University Rankings and Social Science[1]

Simon Marginson

Introduction: global university ranking

Since the first Shanghai Academic Ranking of World Universities (ARWU) in 2003 global rankings have transformed higher education. A 2011 report for the European Universities' Association on *Global University Rankings and their Impact* states:

> . . . 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 (Rauhvargers, 2011, p. 68).

Rankings affect the judgements and decisions of many university leaders and faculty; prospective students, especially international students, and their families; state policy makers and regulators; and industry and philanthropic investors (Hazelkorn, 2008; 2011). Many faculty express disdain for rankings, because they are based on the wrong criteria, or overly narrow the purposes of higher education, or exacerbate competition, and/or are methodologically flawed. Some university leaders claim to stand aloof. Yet this performance indicator exercises an almost hypnotic power. At a time when nations see themselves as 'global competition states' (Cerny, 1997) and global comparisons gather significance in many domains (OECD, 2013a; 2013b), university ranking fills a data gap. It renders the arcane and complex world of universities simple, transparent and compelling. Ranking seems democratizing: it renders accountable institutions that once held themselves above the common herd. It fits with old ideas about university status and modern contests like football league tables. It is easily understood and remembered.

University ranking also speaks to desire. Every research university wants to improve its rank. Many institutions with a primarily teaching mission feel the lack of rank. Faculty want to be associated with prestigious institutions. Students want to be selected by them. The desire to rise is universal, as Adam Smith notes in *The Wealth of Nations*:

> The desire of bettering our condition . . . comes with us from the womb, and never leaves us till we go into the grave (Smith, 1776/1979, p. 441).

Hence also, there is an unquenchable desire for data about relative social position (Hirsch, 1976) as Pierre Bourdieu (1986) notes. Because university rankings order the status of institutions they regulate the relative value of graduate credentials. They affect the social position of many people. They have become an integral part of status culture.

In short, global university ranking is here to stay. In some jurisdictions ranking also has prior national roots, for example the annual *US News and World Report*

[Corrections added on 7 March 2014, after first publication: The editorial office's corrections for this article were inadvertently omitted and have been incorporated in this version of the article.]

ranking in the US (USNWR, 2013). Largely managed by non-state organisations in publishing industry or within universities themselves, ranking has become a form of regulation as powerful in shaping practical university behaviours as the requirements of states. While some national cultures emphasise competition and position more than others, Hazelkorn finds that institutional rank has become a primary performance measure for research university rectors, vice-chancellors and presidents. It is the one universal performance indicator, more important than student numbers (only some universities want to grow) or revenues (money is a means to university status). All executive leaders are impelled to pursue policies designed to enhance their institution's rank on the basis of the performance indicators incorporated in ranking systems, even though few can win a game determined by research outputs or reputational surveys. Ranking reinforces the advantages enjoyed by leading universities. It celebrates their status and propels more money and talent towards them, helping them to stay on top. It is difficult for outsiders, emerging universities and countries to break in. Rankings are not 'fair' to competing universities. The starting positions are manifestly unequal. Yet university ranking insinuates itself in a growing number of places over time. It seems irresistible.

Rankings as Social Science

Yet how sound are these data that fix social position? The most unnerving aspect of global university rankings is not their power to normalise and exclude — many other social systems do that, including 'the economy' — but the shaky methodologies, the arbitrary definitions and scope for manipulation. University status starts to peel loose from its material foundations. Status becomes a circular game in which power makes itself. This highlights the importance of data quality and interpretative validity. If rankings are effectively grounded in real university activity there is potential for a virtuous constitutive relationship between university rank and university performance.

Data quality and rankings validity are also matters of common good. Desire for status is not the only driver of fascination with rankings. Comparative data about universities have many possible uses. Though the potential of comparisons is not exhausted by the present rankings systems, the data already collected include numbers and rates of science publications and related citations by discipline; student enrolments and staffing ratios; internationalisation of students and academic faculty; income levels in selected areas; and so on. Again, this suggests the need to tune global rankings (and other cross-border comparisons of higher education and research) to optimise the information generated.

But where is the constructive discussion about reforming the rankings? There is a large academic and popular literature on university rankings. It is mostly driven by normative considerations, or is descriptive, or provides practical guidance on how to maximise institutional position. There is also continuing (albeit futile) discussion about whether university rankings should exist. Much of the literature is transparently motivated by self or sectional interest; for example attempts to remove indicators that are unfavourable to one or another institution or country and replace them with 'suitable' measures. Academic papers discuss the implications for policy and regulation, and behaviours and systems, and global relations of power (Sauder & Espeland, 2009; Marginson, 2008). Many papers list and review the different rankings systems (Salmi & Saroyan, 2006; Usher & Savino, 2006; Cheng, 2010; Rauhvargers, 2011). Most extant comprehensive reviews are largely

descriptive, or examine rankings primarily in terms of normative policy assumptions rather than principles of social science.

University rankings are critiqued; but surprisingly, they are little critiqued as social science. It is surprising because the techniques used by university rankers are taken from research in sociology, economics, psychology, and business studies, including market research. There is little discussion and debate about what the different rankings measure and how, and related questions of data coverage and validity; the kind of discussion that distinguishes between good and bad rankings on scientific grounds. For example, the reliance on multi-indicators with arbitrary weightings — a method used by all well-known rankings systems but especially vulnerable to critique on grounds of validity — is scarcely mentioned. Why are established social scientists largely silent about university rankings? Experts in social indicators or higher education seem reluctant to single out individual ranking systems for criticism. Some experts are complicit in rankings schemes; and perhaps certain rankers are *de facto* protected by their power to harm institutions.[2]

Given that social science is largely silent, it is unsurprising that the rankings community also fails to take a rigorous approach. The rankings community tolerates all rankings. It seems to be widely accepted that no single ranking can be complete or perfect; and all rankings, being partial, are equivalent in merit — or at least all rankers have an equivalent commercial or intellectual 'right' to practise, as if ranking is an inclusive club. After a decade of university rankings it is apparent that self-regulation by the rankings industry will not distinguish good rankings from bad on grounds of validity of data and methodologies or foster a common culture of improvement based on scientific principles. This was underlined by the decision of the International Ranking Expert Group (IREG) on 15 May 2013 to certify the Quacquarelli Symonds (QS) global ranking (QS, 2013b) despite its problems of methodology (see below) and its 'star' system.[3]

This Article

In an ideal world, it would not matter whether rankings rested on sound social science. Competition for status would be stilled. None would earn social esteem at the expense of others. All universities would be very good. Under these conditions, global university ranking would be not just undesirable but absurd. Ranking exacerbates a market-like competition in higher education that is redundant, wasteful and destructive. Ideally, performance measures should not create a pattern of winners and losers. Yet this not the world we inhabit. Global rankings are a fact of life for research universities even in egalitarian systems with modest tendencies to institutional hierarchy and contest. Rankings are consistent with older ideas of venerable university status; and in some systems, including those in the US, East Asia and Russia, higher education is openly stratified and defined as a 'market'. Research capacity is both unequally distributed and the primary engine of university status. All this sustains the ranking culture.

It is necessary to critique the rankings culture and imagine a world beyond ranking and market simulacra, if such a world is to be achieved (Marginson, 1997, pp. 278–281). Yet there remains the strategic issue of what to do about rankings in the present. The practical question is not how to get rid of university ranking — impossible in the foreseeable future — but how ranking might evolve so that its negative effects are minimised, it better serves the common interest, and provides optimal comparative information.

In other words, while ordinal cross-border comparison is inevitable, the forms of comparison are an open question. Existing rankings are not fixed and new rankings emerge all the time. Arguably, however, unless ranking is grounded in social science, data will be misleading, ranking will have more perverse effects, and the comparative data it provides will have limited value. There is scope for change in ranking systems if there is the will to achieve it. A preferred approach to ranking should join social science foundations to a normative agenda to (1) improve the social science quality of rankings data, and (2) optimise the behavioural effects of rankings in relation to performance, and for the greatest possible number of institutions and national systems, not just some leading or upwardly mobile institutions. It is possible to design a better university rankings system in these terms. Such potentials are the starting point for this article.

The next section proposes criteria for judging rankings. These criteria are then applied to selected global rankings systems. The conclusion suggests a way forward.

Criteria for Evaluating Rankings Systems

At least eight criteria can be brought to bear in judging global rankings systems. The first six relate to social science quality. The final two concern behavioural effects.

Criteria Designed to Improve Data Quality

The six criteria for improving the social science of rankings are materiality, objectivity, externality, comprehensiveness, particularity, and ordinal proportionality.

Materiality: The first test of good rankings is the extent to which they are grounded in higher education realities. The starting point of social science is accurate and consistently recorded observation. Indicators should measure higher education where it can be measured, with maximum precision. Following observation and data collection collation and interpretation of data should minimise the scope for the ranking organisation to over-determine the material phenomena and frame the outcome of comparison.

Objectivity: Following from the materiality principle, rankings indicators should eschew subjective observations using measures such as Likert scales. While surveys may bring research close to the real higher education world, the subjective filter compromises data authenticity and coherence. Surveys provide not data about the real world but data about opinions. These opinions may or may not rest on observation; individual survey respondents are not trained in the observational methods of social science; and subjective observations across a population do not use a common method. It is invalid to combine numbers thus derived within a unitary set. The method is also imprecise. The average response of a group of respondents to a survey is no more able to accurately identify the absolute quality of a university than an average response from the same people could accurately identify the distance between the Earth and Jupiter. Universities can be placed in an order of preference, but it is invalid to attach absolute numbers on the basis of position in that order

All social science entails a subjective element, in that the researcher's preconceptions are brought to bear on the research process. However, one essential task of the researcher is to maximise openness to the material world and minimise the tendency to screen out what can be seen. Further, university ranking should not be a popularity contest. Subjectively-based ranking recycles established status. It reflects an historical hierarchy that may or may not be current. This is known as the

'halo effect'[4], whereby overall preconceptions colour specific judgements. Subjective data in higher education, such as student evaluation of teaching or academic ratings of universities, may be used for other research purposes such as measures of customer satisfaction and university reputation. Subjective measures should not be used for ranking. The bottom line is that survey data can vary on the basis of factors other than variation in the material world. In grounding university ranking materially there must be minimum 'noise' from subjective factors.

Externality: Rankings should be based on data from sources external to the universities being ranked. Data collection and interpretation should not be open to manipulation by parties affected. Ideally, rankings data should be drawn from global collections of data so that all institutions are judged on a common basis.

Comprehensiveness: Rankings of universities should be as comprehensive as possible of university functions. Rankings of a particular aspect such as research should be as comprehensive as possible of that activity, ideally including all disciplines, all types of research input and output, graduate research in quantity and quality, all kinds of use and impact of the outcomes of research, the positive effects of research on teaching, localities, cities and regions, the national economy, other countries, etc. Comprehensiveness should not be achieved by using measures invalid under other criteria such as objectivity.

Particularity: Rankings based on particular qualities should incorporate measures of those qualities consistent with those in the real world. This means that measures of specific quality A should not be discounted, qualified or substituted for by measures of other qualities (B or C, D, etc.), unless quality A is in part or whole identical with quality B or C, D, etc. Furthermore, particular qualities measured for ranking purposes should only be combined into a unitary calculation when it can be demonstrated that these qualities form an identifiable whole in the real world that is proportionate to the ranking model.

The principle of particularity rules out proxy indicators, such staffing ratios as a substitute for measures of the quality of teaching, or data for men as a substitute for data on women and men combined. The principle of particularity also rules out composite indicators based on arbitrary weightings. A ranking that combines budgetary allocations (50%) with research publication (50%) is invalid as a measure of either budgetary allocation or publications. It is also invalid as an overall measure of the position of the university because there is no basis for the assumption that half the social value of a university lies in budgetary inputs and half in research publications, and other elements enter into the social value of a university, such as its teaching and service functions; and because the two qualities are not mutually exclusive. Budgetary input affects publication output and vice versa.

Ordinal proportionality: Ordinal ranks should not exaggerate distinctions between ranked universities. For example, when there are fine differences between institutions in the 300s there seems no good reason to rank them vertically in a league table.

Criteria Designed to Optimise Behavioural Effects on Performance

The two criteria for maximising the behavioural effects generated by rankings are performance alignment and transparency.

Performance alignment: Any ranking system should encourage behaviours that are consistent with the maximisation of absolute performance in all ranked institutions, and install a common dynamic of improvement, without compromising

breadth and quality. This alignment is hard to achieve because (a) most ranking systems are reductionist, favouring some disciplines, institutions and types of work over others; and (b) rankings are structured in terms of relative, not absolute performance. Competition for rankings is zero-sum. In a hierarchy, the number of leading institutions is fixed, as always with positional goods (Hirsch, 1976; Marginson, 1997). Unmodified ranking-oriented competition tends to bifurcate outcomes (winners/losers). In building reputation, rankings help leading institutions to draw a more than average share of the resources necessary for high performance: status, money, talent. This tends to lock out others. League table hierarchies tend to function as partly closed systems that reproduce the oligopoly of leading institutions — unless the rankings system is modified to enhance contestability.

One test of the extent to which rankings encourage all-round performance improvement is the scope for upward mobility, the entry and rise of new institutions. It is impossible to create a ranking system in which the process is entirely 'fair', in the sense of a level playing field (full contestability): historical inequalities ensure different starting positions. Nevertheless the scope for upward mobility can be increased.

Transparency: Indicators, measures and methods of compiling, computing and interpreting ranking data should be fully transparent to all parties. The more open and simple the better. Ideally, it should be possible for any literate person to reproduce the calculations, from raw data to final product, and so understand the basis of ranking outcomes and respond in a straightforward fashion to the incentive structure.

These eight criteria will now be applied to the evaluation of six university ranking systems: ARWU, Leiden, QS, Scimago, Times Higher Education and U-Multirank. The outcome is summarised in Table I. The judgements, which are explained further in the next section, are approximate. No doubt some will be contested.

Evaluation of Six Current Rankings Systems

Academic Ranking of World Universities (ARWU)

The Shanghai Academic Ranking of World Universities (ARWU, 2013) uses a composite multi-indicator. University performances in each of the six indicators that comprise the index are scaled and combined in a single series to enable creation of a league table. The weightings between indicators are arbitrary, though because the indicators are all research focused, performance in each indicator tends to correlate closely to performance in the other indicators (Cheng, 2011). The indicators are Nobel Prizes in the science disciplines and Fields Medals in mathematics won by graduates (10%), the same awards won by current faculty (20%), high citation faculty researchers (20%) (ISI-Thomson, 2008)[5], number of papers in *Science* and *Nature* in the previous five years (20%), number of papers indexed in the Web of Knowledge citation list the previous year (20%), and the above indicators combined and expressed on a per full-time faculty basis (10%). Universities are ranked from one to 100 and then grouped as 101–150, 151–200, 201–300, 301–400, 401–500. This is better ordinal proportionality than in the other multi-indicator rankings discussed here, though perhaps the format still exaggerates distinctions between universities ranked in the second fifty.

TABLE I. Evaluation of six university ranking systems

Evaluation Criterion	Academic Ranking of World Universities	Leiden University	QS	Scimago	Times Higher Education	U-Multirank
Materiality	MEDIUM-STRONG Do graduate Nobels (10%) connect HEI?	STRONG Publications, cites drive outcomes	WEAK Outcomes strongly shaped by methods	STRONG Publications, cites drive outcomes	WEAK Outcomes strongly shaped by methods	MEDIUM Good HEI connect but subjective data
Objectivity	STRONG Prizes, publications, citations only	STRONG Publications and citations only	WEAK Surveys 50% of absolute position	STRONG Publications and citations only	WEAK Surveys 34.5% of absolute position	WEAK All indicators composed by survey
Externality	MEDIUM-STRONG HEIs may influence staff count (10%)	STRONG All data collected by Web of Knowledge	MEDIUM-WEAK Data in a few areas negotiated with HEIs	STRONG All data collected by Scopus	WEAK Data in some areas negotiated with HEIs	STRONG All data collected externally to HEIs*
Comprehensiveness	MEDIUM-WEAK Part-comprehensive of research only	WEAK Limited research data series only	MEDIUM Reputation, research, internationalisation	WEAK Limited research data series only	MEDIUM-STRONG As for QS plus PhDs, financial resources	STRONG Many data teaching, research, services
Particularity	MEDIUM Multi-index though internally correlated	STRONG Stand-alone indicators only	WEAK Proxy, multi-index arbitrary weights	STRONG Stand-alone indicators only	WEAK Multi-index, arbitrary weights	STRONG All indicators can be made stand-alone
Ordinal proportionality	MEDIUM Single ranks to 100 then large groups	STRONG Indicators permit valid single ranks	WEAK Single ranks to 400 then groups to 700	STRONG Indicators permit valid single ranks	WEAK Single ranks to 200 then groups to 400	STRONG Three broad categories, no ranks
Performance alignment	MEDIUM-WEAK Research bias. Nobel block in steep ladder	MEDIUM Research bias. Clear goals. Steep ladder	(MEDIUM) WEAK (Status). Fake range, volatility, mobility	MEDIUM Research bias. Clear goals. Steep ladder	(MEDIUM) WEAK (Status). Fake volatility, mobility	STRONG Teaching, research service all rewarded
Transparency	MEDIUM Scaling opaque for non-specialists	MEDIUM-STRONG Partly opaque for non-specialists	WEAK Surveys and scaling opaque	MEDIUM-STRONG Partly opaque for non-specialists	WEAK Complex, surveys and scaling opaque	MEDIUM Customization but surveys invisible

* Areas where the author's judgement is provisional and more data are needed.
Source of concept and table: author

In terms of data quality, ARWU has no areas of fundamental weakness. Its strong points are materiality, objectivity and reliance on external data. These factors have generated a high level of trust in the ARWU (it also has first mover advantage as the first credible research ranking). ARWU is closely grounded in real world high science research output and impact: leading researchers, primary science publications, citations; though there are no indicators for graduate students or PhDs, and works in the humanities and part of the social sciences and professional disciplines are excluded because they are not globally comparable. The only doubt about its materiality is the weakness of the connection between education of a Nobel winner in the past, and present scientific capacity.[6] All the ARWU data are objective: all can be observed and counted. Only the number of equivalent full-time faculty is open to manipulation by the ranked universities. However, the ARWU is not comprehensive of all university functions and resources. For example, it excludes teaching, social mission, resources and internationalisation.

The incentive structure of the ARWU is transparent to research managers. It focuses the attention of universities and national systems on basic research; and encourages a strong relationship between ranking, performance objectives, and specific research management strategies like hiring Nobel Prize winners and high citation researchers, rewards for faculty publishing in *Nature* and *Science*, and growth of science papers. It elevates research-related objectives relative to teaching-related objectives, though other forces also push research universities in this direction. However, few universities have the conditions and resources to achieve high ARWU rank. The Nobel indicators (30% of the index) are an additional barrier. Past Nobels are mostly confined to USA, Europe, Russia and Japan, constraining possible upward mobility. Top Asian research universities such as the National University of Singapore and Seoul National University in Korea tend to do less well in the ARWU than in other credible research rankings.

Leiden University

The Leiden (2013) ranking is issued by the Leiden University Centre for Science and Technology Studies (CWTS). It is not a composite multi-indicator. Leiden supplies separated rankings of universities, each based on single indicators, including the volume of science papers, volume of citations of those papers, citations per paper, the number of papers in the top 10%of their field by citation rates, and the proportion of the university's papers in the last category. The citation data are provided in both raw form and on a field-normalised basis, whereby the Leiden group adjusts the raw data to account for different rates of publication and citation in the various research fields.

The strengths of both the Leiden and Scimago (below) rankings are materiality, objectivity, and externality. Leiden's data are from Thomson-ISI Web of Knowledge. There is no combined index and no arbitrary weightings. Focusing directly on publication and citation numbers and comparative citation quality, and uncontaminated by multi-indicator weights and subjective surveys, Leiden and Scimago bring data users close to research realities. Leiden offers more than Scimago on citation quality. Both use single indicators so they evade problems of ordinal proportionality. The use of single indicators lifts Leiden and Scimago to a higher level of quality as social science, compared to multi-indicator rankings. These rankings step back and shape the outcome as little as possible. This frees the data user to interpret the data contextually and normatively as desired.

The Leiden data are transparent in meaning and easy to read, though non-specialists may find field normalisation to be opaque. Leiden and Scimago are less well-known in higher education and public circles than ARWU, Times Higher and QS and have a modest role in determining university reputations. However, the Leiden rankings are highly regarded among specialists in research policy and management, and the study of science. Leiden provides an effective foundation ranking in support of strategies for maximising the overall ranked position. Better performance in this ranking feeds into stronger performance in the ARWU and, with some distortion, Times Higher and QS, given that research outputs play an important role in both, especially Times Higher. As with the ARWU, however (though without the Nobel barrier) the targets implied by Leiden are hard to achieve. Only strong research universities can compete. Like all *bona fide* research rankings, Leiden sustains upward mobility primarily among institutions where the investment in capacity is growing. It is not a universal performance driver.

Quacquarelli Symonds (QS)

The QS (2013a) ranking is a multi-indicator ranking that standardises and scales data in five areas to derive a single league table. It uses a steep hierarchy with single ranks from 1–400 and groups institutions in tens and fifties thereafter up to 700: it is weaker in ordinal proportionality than the other rankings discussed here. Small differences in performance have exaggerated effects. Like the Times Higher, which is similar in method, QS aims to be comprehensive of more than just research. Its indicators cover 'academic reputation' for teaching and research (40%), reputation among graduate employers (10%), 'teaching' using the proxy indicator staff-student ratio (20%), citations per full-time academic faculty (20%), and the proportion of faculty (5%) and students who are international (5%). The two reputation indicators are determined by surveys that together constitute half the QS ranking index. The correlation between the different indicators is much weaker than in ARWU (Cheng, 2011). QS uses methods from market research not academic sociology. As often in market research, issues like response rate, survey cells and weights remain opaque. No information has ever been provided about which 'graduate employers' are surveyed.

QS rates poorly in the three areas where the research rankings do well: materiality, objectivity and externality. Fifty per cent reliance on surveys is high. This displaces the ranking process from observable reality; especially given that few survey recipients can make comparative judgements about more than two or three universities. The QS ranking is not a pure reputation ranking either because half the index is comprised by objective elements. But these observations of material higher education pass through a series of filters entailing QS decisions, and these filters shape the ranking outcomes, for example the arbitrary weighting between the indicators, and the scaling process: the conduct and compilation of surveys, including questions used, and interpretation of survey returns, for example whether to compensate for uneven returns by area; and also negotiations with institutions on number of faculty, and numbers of international staff and students.[7] Particularity is compromised by the multi-indicator and by the use of staff-student ratios as proxy for teaching quality (note also that a pure quantity measure is a poor proxy for quality). Some specific QS indicators provide valuable information, like the student-staff ratio and citations per faculty. The multi-indicator approach buries those useful data.

The QS ranking is subject to much annual fluctuation, as is the Times Higher ranking. What determines this volatility? Does it indicate that QS permits a higher degree of upward mobility than, say, ARWU? This volatility is a product of two elements. First, the multi-indicator character of the ranking, in combination with many institutions close to each other in score and low correlations between indicators. Small changes in one indicator can sharply affect position. With poorly correlated indicators moving inconsistently in relation to each other, there is much potential for random effects. Second, surveys provide half the ranking and are subject to year-by-year changes in the returns. In some years a third element adds volatility: methodological changes. This is fake volatility driven by ranking method, not real-world volatility based on openness and contestability.

The large role played by the surveys also enables universities to improve their QS ranking following marketing and other reputational building activities, including gaming of survey returns. To this extent, universities have better prospects of improving in this ranking in a short time than in ARWU, Leiden or Scimago. In this one respect, QS provides real rather than fake mobility. However, this upward mobility can be achieved only in relation to status and the effect on status is temporary. There is no necessary link from the improvement in status back to improved real performance, of the kind that delivers a sustained improvement in ranking, as with a lift in ARWU position. But in any case, in QS, especially below the top 30–40 universities, the relationship between performance and ranking is not transparent. It is crowded out by the many ways QS shapes the outcome. The relationship between performance and ranking may not be very strong at all.

The QS rankings régime cannot necessarily lead to improved all-round performance for three reasons. First, there is no given theoretical base for the chosen indicators and their internal relations and weights, for example no coherent definition of production or performance. Second, if the arbitrary weights are shifted (e.g. citations per faculty is 10% and internationalization is 20%) there are major shifts in ranking positions without any concurrent change in real performance (Gladwell, 2011). Third, the relationship between real effort and real performance is filtered by QS methodologies and liable to be over-determined by surface fluctuations in position. This means that if there are positive effects in relation to performance generated by managing a university so as to maximise the QS ranking, these effects are essentially accidental. In that case, it is difficult to see what benefit is served by the QS global rankings, aside from the commercial benefits to QS itself in marketing its higher education-related business.

Scimago

The Scimago (2013) data are based on the Scopus collection by Elsevier. Scimago provides data on all research organisations, not just higher education. It provides a much longer list than Leiden. The data can be adjusted so as to include only higher education institutions. There are also more research-related indicators than provided by Leiden but the data as presented are less accessible and explanations are more complex and less transparent. The most useful indicator of quality is 'normalised impact' that measures average citations. For the most part the issues in relation to the Scimago ranking are similar to those discussed in relation to Leiden. Scimago is strong in materiality, objectivity, externality, particularity and, because it is a set of single indicators and not a composite multi-indicator ranking, ordinal proportionality. Shaping by the ranking agency is transparent and modest.

The data are confined to research and to a limited set of information about research, focused on primary scientific outputs and the global disciplines, and the academic impact of research in those disciplines as expressed through citation patterns.

As with the Leiden ranking and ARWU, Scimago provides a clear picture of the hierarchically distributed character of research outputs. Only strong research universities can compete successfully for upper level Scimago positions. One of the virtues of this very long list of research producers is that improvements lower down the scale are made visible. In that respect Scimago is more inclusive than Leiden and ARWU, which confine themselves to top 500s. Nevertheless, like the other research rankings, Scimago drives upward mobility primarily among institutions with growing investment in capacity.

Times Higher Education (THE)

The Times Higher Education ranking (THE, 2013) is similar in form to QS ranking and invokes similar issues; though arguably, it is managed by Thomson-Reuters at a higher level of competence than QS. For example, survey returns seem larger and the geographic coverage more effective (for QS survey problems see Sowter, 2007). The main strength of Times Higher is comprehensiveness. Thomson-Reuters uses 13 separated indicators that are weighted, scaled and pressed into the final unitary number. Reputational surveys for research and teaching constitute 34.5%, bibliometric indicators 34.5%, income indicators 10.75%, PhD studies 8.25%, internationalisation 7.5% and the student-staff ratio 4.5%. Rankings in these different areas are not closely correlated (Cheng, 2011). As with QS, there is no direct measure of teaching quality or learning achievement: there is a survey of reputation for teaching and the student-staff ratio. The data on income for research purposes are incomplete and difficult to standardise across borders, but, arguably, are unique and interesting. Unfortunately, they are lost in the composite indicator. In total, 73.25% of the Times Higher rankings are constituted by one or another aspect of research performance: research reputation via survey, citations, research volume, research-related income, international research collaboration and PhDs. Comparative research data are easier to gather in relation to research than teaching and service functions. Research is also central in shaping university reputation — as with QS, the Times Higher ranking appears to have been designed as a reputational table not a performance table.

Like QS, the Times Higher ranking is weak in particularity, objectivity (one third reliance on survey data) materiality and externality. Data standardisation and weighting trump empirical observation. Externality is compromised by the scope for institutional influence across the 13 indicators. There are single rankings to 200 and then groupings to 400: ordinal proportionality is stronger than QS but weaker than ARWU. As with QS, there is annual fluctuation in ranking positions. Except to the extent that marketing-related factors shape position, this fluctuation does not indicate *bona fide* capacity for mobility. Below the top group of universities the relationship between performance and ranking is over-determined by many elements. There are many equally possible rankings on the basis of the data collected. An additional difficulty is that the index is complex. Even if random volatility was not a factor it would be difficult to design systems that optimise Times Higher ranking outcomes. Most universities seem to focus attention on boosting their position in the survey and negotiating with Thomson-Reuters to

secure favourable data interpretation in areas like internationalisation. There is no evidence that these behaviours constitute a régime of global comparison that fosters all round improvement.

U-Multirank

U-Multirank (2013) has several novel features. It provides the most comprehensive information and focuses on the information user (van Vught & Ziegele, 2011). It does not provide league tables but groups institutions in three performance bands. This is a very good outcome in terms of ordinal proportionality, though it fails to satisfy desires for a hierarchical set. U-Multirank collects and provides data in a large number of areas of teaching and learning, research and services, for both disciplines and institutions. There are no problems of multi-indicator weightings in that all indicators can stand alone. It is planned that the U-Multirank website will allow users to choose their own criteria and weightings when making comparisons. This radically reduces the shaping role of the rankings agency when compared with the Times Higher, QS and even ARWU.

By covering a broad set of aspects of higher education U-Multirank takes its users close to the materiality of higher education. The fly in the ointment is that its comparison is especially weak in objectivity. All data are generated by survey, undercutting materiality, though externality is strong. The U-Multirank data explain the standing of institutions and disciplines, focusing on academic reputation and, more indirectly, student satisfaction, but do not provide comparative data on quantity or quality. Transparency is also weak — survey data gathering, standardisation and interpretation are opaque to the user — though U-Multirank gains points for its use of customised indicators and non-normalising weights. A performance régime based on U-Multirank is likely to improve customer satisfaction. Whether underlying performance also improves is unclear. On the plus side, U-Multirank ensures that institutions focus on the full range of activities rather than only on research; and by setting aside single league tables it discards the zero-sum element in ranking. A U-Multirank system of comparison fosters across-the-board improvement in customer satisfaction and academic standing (though not necessarily in the substance of activity) in all institutions. It constitutes the most inclusive performance régime of all the rankings discussed.

Conclusions

It is highly desirable that social scientists in higher education studies or other fields become more active and proactive in analysing, critiquing and proposing alternatives to present university ranking systems. This should not be left to the rankers.

In general, it is preferable to use comparisons of universities that focus on the substance of performance, not reputation. Rankings have an irreducible reputation-making role. It is highly desirable to ground that role in the materiality of performance transparently — rather than use comparisons in which reputation drives reputation in a circular effect that is unmediated or only partly mediated by material performance. Circular reputational rankings function as a competition game that is an end in itself. This benefits only the leading institutions and does nothing to enhance teaching, research and service overall. Survey-based data on reputation are interesting in themselves. They contribute to the picture of higher education as a positional market. However, such data should be provided entirely separately from rankings based on quality, output or performance.

The most useful rankings are those that are closely lodged in the realities of higher education and use concrete measures (materiality), not filtered by survey opinion (objectivity), and not using data from the universities affected (externality). Within these constraints, the more comprehensive the comparative data, the better. However, different indicators should not be combined into single multi-indicators using weights. While, nominally, arithmetic can combine heterogeneous data sets into one standard set, this violates the relationship between observation and the material higher education world. *What is apprehended is not the material world but the artificial model-world of the ranker.* Again, this pushes the process back towards the dystopia of the ranking game as end in itself, breaking the virtuous nexus between ranking and performance.

When rankers use broad bands not league tables, this reduces the normalising effects of rankings, the tendency to bifurcate the ranked units between winners and losers, and the tendency to close off ranking competition to all but the established institutions. However, it is doubtful if this can still the widespread desires for positional hierarchy. One way to provide space for emerging institutions is to develop separate rankings of such institutions. For developing countries with, say, a per capita income of USD $5000 p.a., what is to be gained by competing with countries with USD $40,000 or more?

At present, in university ranking, the strongest comparisons are confined to research. There, global ranking has driven a widespread dynamic of increased investment in research capacity (Hazelkorn, 2011). Of the existing rankings, Leiden and Scimago provide the best research data in terms of social science. The ARWU collection is also useful, though it would be better if it were Nobel-free. U-Multirank is state of the art in some ways — customised comparisons and comprehensive coverage — but sadly stymied by its reliance on surveys.

All three of Leiden, Scimago and ARWU are limited by their focus on research. At this stage, there are no available valid indicators of comparative teaching quality or learning achievement. Because multi-indicator rankings that conflate heterogeneous functions into one number are invalid, it would be better if Times Higher and QS (and also US News and World Report, which pioneered the approach) stopped producing single league tables. This does not mean the data collected by multi-indicator ranking organisations lack value. If the Times Higher, QS and ARWU provided disaggregated data in the form of separate league tables, in the manner of, say, Leiden, there would be much interest in those data. More specific data would facilitate the evolution of targeted strategies for institutional improvement and foster a more coherent relationship between ranking and performance. This would also encourage mission specialisation, rather than the normalised institutional form now generated by the combined indicator league tables. That is the institution that pretends to be all things to all people, grows as large as possible, and imitates the practices of the ideal Anglo-American research university.

Simon Marginson, Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, United Kingdom, s.marginson67@gmail.com

NOTES

1. Disclaimer: In ascending order of closeness of affiliation, the author is a member of the Editorial Board of the Times Higher Education (i.e. the

publication rather than the ranking), a member of the International Advisory Board of the Shanghai Academic Ranking of World Universities (ARWU), and one of the authors of the U21 ranking of higher education systems, which is not discussed here. He considers himself free to comment without fear or favour in relation to all rankings, a right supported by all three organisations named. This article contains criticisms of both ARWU and the Times Higher ranking.

2. Opaque methodologies enable rankers to fix the exact position of one or another university by small changes in standardisation and scaling, weights, interpretation of survey data, inclusions/exclusions of survey particular returns, etc. In addition, QS has sought to protect itself from academic criticism by threats of legal action against publications carrying such criticism. The author has experienced this twice.

3. QS awards individual universities four or five stars provided they undergo a QS 'evaluation' at a cost reported to be $15,000–20,000. These stars are prominently displayed on university websites as if they are a form of quality certification. QS has been able to create this marketing service, in the transparently bogus form of a formal evaluation, because of its role in global rankings.

4. A notion attributed to the educational psychologist Edward Thorndike (Wikipedia, 2013).

5. This indicator is no longer collected by ISI-Thomson. It will be replaced by a new indicator yet to be announced at the time of writing.

6. Oddly, ARWU (2013) claims the Nobel alumni measure as an indicator of teaching quality.

7. This means that QS staff have much scope to fix the position of any institution in the ranking if they so desire.

REFERENCES

ACADEMIC RANKING OF WORLD UNIVERSITIES (ARWU) (2013) Shanghai Jiao Tong University Graduate School of Education. www.shanghairanking.com/

BOURDIEU, P. (1986) *Distinction. A Social Critique of the Judgment of Taste* (London, Routledge and Kegan Paul).

CERNY, P. (1997) Paradoxes of the competition state: the dynamics of political globalization, *Government and Opposition*, 32, pp. 251–274.

CHENG, Y. (2010) *To better measure social sciences performance. A review of existing ranking indicators*. IREG 2010 conference, Berlin, 6–8 October. www.iregobservatory.org/index.php?option=com_content&task=view&id=92&Itemid=86

CHENG, Y. (2011) *The History and Future of ARWU*. Paper presented to the inaugural meeting of the ARWU International Advisory Board, Shanghai, 30 October.

GLADWELL, M. (2011) The order of things, *The New Yorker*, 14 February. www.newyorker.com/reporting/2011/02/14/110214fa_fact_gladwell

HAZELKORN, E. (2008) Learning to live with league tables and ranking: the experience of institutional leaders, *Higher Education Policy*, 21, pp. 193–215.

HAZELKORN, E. (2011) *Rankings and the Reshaping of Higher Education: the battle for world-class excellence* (Houndmills, Palgrave Macmillan).

HIRSCH, F. (1976) *Social Limits to Growth* (Cambridge, Harvard University Press).

INSTITUTE FOR SCIENTIFIC INFORMATION, ISI-THOMSON (2008) Data on highly cited researchers, http://isihighlycited.com

LEIDEN UNIVERSITY (2013) (Centre for Science and Technology Studies, CWTS) *The Leiden Ranking 2013*. www.leidenranking.com/default.aspx

MARGINSON, S. (1997) *Markets in Education* (Sydney, Allen and Unwin).

MARGINSON, S. (2008) Global field and global imagining: Bourdieu and relations of power in worldwide higher education, *British Journal of Sociology of Education*, 29, pp. 303–316.

OECD (2013a) *Education at a Glance 2013* (Paris, OECD).

OECD (2013b) *Program of International Student Assessment*. www.oecd.org/pisa

QS (2013a) *QS World University Rankings*. www.topuniversities.com/university-rankings

QS (2013b) *QS World University Rankings Approved by IREG*. www.topuniversities.com/university-rankings-articles/world-university-rankings/qs-world-university-rankings-approved-ireg

RAUHVARGERS, A. (2011) *Global University Rankings and their Impact*. A report for the European Universities Association (EUA) (Brussels, EUA). www.eua.be/pubs/global_university_rankings_and_their_impact.pdf

SALMI, J. & SAROYAN, A. (2006) *League tables as policy instruments: uses and misuses*. OECD/IMHE and Hochschulrektorenkonferenz, Germany, Workshop 4–5 December, Bonn. OECD, Institutional Management in Higher Education program. www.oecd.org

SAUDER, M. & ESPELAND, W. (2009) The discipline of rankings: tight coupling and organizational change, *American Sociological Review*, 74, pp. 63–82.

SCIMAGO (2013) *Scimago Institutions' Rankings*. www.scimagolab.com

SMITH, A. (1776/1979) *The Wealth of Nations* (Harmondsworth, Penguin).

SOWTER, B. (2007) *THES-QS World University Rankings*. Symposium on International Trends in University Rankings and Classifications. Griffith University, Brisbane, 12 February. www.griffith.edu.au/conference/university-rankings

TIMES HIGHER EDUCATION (THE) (2013) *Times Higher Education World University Rankings*. www.timeshighereducation.co.uk/world-university-rankings/

U-MULTIRANK (2013) *U-Multirank Portal*. www.u-portal.org/u-multirank

USHER, A. & SAVINO, M. (2006) *A World of Difference: a global survey of university league tables*. www.educationalpolicy.org

US NEWS AND WORLD REPORT, USNWR (2013) *National University Rankings*. http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/national-universities

VAN VUGHT, F. & ZIEGELE, F. (Eds) (2011) *U-Multirank. The design and testing the feasibility of a multi-dimensional global university ranking*. Final Report. Consortium for Higher Education and Research Performance Assessment, CHERPA-Network. http://ec.europa.eu/education/higher-education/doc/multirank_en.pdf

WIKIPEDIA (2013) *The Halo Effect*. http://en.wikipedia.org/wiki/Halo_effect