17] Given the vicissitudes of the survey, numerous changes in data and methodology, and the lobbying factor, it is not surprising that the Times Higher rankings have proven highly volatile. The volatility seems to be more extreme in 2007 than in previous years. The 2007 results were released two days ago. This year Stanford slipped to 19 in the world after being 6 last year. The National University of Mexico, UNAM, tumbled from 74 to 192. The National University of Singapore moved from 19 to 33. Think 'yo-yo' and you've about got it.

[18] CHE comparisons: Student-teacher ratios are not an adequate indicator of teaching quality and research performance has no necessary implications for quality of teaching. There are widespread desires to provide better data in relation to comparisons of teaching quality. In Europe there is cross-country support for the approach to comparison developed by the Centre for Higher Education Development (CHE) in Germany. This dispenses with holistic (summative) rank ordering of institutions in league tables. CHE surveys 130,000 students and 16,000 academic faculty in almost 250 higher education institutions, focusing on student experiences, student satisfaction, and academic recommendations on the best places in each field of study. It thus includes student judgments about teaching and extends into student judgments concerning services, and academic faculty judgments concerning research. It supplements the surveys with independent sources comprising one third of the total data base. No data are taken from institutions.

CHE ranks institutional departments according to each separate indicator of academic and service quality, assigning them to top third, middle third or bottom third of all institutions. It provides neither discipline-based league tables nor institutional league tables. It refuses to integrate the different indicators into a single indicator for each institution because there is no 'one best university' across all areas, and 'minimal differences produced by random fluctuations may be misinterpreted as real differences' in holistic rankings systems.

The CHE data are made available to prospective students and the public free of charge through an interactive web-enabled database.

Any person can interrogate this data base by investigating and rank identified disciplines and administrative services, using their own combination of criteria, thereby creating weightings and rankings themselves. CHE acknowledges that the definition of 'quality' is purpose-driven and open to variation, and passes power over the definition from the ranker to the consumer. A difficulty is that the CHE approach requires sufficient homogeneity between institutions and their programs and services to enable the comparisons. As European systems develop more inter-institutional diversity, and given also the existing diversity between nations, in any Europe-wide system of comparison it is probable that it would be necessary to develop several sets of institutions for CHE-style comparisons much as the United States has done.

A limitation of the CHE data on teaching and curricula is that they are subjective in character. The OECD is in the early stages of developing standardizable measures for measuring the academic outcomes of programs on a comparative international basis, possibly a combination of competency tests of generic skills and academic performance in a small number of key disciplines. It is possible that 'value-added' measures will be developed also. The OECD has stated that it does not intend to produce a ranking using these data but if such data are developed (and it is not yet certain that viable indicators are possible) it would be inevitable that they would be fed into rankings systems issued by other agents.

[19] Impact: The first four years of experience of global rankings testify to the potential impact of inter-institutional comparisons. There are signs that outside the United States that global rankings are beginning to exert significant system-shaping and institution shaping effects, particularly through their capacity to determine and alter institutional reputation. Rankings have normalized the idea of a worldwide market in higher education and exacerbated competitive pressures within and between nations.

More specifically, in some countries such as Germany and the Netherlands the Jiao Tong University research rankings have focused national government attention on actual or possible policies designed to increase the concentration of research activity in a small number of universities, including recruitment of additional high citation

researchers, a group which significantly impacts university performance in the Jiao Tong University rankings. In Europe global rankings are also associated with the formation of the European League of Research Universities and the development of a typology of European institutions. In East and SE Asia there are discussions concerning the possibility of both a regional typology and a regional ranking system. In some nations new national rankings systems are emerging. Market research suggests that foreign student choice-making regarding country and institution of study is affected by university rankings data.

It is likely also that as with *US News* in the United States, the global rankings systems are also affecting the flows of doctoral students, elite researchers and the philanthropic and corporate dollar.

[20] Problems of validity and effect: The rise and rise of rankings has led to a widespread interest in well grounded rankings that have positive effects on institutional performance. While the Berlin Principles established a basic code for rankings, in areas such as transparency of data compilation method, there are significant grey areas in relation to validity in combination with effect.

Three problems of validity have emerged. The first is the reliability and accuracy of the data, where the principle negative example is that of the *Times Higher*. A 1 per cent survey response rate, together with the sourcing of some data from interested parties, the institutions themselves, does not constitute a valid data compilation.

A second problem is the arbitrary character of the weightings used to construct composite indexes covering different aspects of quality or performance, the means by which ratings agencies construct a total picture of the institutions that are ranked against each other.

A third problem of validity is that reputational rankings tend to be both ill-grounded and circular: In reputation-based rankings known university brands generate 'halo' effects. The *Times Higher* favours universities already well known regardless of merit, tending to recycle existing reputations while blocking newcomer institutions or nations. There is no means of verifying the soundness of subjective judgements of reputation, for example ensuring that they are

grounded in actual comparative knowledge, or address the fundamentals such as the quality of teaching and research. One study of ranking found that one third of those who responded to the survey knew little about the institutions concerned apart from their own. The classical example of these problems is the American survey of students that found Princeton law school was ranked in the top ten law schools in the country. But Princeton did not have a Law school.

Likewise we can note three problems of use of the data, in combination with validity issues. First, rankings, especially reputational rankings, become an end in themselves and protected from critical scrutiny, 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. Often institutions are rank ordered even where differences in the data are not statistically significant. Moreover, the illusion is created that all institutions have the same capacity to succeed even though their circumstances are often vastly different. Consider for example the difference between a leading university in the USA and a leading university in Indonesia. Population sizes of the two countries are the same order of magnitude. But that's where equality of capacity stops. In 2001 the USA published one thousand times the number of scientific papers that Indonesia published.

A second problem of use of rankings data is that when the data are not solidly grounded, as in the case of the Times Higher ranking, changes in the rankings do not necessarily reflect changes in actual performance. There is no virtuous link between competition, performance and ranking. The worst case scenario is when institutions receive an undeserved 'hit' in the data, so that the effect of the rankings is capricious and destructive. The now famous example of the University of Malaysia in Malaya is a case in point. In 2004 the University of Malaya was ranked by the *Times Higher* at 89 in the world. This was seen as a very positive achievement within Malaysia; for example the University's Vice-Chancellor ordered large banners declaring 'UM a world's top 100 university' placed around the city, and on the edge of the campus facing the main freeway to the airport where every foreign visitor to Malaysia would see it. But the next year in 2005 an error in the classification of foreign students at the University was corrected and the outcomes of the *Times* two

reputational surveys changed; both changes were to the disadvantage of the University. The University dropped from 89 to 169 in the Times ranking, without any necessary change in its actual performance. The Vice-Chancellor was pilloried in the Malaysian media and when his position came up for renewal by the government in March 2006 he was not reappointed.

A third problem of use is that singular rankings systems encourage institutions to reduce the emphasis on those activities that do not contribute to rankings performance; and more generally, leads to convergence of behaviour between institutional types and between national systems (and languages of use). Unless there is a broad range of rankings systems with no one system dominant, all else being equal rankings tend to work against diversity of provision. This is a serious difficulty, much remarked upon, with no solution in sight.

[21] Conclusions: This is a fast changing area and there is potential for improvement in practices. The foregoing discussion suggests a number of conclusions:

- Rankings based on surveys of reputation as such should be avoided as there is no necessary links with fundamental capacity or performance, and reputational rankings generate circular reputation-forming effects;
- Rather than composite 'omnibus' rankings that in reality leave much uncovered and involve arbitrary decision about weightings, specialist rankings specific to purpose (such as rankings of research, rankings of student achievement, etc), grounded in data specific to the purpose, should be used, OR comprehensive data bases that can be broken down to specific questions such as the CHE data base;
- All else being equal, the more the number of rankings systems, and the more diverse the qualities included in them, the better. Diverse multiple rankings produce more information of use to more people, and undermine the potential of any one ranking to obtain supreme status thus becoming a de facto reputational ranking;
- Ideally rankings should be developed and maintained by independent agencies funded by foundations or governments, situated at arms length from the funders. The next best option is for rankings to be managed in university research centres

- providing that they are not contaminated by institutional interest and a completely separated from marketing departments;
- Rankings should not be run by newspaper companies because their purposes are unsuitable to the production of valid rankings. They do not have a vested interest in valid social science or the long term healthy development of higher education. **[22]**