

World university ranking methodologies: stability and variability

Brian Fidler and Christine Parsons

There has been a steady growth in the number of national university league tables over the last 25 years. By contrast, 'World University Rankings' are a more recent development and have received little serious academic scrutiny in peer-reviewed publications. Few researchers have evaluated the sources of data and the statistical approaches used. The present article seeks to address this gap.

The authors explain and evaluate the methodologies used by the Times Higher Education Supplement¹ and Shanghai Jiao Tong University, highlighting differences in their outcomes and in their degree of stability over time. A range of concerns must be addressed if such rankings are to inspire a level of confidence which transcends the established 'infotainment' value of league tables (Bowden, 2000).

Keywords: university ranking, university league table, world class university

For some 25 years there has been a steady growth in the number of national rankings of universities. These started as commercial undertakings. The first, for US universities, was in 1983 by *US News and World Report* (Merisotis, 2002). UK newspapers then began to develop their own approaches to ranking UK universities. The most developed is that jointly used by *The Times* and *Times Higher Education Supplement* from 1993. It was clearly commercial concerns that led to what has become an annual event (Jobbins, 2002). Ostensibly, the prime audience for such lists of national rankings was potential undergraduates and subsequently potential postgraduates, although there is little evidence that these rankings influence student choice in either the UK or the USA (Eccles, 2002). The originators of university rankings had nevertheless identified a topic of human interest. In a similar way to other popular news stories that the media choose to publish, league tables are a potent

source of ‘infotainment’, where a certain amount of information provides a disproportionately large degree of entertainment (Bowden, 2000).

While the first national rankings were controversial and attracted much legitimate criticism (Berry, 1999; Provan and Abercromby, 2000; Clarke, 2002), they have persisted and developed, and the producers have taken this as acceptance by the academic community of such rankings (Merisotis and Sadlak, 2005). Indeed they cite the discussion of the results of such rankings by academics as confirmation of their value.

Drivers to produce international league tables

There are a number of interconnected influences that are driving the growing appetite for international comparisons. Universities have long been considered international organisations and the broad trend is one of increasing internationalisation (Parsons and Fidler, 2004). One of the aspects of globalisation has been to create competition between universities in different countries rather than within a single country. Staff and students are more internationally mobile, the latter particularly at postgraduate level, and for some purposes universities are in an international market. Those making such choices need guidance.

Perhaps the key underlying driver is the rise of the knowledge society and its economic impact. In a knowledge society, knowledge replaces physical resources as the main driver of economic growth (Wooldridge, 2005). It is widely recognised that higher education has an important role to play in the creation and transfer of knowledge to the economy (see for example, Universities UK, 2006). While all universities might potentially make some contribution to the economy, ‘world class’ universities, with their strong science and innovation capabilities, are likely to generate ‘outsize economic benefits’. For example, in the USA, Stanford University helped to incubate the search engines Google and Yahoo. The University of Texas at Austin has assisted the creation of a high technology cluster that employs around 100,000 people in some 1,700 companies. In 2000, the eight research universities in Boston provided a \$7.4 million boost to the region’s economy (*The Economist*, 2005). At the same time, ‘science’s appetite for money and manpower’ requires high levels of resource for this contribution to be achieved and sustained. Massachusetts Institute of Technology’s Lincoln laboratory employs nearly 2,400 people and spends \$450m on research (*The Economist*, 2005).

Whilst there are major differences between countries in the approach to building and developing a knowledge-based society and economy, an analysis of global investment in research and development shows that

‘no single country has succeeded in achieving and sustaining high levels of prosperity ... without investing in science and technology and exploiting them’ (Westholm *et al.*, 2004: 28). Thus scientific and technological innovation, and its application, can be seen as keystones of the knowledge society and catalysts for economic growth.

Interest in comparing national performance with the best international practice fits with the ‘world class’ aspirations of governments. The UK government has identified with world class services in education since 1997 (Barber, 1998; Barber and Sebba, 1999). The permanent secretary of the Department for Education and Science, David Bell, has publicly committed the department to securing ‘a truly world class education system’ (DfES, 2006: 2). ‘World class’ can be defined as ‘of or among the best in the world’. Thus ‘world class’ implies an international comparison involving all other countries. Public statements from governments in other countries are similar. Such policies lead to pressure on the leading national universities to demonstrate that they are performing among the top ranking international universities.

World rankings

While national rankings achieve some aims that the earlier discussion identified, they do not deal with the globalisation dimension at all. They do not indicate to national governments how their universities compare with those of other countries, nor do they indicate to potential students who are internationally mobile, which are the most desirable universities in the world. Hence there are pressures for international or world rankings. However, they cannot be achieved by combining national rankings, since the criteria used to produce these are not internationally comparable. World rankings, if they can be compiled in a way that inspires confidence, would be highly desirable.

There are two current schemes that have achieved widespread publicity and that are already having some impact on UK universities, even if only at the level of their publicity. The two schemes – Academic Ranking of World Universities (ARWU), carried out by two academics from Shanghai Jiao Tong University in China, and the *Times Higher Education Supplement* rankings (hereafter referred to as *Times Higher*) – use contrasting methodologies both in the criteria used and the type of organisation producing the league tables. The *Times Higher* covers only the top 200 institutions, while the ARWU lists the top 500. However, the large number of institutions with tied ranks means that only the top 100 appear in rank order in the ARWU tables. The rest appear in alphabetical order in bands of 100 institutions, with up to 107 tied ranks in 2007. In

this article, the main interest is in whole institution rankings, although both organisations also provide university rankings for a range of specialist fields.

Academic Ranking of World Universities (ARWU)

Although this ranking emanates from Shanghai Jiao Tong University, as the lead researchers explain, 'The ARWU is academic research driven by personal interest, and carried out independently without any external support' (Liu and Cheng, 2005: 135). Indeed, their website (<http://ed.sjtu.edu.cn/ranking.htm>) now has a pop-up disclaimer to that effect.

The information about the derivation and results of their approach to international university league tables is displayed on the university's website and has been described in a number of articles (Liu and Cheng, 2005). These describe the rationale for the basic approach, but the more detailed decisions involved in the choice of data and their combination are only briefly outlined, without any supporting justification. The ARWU was pioneered in 2003 with the intention of trying to estimate the gap between Chinese universities and world class universities (Liu and Cheng, 2005). The basic stance was to use only publicly available data that could mainly be compiled from an international citation database. The only one available at the time was that of ISI/Thomson Scientific. No data were to be included that were supplied by universities themselves, since such data could not be independently verified. Further rationale for the particular choice of data is not given. Choices about particular journals in which to privilege publication appear to be arbitrary, but to some extent these restrictions follow from the limitation to use only publicly available, internationally comparable data in seeking to discriminate performance at the highest level.

This seemingly innocuous requirement necessitates the exclusion of areas of activity in universities which are not covered by objective, comparable, publicly available data. The predominant components of the rankings are particular aspects of research output. The compilers mainly count various combinations of citations collated by ISI. This decision, seemingly demanded on the grounds of objectivity, in its turn produces bias. The databases used are almost all in English and produced by ISI/Thomson Scientific, a commercial company in the USA (<http://www.isinet.com>). They consist of a suite of indices: the Science Citation Index-Expanded (SCIE), Social Sciences Citation Index (SSCI) and other specialist indices.

Whatever data are used, if there is more than one figure, then the various components have to be aggregated in some way to achieve an overall ranking of universities. This introduces issues of combining different types of data and deciding on a system of weighting of the various types.

TABLE 1 ABOUT HERE

Table 1 summarises the way that the ARWU index is compiled. Overall it appears that measures have been chosen that single out small numbers of universities. These are outliers to the general distribution. A decision has also been made to collate data over long periods of time to form the components leading to the ranking. Although there are six measures, the three measures involving citations are highly correlated (between 0.65 and 0.88 for the top 100 in 2007). As the period over which most of the indicators are calculated is long, yearly changes will make only small differences. This makes the results of this formulation of rankings very slow to change over time. Indeed, there has been only one ‘newcomer’ to the institutions listed in the top 30 in the last three years; a tied rank in 2007 allowed 31 institutions to be listed. As most universities will only contribute to the indicators for ‘research output’, the changes in positions will only be at much lower rankings and these will be marginal in terms of the changes to the underlying variable.

The intention of this article is not to assess the general validity of metrics-based approaches to institutional ranking using various forms of citation, rather it is to examine the particular choices of indicators that are used in the ARWU tables and to consider this approach as an example of using metrics. There are a number of issues.

There are issues concerned with attributing institutions to alumni who have gained degrees at more than one university. There are also problems when staff carried out the work, leading to a prize at a different institution from the one at the time the prize was awarded (van Raan, 2005).

The citations attributed to staff and institutions are compiled from ISI/Thomson sources. The increasing use of this data source to assess the productivity and quality of university output has led to much greater scrutiny of its weaknesses for such purposes. The English language bias inevitably favours universities in the UK, USA, Canada and Australia, as shown in Table 2. There are further biases: citations cover only journal articles and not books or research monographs; the citations indices are much weaker outside the natural sciences; and the choice of journals

from which the citations are taken are heavily dominated by those published in the USA. Furthermore, there are concerns about the types of article included, the cleaning of data and the attribution of citations to institutions (van Raan, 2005). There are examples of tracing errors between seven and 30 per cent in different contexts (Moed, 2002). Since ‘highly cited’ status draws on two decades of citations, the ranking is rooted in history and is likely to be a poor reflection of current performance. An analysis of the authors of the 10 most highly cited articles published between 1996-1999 and 2000-2003 in the ISI/Thomson Scientific database shows that five had changed institutions by 2006 and two had died (Ioannidis *et al*, 2007).

TABLE 2 ABOUT HERE

The ISI/Thomson Scientific databases attribute citations to multiple authors equally. Although the numbers of authors of an article has been observed to vary from one to 865, they all receive equal credit. Thus it is an inconsistency that attributions made by the ARWU team, with respect to articles in *Nature and Science*, involve differential attributions by the number and positioning of the author in the listing.

There are obviously differences in comparing multi-faculty institutions, such as the Universities of Oxford or Cambridge, with monotechnics or specialist institutions, such as London School of Economics. There is also the related issue of size of institution. For some purposes it may be aggregates that are required, while for judgements about efficiency, cost effectiveness or productive considerations, size needs to be taken into consideration. Recent work by Leiden University in the field of European University rankings reinforces the point that rank order varies substantially according to whether research output takes account of size of institution (Center for Science and Technology Studies, 2008). Only one factor in the ARWU index takes account of size; 90 per cent of the data are unweighted in this respect.

The degree to which the ARWU results are reproducible has been the subject of recent investigation. Using the specified data sources, Florian (2006, 2007) tried to reproduce the 2005 results. He found ambiguities in the calculation of the number of staff in institutions but more worryingly found that ‘the values for an objective indicator such as SCI [Science Citation Index] cannot be reproduced using the published methodology’ (Florian, 2006: 5). Correspondence in the name of ‘The Ranking Team’ from Shanghai Jiao Tong University admitted to ‘statistical treatment’ having been applied, maintaining that this did not

affect reproducibility and refusing to provide raw data for comparison (*op cit*: 6, 7).

An 'International Ranking Expert Group' (IREG), founded by UNESCO European Centre for Higher Education in Bucharest and the Institute for Higher Education Policy in Washington, DC in 2004, advises the ARWU compilers on matters including methodology, accountability and quality assurance. Members have generated and discussed a range of papers and presentations since its inception (see <http://www.arwu.org>), and in 2006 they drew up a 'set of principles of quality and good practice' in university rankings (CHE, 2006: 1). However, any refinements to the ARWU approach to date appear to be marginal and inadequately documented.

Times Higher Education Supplement World University Rankings

The basic approach to these rankings, which has been refined since inception in 2004, is explained on its website (<http://www.timeshighereducation.co.uk>), in the supplement containing the results and in a number of articles (for example, Jobbins, 2005). This is a very different approach from that of ARWU and much of the detail of the methodology is sketchy.

The rankings attempt to reflect a wider perspective of university performance than the ARWU rankings, by combining subjective judgements and objective indicators. The method of introducing judgemental indicators is by the use of international peer review by academics and employers. Large panels are identified and asked to nominate the best universities in their field. These are then conflated in some way to produce the international ratings. Data are also collected from universities on numbers of international staff and international students, and staff:student ratios.

Times Higher has employed a business organisation, QS Quacquarelli Symonds, to derive the data. It appears to be an organiser of international exhibitions of university courses to attract students, and exhibitions of employers to attract university graduates (<http://www.qsnetwork.com/>). This is claimed to give it unprecedented knowledge of both sectors. Not all universities are covered in the rankings, only institutions teaching undergraduates with a broad but not necessarily a full spread of subjects. Institutions in a federal structure are separated, if possible, but no multi-campus institutions are included.

The detail given for the categories of data and the means by which they are built up is shown in Table 3.

TABLE 3 ABOUT HERE

This article does not attempt to assess the merits of including peer review as the basis for producing rankings. It is a widely used method for making assessments when there are no absolute criteria on which to base a valid judgement (Woodhouse, 1994). Like democracy, peer review is often justified on the basis that though it may be imperfect it is better than the alternatives. In this article, attention is directed at the particular approach to incorporating peer assessments into the creation of international institutional rankings.

Compared to the specific, if limited, sources of data used and the publicly available information that is used in the ARWU approach, the way in which the *Times Higher* rankings are compiled, and the results, have to be taken almost entirely on trust. Any potential bias, real or perceived, of having the process conducted by a company that has a commercial interest in selling services to universities and employers is entirely unacknowledged.

Peer review carries a 50 per cent weighting (40 per cent for academics, 10 per cent for employers). However, the process lacks rigour and transparency. The survey is emailed to 190,000 potential academic respondents drawn from two databases: 'World Scientific', based in Singapore, and 'Mardev', focused on Arts and Humanities. It is not specified whether respondents are themselves research active. In 2006, this generated only 1,600 responses, which were combined with those from the previous two years to yield a total of 3,703 responses (Sowter, 2006). A three-year 'latest response' model means that only the most recent response will be taken from any given peer. In 2007, responses grew to 3,069 (Sowter, 2007a), yielding 5,101 in total across the period 2005-7 (Ince, 2007). Even taking into account these increases, if the 190,000 represented an appropriate sample from which to collect judgements, the degree of self-selection, indicated by a 1.6 per cent response rate, introduces an enormous amount of bias. Peers can register judgements on more than one of the *Times Higher*'s five designated subject areas and on more than one geographic region; no criteria are provided on which to base the judgement. It is not clear how far the surveys assess reputation that the respondents may have acquired from other sources and how far they rely on actual contact with universities (Williams, 2005).

For the rankings in 2007, *Times Higher* made a number of changes to the World University ranking methodology (Ince, 2007), specifically,

use of z-scores, change of citation database, reform of the peer assessment.

Perhaps the least controversial was to adjust the raw proportional scores on each of the six components making up rankings, and transform them into z-scores. This is a way of harmonising distributions that is particularly appropriate when scores are to be combined. It is the technique applied in any sophisticated approach to combining examination scores for different subjects that have very different ranges. A criticism of ARWU is that it does not use z-scores in its calculations.

The z-score adjustment before combining scores to achieve the final rankings was not made for the *Times Higher* rankings in 2006. However, this adjustment can be made to the scores to investigate the effect if it had been made at the time. The rank order correlation coefficient between the z-score ranking and the raw score ranking for the top 100 universities in 2006 is 0.98. This indicates a very high measure of overall agreement between the two approaches. Only five institutions leave the top 100 but some institutions would move rankings considerably. The largest upward movement is that of Erasmus University in Rotterdam which moves up 23 places from 92 to 69, and the largest downward movement is Otago University in New Zealand, moving down 26 places from 79 to 105. There is a slightly less high correlation for the top 50 of 0.976 but the changes of rankings are far smaller. The largest upward movement is 11 places by the Indian Institutes of Technology, and the largest downward movement is eight places by the London School of Economics.

The change of database for citations is more controversial for those who have developed great loyalty to ISI. ISI/Thomson Scientific is a commercial operation in the USA that has been developed since the 1960s, but it is also a publisher of journals. This apparent conflict of interest, its US and English language bias, and its limited journal coverage need to be considered. This is not the place for a detailed consideration of the alternatives that have emerged since 2004 – Scopus, produced by Elsevier and Google Scholar – but papers are beginning to appear (Meho, 2007; Bakkalbasi *et al*, 2006; Jasco, 2005) that show different citation results from the three sources in particular fields and there is no clear winner. *Times Higher* has chosen in 2007 to use Scopus; this covers additional journals compared to ISI and is reported to be less biased towards English language publications.

The final major change in methodology for the *Times Higher* World University Rankings in 2007 has been to ‘strengthen measures’ to prevent peers from voting for their own institution (Ince, 2007: 7).

Comparison overview of the ranking results

There are striking differences in the latest results from the two procedures (see Table 4). Table 2 has shown that while both rankings favour the English speaking world, there remain substantial differences between the rankings in the numbers of institutions from individual countries. This in turn affects which region can claim the most ‘world class’ universities. While North America has the most universities in the top 100 in both rankings, it is in second place to Europe in *Times Higher* if the top 200 institutions are considered, but retains first position in ARWU. The Asia Pacific region is a strong contender in the *Times Higher* rankings, but accounts for only 10 per cent of the ARWU top 200 institutions.

At the level of individual institutions, the two rankings have seven of the top 10 institutions in common in 2007, and the top 100 just 56, with a Spearman correlation coefficient of 0.62. In 2006, 133 universities appeared in the top 200 in both rankings; however, four which appeared in the top 50 in ARWU were completely absent from the *Times Higher* (Ioannidis *et al*, 2007). These discrepancies and differences cannot be solely attributed to the fact that *Times Higher* excludes institutions with no undergraduate provision. The two rankings reflect two different approaches to the task. We discuss the position of individual UK universities in more detail in the next section.

TABLE 4 ABOUT HERE

Shanghai Jiao Tong University has made very rapid progress in both rankings, as shown in Table 5. The recognition of this university following the 2003 publicity would appear to have followed very quickly in the *Times Higher* ranking, with its greater reliance on peer review and a contribution from name recognition. The absolute figures are not published to indicate the size of the changes in the data which led to such large changes in the ARWU positions, but it is likely that they were quite small, since bands of 100 plus institutions are shown with tied ranks.

TABLE 5 ABOUT HERE

How well do the two methodologies reflect changes over time?

In the period 2004-07, the number of UK universities in the top 100 has remained the same in the ARWU rankings (11), but in the same period has risen from 13 to 19 in the *Times Higher* rankings, and from 15 to 19 between 2006 and 2007. Most of the UK universities in the top 100 in

2006 have an improved position in 2007 and six new UK entrants to the top 100 have leapt between 33 and 61 places to their new positions. Those on the way down are the London School of Economics, falling from 17 to 59, Queen Mary, University of London, falling from 99 to 149, and the School of African and Oriental Studies, at 70 in 2006 and unlisted in 2007. This brings to the fore an issue that had been observed in earlier rankings (see for example, Marginson, 2007), namely that the rankings produced by *Times Higher* appear much less stable from year to year than those of ARWU. Table 6 demonstrates this, using the example of UK institutions.

TABLE 6 ABOUT HERE

Whilst the ARWU rankings can be seen to have an advantage of stability (at least in the top 100), this can also be seen as created on a quite spurious basis, when the make-up of the individual component scores is analysed. We have shown in Table 1 the extremely lengthy period over which some of the ARWU data are cumulated. The few Nobel prizes in any year will make only a small difference to the score over such a long time period. Similarly, the 20-year period to cumulate high citations leads to a similar effect. The correlation coefficients between the 2006 and 2007 components are Alumni 0.95, Awards 0.95, and highly cited (HiCi) 0.96. When this mass of historical data is also considered for its validity in indicating a current ranking, there is a clear mismatch. Nobel prizes are attributed to institutions of the holder when the prize was awarded. This may be different from both the institution where the work was done and the current workplace of the holder. The case is similar for highly cited authors. ISI allocates this to the institutional designation given in the cited paper and not the current workplace of the author. Thus the ARWU emphasises measures that are historical and whose validity for current quality are in some doubt.

Since the *Times Higher* methodology has been refined each year, this may be responsible for some of the yearly variation. However, as we have shown, the substantial variation between 2006 and 2007 is only explained to a small degree by the changes to z-scores. As citations from ISI/Thomson Scientific used by ARWU are not expressed in the same form as the Scopus ones used by *Times Higher*, it is not possible directly to assess differences generated by the change of database. In view of the relatively small weighting of the citation factor in both methodologies (20 per cent), it is highly unlikely that this factor fully accounts for the variation.

Some of the concern about apparent ‘volatility’ in the *Times Higher* tables (Marginson, 2007) may not take into account the features of rankings. There needs to be a better understanding of the limitations of the decision to express the results in rank order. The most obvious aspect of rankings is that they convey less information than other forms of scale, in that very small changes in the underlying data can produce large changes in the resulting rankings. This is an artefact of rankings and not necessarily a failure of the underlying methodology.

There are further issues concerning the limitations that follow from the limited accuracy of the basic component data from which the rankings are calculated. These are expressed to three significant figures in the case of ARWU and mainly two significant figures for *Times Higher*. When scores are combined and the rankings calculated from the resulting data expressed to three significant figures, there are many tied ranks and changes of rankings following very tiny changes in the underlying data. An alternative form of presentation of the results would be to express them as scores, with appropriate confidence limits. However, it would be much more difficult for casual users of the results to understand the findings. For any serious purpose, however, this should be a strong consideration.

Viewed from an ‘infotainment’ perspective (Bowden, 2000), the variation shown by *Times Higher* may be highly desirable, as the differences generate interest and discussion each year. However, it is scarcely credible that there could be such large changes in institutional quality in such short time periods. This is very important as it runs the risk of discrediting peer assessment as a substantial contributor to international rankings. Thus this form of peer assessment should be developed and made more robust if it is to offer an alternative and complementary approach to identifying world-class institutions.

All of this leaves rather open any consideration of how much yearly variation in the rankings should be expected from year to year and across a period such as five years. If any methodology to produce world rankings was both reliable and valid, this would be a redundant question as there would then be an expectation that the answer would be empirical and emerge from the data. However, both of the methodologies here have been formulated on an atheoretical basis and represent ad hoc compilations of data. Undoubtedly, both would have been created with some expectation that changes would be demonstrated in the rankings year by year. This expectation would be of the changes as a whole and not expectations of particular institutions.

It is clear that the ARWU is so weighted by historical data, that it does not reflect current quality or changes well. Any changes from year to year are at much lower rankings where changes in the basic citation data over a short period have an effect. On the other hand, the substantial changes in rank positions for many institutions each year in *Times Higher* are unlikely to be truly indicative of underlying changes in quality. They are more likely to represent changes in the perceptions of the peer reviewers and chance variation in their ratings, reflecting the substantial bias arising from a response rate of less than 2 per cent in any one year.

Changes to indicator systems should be made infrequently as each change breaks the continuity of methodology that provides the basis for a valid historical comparison of changes over time. Against that, however, in the early formative stages of a new indicator system it is important to modify the working of the system so that it functions as intended. In this way the system is modified to deliver longer term comparability of the results.

Conclusions and recommendations

Previous academic investigations of the methodologies used in university league tables have predominantly focused on national rankings. Typically, concerns are raised regarding reliability and statistical validity. This article contributes to and broadens the literature by exploring the two principal methodologies for producing world class university rankings, and seeks to promote further discussion and research in this area. Issues of reliability, validity and utility remain. Together with the problems created by the need for, and limitations of, internationally comparable data, such issues are inevitably more complex to resolve. However, since both *Times Higher* and ARWU rankings are still at a formative stage, it is urgent that revision and reform takes place, before the compilation processes become institutionalised.

The differences between the *Times Higher* and ARWU approaches could be viewed simply as indicating differences between the research performance of universities and a more balanced view of quality performance. This would make a virtue of the application of two different methodologies. However, both have weaknesses that need to be addressed before they could be accepted as valid indicators of the facets of world class universities. Currently, the production of data is unregulated and there are limited or varying levels of transparency in both the data collection and data analysis processes. In the *Times Higher* ranking, the directions given to peer reviewers are imprecise. It was not

until the fourth year of operation that measures were 'strengthened' to prevent peers voting for their own institutions (Ince, 2007: 7), suggesting a lack of scrutiny of their feedback. The role and weighting of peer review clearly needs careful consideration. Well-established prestigious universities generate a 'halo' effect, whereby existing reputation continues to be recycled even if relative performance has changed (Marginson, 2007).

In addition to the need for validity in the rankings by external users, if the rankings are to provide an incentive for institutions to improve their position, then they need to be seen as valid and stable. There would be little point in a university management team seeking to improve their university's ranking if they did not regard the variables making up the rankings as indicators of underlying quality processes. Nor would they be wise to do so, if the basis for calculating the rankings changed from year to year or if historical performance is substantially advantaged over current work. The large movements in the rank positions of certain universities from year to year in *Times Higher* do not inspire confidence in the utility of rankings as a basis for developing direction and strategy.

Further refinements and improvements to the presentation of findings could be implemented, once reforms to the existing methodologies have been prioritised. Bowden (2000) proposes that for league tables to increase their utility to users beyond mere 'infotainment' value, a web-based 'one-stop shop' could be created, where users could define their own searches based on their own priorities. Steps towards this approach have been taken in Germany whereby users can weight criteria according to preference, obviating the application of arbitrary weightings to combine the criteria into a single result (Federkeil, 2002). The process can be viewed at <http://www.daad.de/deutschland/hochschulen/hochschulranking/06543.en.html>. A similar system is likely to be introduced in a number of other European countries, including Switzerland, Austria, the Netherlands and Belgium (Marginson, 2007; Usher and Savino, 2006). The choice of variables used will influence a university's position in a league table (Yorke, 1997; Bowden, 2000) and so, if users can select and weight their own criteria, utility will be enhanced. We have found that whilst it is possible to manipulate the data published on the *Times Higher* website using statistical software; the absence of absolute figures such as, for example, the number of international students and staff, greatly limits transparency and the calculations that can be made. It would also be advisable for the component scores to be given to the appropriate degree of accuracy of the underlying data.

Limitations of the German approach are that data relate to individual academic disciplines rather than whole institutions, and are only collected nationally. The ranking of whole institutions assumes that all sections are of equal quality and yet it is knowledge of the quality of specialist areas that stakeholders, including students, businesses and research councils, typically require. Such discipline-based and specialist rankings offer another route for the further development of ‘world class’ rankings and these are already well developed, for example, in the field of management education, with respect to business schools (Wedlin, 2006).

Whilst Bowden (2000) points to the ‘infotainment’ value of rankings, the salience of league tables has substantially increased in the last decade. Rankings and league tables have become ‘part of the higher education landscape’ and their impact now extends beyond student choice to ‘institutions’ reputations and ... the behaviour of academics, business and would-be benefactors’ (Eastwood, cited in HEFCE 2007). Moreover, the pace of globalisation continues to accelerate, fuelling interest in international comparisons and reinforcing the need for international competitiveness. Since institutional position in league tables increasingly matters, further research is needed regarding how methodologies can be improved, to increase the validity and reliability of world university rankings.

Note

¹ Times Higher Education Supplement was renamed Times Higher Education in January 2008. Rankings discussed in the article were published under the former title.

Address for correspondence

Dr Christine Parsons, School of Business and Management, Buckinghamshire New University, Chalfont Campus, Goreland Lane, Chalfont St Giles, Bucks, HP8 4AD. Email: Chris.Parsons@bucks.ac.uk

References

Bakkalbasi, N, Bauer, K, Glover, J and Wang, L (2006) ‘Three options for citation tracking: Google Scholar, Scopus and Web of Science’, *Biomedical Digital Libraries* 2006, 3:7, available: <http://www.biomedilib.com/content/3/1/7> (access date: 4 March 2008)

- Barber, M (1998) 'Creating a world class education service', text of a speech delivered at The North of England Education Conference, Bradford 5-7 January 1998, available: <http://www.leeds.ac.uk/educol/documents/000000442.doc> (access date: 4 March 2008)
- Barber, M and Sebba, J (1999) 'Reflections on progress towards a world class education system', *Cambridge Journal of Education*, 29 (2): 183-93
- Berry, C (1999) 'University league tables: artefacts and inconsistencies in individual rankings', *Higher Education Review*, 31 (2): 3-10
- Bowden, R (2000) 'Fantasy higher education: university and college league tables', *Quality in Higher Education*, 6 (1): 41-60
- Center for Science and Technology Studies (2008) *The Leiden Ranking*. Available: <http://www.cwts.nl/cwts/LeidenRankingWebSite.html> (access date: 4 March 2008)
- CHE (2006) *Berlin Principles on Ranking of Higher Education Institutions*. Centrum für Hochschulentwicklung, Hannover, available: http://www.che.de/downloads/Berlin_Principles_IREG_534.pdf (access date: 4 March 2008)
- Clarke, M (2002) 'Some guidelines for academic quality rankings', *Higher Education in Europe*, XXVII (4): 443-59
- Department for Education and Science (2006) 'Delivering a World Class System – David Bell', *Press Notice* 19 July 2006, Department for Children, Schools and Families, available: http://www.dfes.gov.uk/pns/DisplayPN.cgi?pn_id=2006_0106 (Access date: 4 March 2008)
- Eccles, C (2002) 'The use of university rankings in the United Kingdom', *Higher Education in Europe*, XXVII (4): 423-32
- Economist, The* (2005) 'The best is yet to come', *The Economist*, Vol 376, Issue 8443: 20
- Federkeil, G (2002) 'Some aspects of ranking methodology – the CHE-Ranking in German universities', *Higher Education in Europe*, XXVII (4): 389-97
- Florian, R (2006) 'Irreproducibility of the results of the Shanghai academic ranking of world universities', available: http://www.ad-astra.ro/journal/8/florian_shanghai_irreproducibility.pdf (access date: 4 March 2008)
- Florian, R (2007) 'Irreproducibility of the results of the Shanghai academic ranking of world universities', *Scientometrics*, 72 (1): 25-32
- HEFCE (2007) 'Research commissioned to throw light on university league tables and their impact on institutional behaviour', *Press Release*, 24 July, available: <http://www.hefce.ac.uk/news/hefce/2007/leagtab.htm> (access date: 4 March 2008)
- Ince, M (2007) 'Fine tuning reveals distinctions', *Times Higher Education Supplement*. World University Rankings Supplement, 9 November 2007: 7
- Ioannidis, J P A, Patsopoulos, N A, Kavvoura, F K, Tatsioni, A, Evangelou, E, Kouri, A, Contopoulos-Ionannadis, D G and Liberopoulos, G L (2007) 'International ranking systems for universities and institutions: a critical

- appraisal', *BMC Medicine*, available: <http://www.biomedcentral.com/1741-7015/5/30> (access date: 4 March 2008)
- Jasco, P (2005) 'As we may search – Comparison of major features of the *Web of Science*, *Scopus*, and *Google Scholar* citation-based and citations-enhanced databases', *Current Science*, 89 (9): 1537-47
- Jobbins, D (2002) '*The Times/The Times Higher Education Supplement* – League tables in Britain: An insider's view', *Higher Education in Europe*, XXVII (4): 383-88
- Jobbins, D (2005) 'Moving to a global stage: a media view', *Higher Education in Europe*, 30 (2): 137-45
- Liu, N C and Cheng, Y (2005) 'The academic ranking of world universities', *Higher Education in Europe*, 30 (2): 127-36
- Marginson, S (2007) '*Global university comparisons: the second stage*', paper presented at the Griffith University/IRU Symposium, 'International Trends in University Rankings and Classifications', 12 February 2007
- Meho, L I (2007) 'The rise and rise of citation analysis', *Physics World*, 20 (1): 32-36
- Merisotis, J (2002) 'On the ranking of higher education institutions', *Higher Education in Europe*, XXVII (4): 361-63
- Merisotis, J and Sadlak, J (2005) 'Higher education rankings: evolution, acceptance, and dialogue', *Higher Education in Europe*, 30 (2): 97-101
- Moed, H F (2002) 'The impact-factors debate: the ISI's uses and limits: towards a critical informative, accurate and policy-relevant bibliometrics', *Nature*, 415: 731-32
- Parsons, C and Fidler, B (2004) 'De-internationalisation in higher education: the case of UK plc', *Higher Education Review* 36 (3): 13-32
- Provan, D and Abercromby, K (2000) '*University League Tables and Rankings: A critical analysis*', CHEMS Paper 30, available: <http://www.acu.ac.uk/chems/onlinepublications/976798333.pdf> (access date: 4 March 2008)
- Sowter, B (2006) THES-QS World University Rankings – Methodology, available: http://www.topuniversities.com/worlduniversityrankings/university_ranking_s_news/article/thes_qs_world_university_rankings_methodology (access date: 4 March 2008)
- Sowter, B (2007a) Selection of the initial list, available: http://www.topuniversities.com/worlduniversityrankings/university_ranking_s_news/article/selection_of_the_initial_list (access date: 4 March 2008)
- Sowter, B (2007b) Methodology: The Peer Review, available: http://www.topuniversities.com/worlduniversityrankings/university_ranking_s_news/article/methodology_the_peer_review (access date: 4 March 2008)
- Sowter, B (2007c) 2007 Peer Review Response Analysis, available: http://www.topuniversities.com/worlduniversityrankings/university_ranking_s_news/article/2007_peer_review_response_analysis (access date: 4 March 2008)

- Universities UK (2006) *The Economic Impact of UK Higher Education Institutions*, available: <http://bookshop.universitiesuk.ac.uk/downloads/economicimpact3.pdf> (access date: 4 March 2008)
- Usher, A and Savino, M (2006) *A world of difference: A global survey of university league tables*, available: <http://www.educationalpolicy.org> (access date: 4 March 2008)
- van Raan, A F J (2005) 'Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods', *Scientometrics*, 62(1): 133-43
- Wedlin, L (2006) *Ranking University Business Schools: Forming Fields, Identities and Boundaries in International Management Education*, Edward Elgar: London
- Westholm, G, Tchatchoua, B and Tindemans, P (2004) 'The great global R and D divide', *Multinational Monitor*, July/August 2004: 24-28
- Williams, R (2005) 'Broadening the criteria: lessons from the Australian rankings', paper presented at the First International Conference on World Class Universities, Shanghai Jiao Tong University, 16-18 June 2005, available: <http://melbourneinstitute.com/research/micro/downloads/Educ%20Page/ShanghaiWCU-1.pdf> (access date: 4 March 2008)
- Woodhouse, D (1994) 'International peer review in Hong Kong', *Higher Education Review* 26 (3): 19-26
- Wooldrige, A (2005) 'The brains business', *The Economist*, Vol 376, Issue 8443: 3-4
- Yorke, M (1997) 'A good league guide?', *Quality Assurance in Education*, 5 (2): 61-72

TABLE 1
Compilation of Academic Ranking of World Universities in 2007

Category	Quality of education	Quality of staff		Research output		Size
Detail	Alumni winning medals and prizes	Staff winning medals and prizes	Highly cited researchers	Articles in <i>Nature and Science</i>	Articles cited in SCIE and SSCI	Academic performance with respect to size
Weighting	10%	20%	20%	20%	20%	10%
Period	Cumulated since 1901 (weighted by elapsed time since graduation)	Cumulated since 1911 and working at university when prize awarded	Cumulated over a rolling 20-year period. The credit goes to the institutional designation of the author at the time of publication	Previous five years (multiple author shares: 50% to first author)	Articles published in SCI and SSCI in the previous year. SSCI articles are weighted 2. Only publications of an article type are included	The sum of the weighted scores of the remaining indicators is divided by the number of full-time equivalent staff (if it can be obtained)

Source: Derived from <http://www.arwu.org/rank/2007/ARWU2007Methodology.htm>

TABLE 2
*Dominance of the English-speaking world in the top 100 universities
 by number of institutions*

	ARWU			<i>Times Higher</i>		
	2007	2006	2005	2007	2006	2005
UK	11	11	11	19	15	13
US	54	54	53	37	33	31
Canada	4	4	4	6	3	3
Australia	2	2	2	8	7	12
Total	71	71	70	70	58	59

TABLE 3
Compilation of *Times Higher* World University Rankings in 2007

Criteria	Research Quality		Graduate Employability	International Outlook		Teaching Quality
Category	Peer review	Citations per academic	Recruiter review	Int'l faculty	Int'l students	Faculty per student
Detail	Academics	Citations per staff member	Employers frequently recruiting internationally or large numbers nationally	Percent of int'l staff	Percent of int'l students	Staff numbers/ student numbers
Weighting	40%	20%	10%	5%	5%	20%
Selection	Two databases, 'World Scientific' and 'Mardev', plus previous respondents	Scopus	Quacquarelli Symonds (QS) or nomination by universities	Not specified	Not specified	Not specified
Method	Asked to name up to 30 universities in their subject area and geographical area	Citations compiled from Scopus. Staff numbers – QS obtains data	Asked to nominate universities from which they like to hire graduates	QS obtains data from national bodies or directly from	QS obtains data from national bodies or directly from universities	QS obtains data from national bodies or directly from universities

	of their expertise, from a list of 540	from national bodies or directly from universities		universities		
Period	Most recent response from any given peer in last three years	Cumulated over previous five years (2002-6)	Not specified	Not specified	Not specified	Not specified

Source: Derived from Ince (2007), Sowter (2007a, 2007b, 2007c, 2006)

TABLE 4
Performance by major world regions in the top 200, by number of institutions

	Rank	ARWU			<i>Times Higher</i>		
		2007	2006	2005	2007	2006	2005
Asia-Pacific	1-100	8	8	7	22	22	28
	101-200	13	12	12	19	26	25
	total	21	20	19	41	48	53
Europe	1-100	33	33	34	35	40	35
	101-200	46	44	44	51	44	46
	total	79	77	78	86	84	81
North America (US and Canada)	1-100	58	58	57	43	36	34
	101-200	36	37	41	24	26	28
	total	94	95	98	67	62	62
Rest of World	1-100	2	2	2	0	2	3
	101-200	6	6	5	6	4	2
	total	8	8	7	6	6	5

Notes: Rest of World includes South America, Israel, Russia, South Africa

Totals for all regions in the top 100 or 101-200 may exceed 100 due to joint positions

TABLE 5
Position of Shanghai Jiao Tong University in ARWU and Times Higher rankings

	ARWU	<i>Times Higher</i>
2004	404-502	Not in top 200
2005	301-400	169
2006	201-300	179
2007	203-304	163

TABLE 6
Rank positions of UK institutions in the *Times Higher* top 100 in 2004
and subsequent performance

	2004	2005	2006	2006(z)	2007
Oxford	5	4	3	5	2
Cambridge	6	3	2	2	2
London School of Economics	11	11	17	25	59
Imperial College London	14	13	9	10	5
University College London	34	28	25	29	9
Manchester	43	35	40	43	30
School of African and Oriental Studies	44	103	70	83	not listed in top 200
Edinburgh	48	30	33	33	23
Sussex	58	100	105	123	121
St Andrews	70	136	109	112	76
Warwick	80	77	73	73	57
Bristol	91	49	64	62	37
King's College London	96	73	46	49	24
Queen Mary, University of London	100	112	99	122	149

(z) = adjusted for z score

