Methodology, Meaning and Usefulness of Rankings

Ross Williams
Melbourne Institute of Applied Economic and Social Research
University of Melbourne
www.melbourneinstitute.com

****
Paper prepared for AFR Higher Education Conference, Sydney, 13-14 March 2008

> *"I imagine that all university heads broadly share my own view of these [league] tables. They are terrific and unquestioned when you score well and better than last time. They are fatally flawed and fundamentally unfair when you move in the opposite direction."*
>
> Howard Davies, Director, *London School of Economics*

Why is Ranking Universities such a Growth Industry?

For one hundred years after the establishment of Australia's first university in Sydney in 1850, higher education was a colonial/state based system. Students were residents of the state, graduates were employed by locally-based firms, and the operations of universities were financed by state government monies, local private benefactors and student fees. In the absence of competitive pressures there was little need for benchmarking, KPIs or ranking. Benchmarking was largely confined to that implicitly undertaken by professors hired from Britain and Ireland who had maintained their overseas links.

The establishment of second and more universities in most states created competition for students, staff and funds, but it took some time for competitive behaviour to emerge. In the beginning, the demographic pressures on the old universities meant that they were keen to see new institutions develop quickly.

The relative standing of a university within Australia became important as the employing firms became national, as students began to explore interstate options, and as federal government funding replaced state government funding. The development of *The Good Universities Guide* was a response to these new demands.

Globalisation, assisted by deregulation, has created the demand for international rankings. The demand originates from a range of stakeholders: students, employers,

supranational institutions, scholars, funding agencies, and governments. In addition, there is public interest in rankings for their own sake, whether it be the world's most liveable city or an international ranking of the quality of financial newspapers. At the same time as this expansion in demand, developments in technology, most noticeably the world wide web, facilitated the supply of information to meet demand.

In Australia, over the last two decades the lifting of restrictions on the enrolment of fee-paying international students combined with a freezing of funding for government subsidised students has made universities heavily dependent for growth on income from international students. These students, being located far from the supplying source, need independent advice on which to base their choice of university. International rankings supply some of this need.

I will concentrate on approaches that include as output the ranking of institutions as a whole. This is for two reasons: it is a fast growing area of research and the influence of these rankings is new and important.

Rankings Matter

International rankings are influencing decision making within institutions and even affecting national systems of education. France and Germany suffer in international rankings because quality research performance is spread over many institutions; these are often specialized and a significant number are not universities. The rankings have provided much motivation for the current policy in these countries of linking or consolidating institutions to establish larger entities.

Salmi and Saroyan (2007) report that in some countries authorities restrict scholarships for studies abroad to students admitted to highly ranked institutions; donor agencies and foundations also look at international rankings to inform their decision making.

Rankings matter to CEOs of universities, as is documented by Ellen Hazelhorn (2007) in her international survey of leaders and senior university administrators. Fifty-six per cent indicated that their institution had a formal internal mechanism for reviewing their rank. Respondents also indicated that league tables played an important role in deciding on international collaborations.

The effect of league tables on student choice is more complex. The consensus seems to be that for rankings targeted at school leavers their direct influence is greatest for high achievers. It seems it is overall reputation which matters for undergraduate student choice and rankings are one factor feeding in to that perception. However, Marginson (2007) notes that market research and anecdotal evidence from educational agents indicate that the international rankings published by Shanghai Jiao Tong University are feeding directly into student choice at all levels, even though the rankings are based solely on research performance. Increasingly the international rankings are being interpreted as measuring the international standing of an institution.

Positives for Universities

An obvious marketing benefit accrues to a university which is highly ranked in a study. But as with all forms of external appraisal there are a number of more indirect benefits. Rankings provide an incentive for better data collection within institutions, they can expose pockets of institutional weakness and confirm areas of strength, and they are useful for benchmarking against like institutions. Rankings encourage institutions to re-examine mission statements.

For the university system as whole, poor performance can be used to prod governments into action.

Ranking Methodologies

The Berlin principles (see Appendix) set out some of the guidelines on what we should measure: outputs rather than inputs, be transparent, use verifiable data and recognize diversity of missions. What criteria should be used to rate or rank a university's performance? Candidates include research output and its influence, the quality of teaching and research training, and contribution to the formulation and implementation of national policy. Different groups of stakeholders will have different interests; this implies that ratings should be undertaken separately for the different attributes before they are combined into a single measure.

The methods used to measure *research performance* in universities form a spectrum: from a survey of peers at one end to the use of quantitative measures of output only, such as publications and citations, at the other end. In the middle of the spectrum lies evaluation obtained by providing peers with representative publications and detailed quantitative information.

In evaluating the *quality of teaching* the methodology spectrum ranges from surveys of students and employers to quantitative measures such as progression rates, job placements and starting salaries of graduates. There is, however, much less agreement about the appropriate quantitative performance measures for teaching and learning than there is for research.

A university should be ranked highly if it is very good at what it does. This implies that in order to recognize institutional differences whole-of-institution rankings should either be conducted separately for different types of institutions or be obtained by aggregation of rankings at a sub-institutional level. In Australia, where all universities offer Ph.D. programs and have similar mission statements, categorization into types, such as is done by the Carnegie Foundation in the U.S. and Maclean's in Canada, is less useful.

We are then left with the option of ranking by sub-institutional unit, most commonly discipline, and aggregating. Rankings by discipline are of value in themselves, especially to academics and postgraduate students. The downside of the aggregating up approach is

that it requires much more detailed information, including measures of the importance of each discipline (or some other sub-institutional unit) in the university. Not to allow for scope will bias overall rankings in favour of institutions which have disciplines where the number of publications produced per academic is large, such as in medicine. To illustrate, over the last ten years 22 per cent of Australian publications in Thomson ESI were in clinical medicine. How much of the desire by universities to have a medical school is driven by the knowledge that this is a sure way of providing a large boost to research output with resultant dollar flows from research funding formula? In our work at the Melbourne Institute (Williams, 2007) we found that allowing for scope improves the ranking of the more technologically oriented universities and ANU.

Disaggregation can be at various levels: research groups, disciplines, departments and faculties. It is inevitable that international rankings will be at the discipline or institutional level, especially if the rankings are based on publicly available information. Only at this level can the independent ranker sitting at her lap top obtain data on a consistent basis. In general, departments and faculties do not translate well across national frontiers: organizational structures differ too much and departmental affiliations of authors are not always known. While there are international rankings of MBA programs, these require most information to be collected from institutions, which raises issues of consistency. While national research funding agencies may rate research groups, this requires too much detailed information for international comparisons.

The proposed federal government Excellence in Research for Australia (ERA) initiative proposes to use discipline as the sub-institutional unit for measuring research. It will have the added benefit of encouraging universities to look at their internal departmental structures.

Categories of Data

There are three categories of data: survey data, data supplied by universities, and data from third party sources such as government agencies and private sector citation data banks. The weakness of collecting data direct from universities without external moderation is that definitions may vary across institutions (for example, how part-time students are counted or whether honorary staff are included in staff numbers) and the data are subject to game playing.

Data deficiencies exist in many areas and conceptual differences exist, especially in the evaluation of teaching. This is not a reason for refusing to take rankings seriously, rather it should act as spur for people to come up with better measures and collect additional data.

To be useful, survey data must meet statistical standards with respect to choice of population, questionnaire design and response rate. When using surveys for evaluating standing, the validity of the results depends critically on the knowledge possessed by the respondent. Respondents with little direct knowledge of an institution will be reflecting

reputation as much as current performance. The group that will be most informed about research performance are scholars in the same discipline. At the Melbourne Institute (Williams and Van Dyke, 2006), we compared such responses with quantitative measures of research performance for seven discipline groups in Australian universities; the rankings were broadly similar by the two methods.

Quantitative Measures of Research Performance

The obvious point to make is that the nature of research output varies greatly across disciplines. Most rankings give particular importance to publications in refereed journals, which can impart a discipline bias to rankings. Table 1 contains estimates of the percentage of weighted output of the G08 universities in selected disciplines in 2003-04. The percentages range from 95 per cent for medicine and chemistry to 25 per cent in computer science and electrical engineering (where conference papers dominate). The degree to which these differences matter in a straight count of articles published depends on the extent to which, for a given discipline, the ratio of articles to other publications differs across institutions. In any event, for a national ranking it is feasible to obtain discipline-based data on each form of output. The greatest difficulties lie in areas which are not represented in table 1, particularly the creative and performing arts. New measures need to be developed here, preferably by those in the disciplines.

Table 1: Estimates of percentage share of weighted output in the form of journal articles, Go8 universities, 2003-2004.

Discipline	% output articles
Chemistry	95
Medicine	95
Mathematics/Statistics	85
Accounting	80
Physics	80
Behavioural Science	80
Finance	75
Earth Science	75
Chemical Engineering	70
Law	70
Economics	65
Philosophy	60
Education	50
Civil Engineering	40
English	40
History	40
Political Science	35
Computer Science/Elect Eng	25

Citation counts play an important role in most ranking schemes; either directly through citation counts or indirectly by using them to define high quality journals. Again this methodology is most useful in the sciences where impact is more immediately clear and publication lags are short. Publication delays are a major problem in other areas: in accounting and economics, for example, citations in the top journals are not common under three or four years from the time of submission of the original article, although these lags are diminishing with greater access through the web to working papers and forthcoming articles.

Measures of Learning and Teaching

Few measures of performance in learning and teaching are available on a comparable basis internationally. Staff-student ratios have been a traditional measure of resources devoted to teaching but are becoming less useful with technological change. Similar remarks apply to other input measures such as library holdings. Resources devoted to teaching and research training is probably the best input measure but is one which requires some standardization of budgets to make it operational.

Output measures such as progression of undergraduate students to higher degrees and placement of PhD graduates have merit as indicators of international academic standing. The former measure is an important one for U.S. liberal arts colleges which act as feeders to graduate programs. Other output measures used in national rankings are employment rates and remuneration on graduation but these can be seriously affected by regional factors.

Surveys of current and past students and of employers are useful provided they satisfy statistical standards in design and responses. Write-in evaluations of teachers by students, such as is available in the US at www.RateMyProfessors.com, provide some information to prospective students in a course but are not suitable for cross-institutional comparisons.

International comparisons of the quality of graduates will probably require initiatives from international agencies such as the OECD or World Bank. The OECD is well placed to undertake this work because of its experience in measuring student performance at the school level.

One fundamental difficulty will always remain: the quality of graduates will be reflected in their development over several decades, but this reveals little about the current quality of teaching.

Presentation of Results

Some critics object to the calculation of whole-of-institution rankings on the grounds that they do not adequately reflect different institutional characteristics. A few evaluators,

most notably CHE in Germany, do not give overall rankings. The validity of the criticism depends on the methodology used; aggregating up performance at the discipline level goes a long way to meeting this objection. In practice, there is a large market for a simple rating or ranking of an institution which can be obtained without additional calculation by the user. It is important however that rankers give details of the ratings/rankings on different attributes and be quite explicit about the weights so that users can use alternative forms of aggregation.

Unless evaluation is entirely by survey, an overall ranking requires the use of weights. In order to provide some objectivity in choosing weights we surveyed CEOs of the world's leading research universities and all Australian and New Zealand universities (Williams and Van Dyke, 2007). The average response gave a weight of nearly one-half to research and research training. The exact weights were: 40 per cent on quality of staff as measured by research performance, 16 per cent on quality of graduate programs, 14 per cent on quality of undergraduate programs, 11 per cent on each of quality of undergraduate intake and resource levels, and 8 per cent on peer opinion.

Rankings exaggerate small differences in performance scores and for this reason some prefer banding results as is done in the allocation of the Learning and Teaching Performance Fund. The downside of banding is that it usually exaggerates differences between the lowest ranked institution in one grade and the top institution in the grade below. Rating performance on a scale of say 1 to 5 is another banding technique.

The Rankers

National rankings were originally supplied by newspapers and journals, particularly *US News and World Report* in the US, *The Times* ranking in the UK, *Maclean's* in Canada and *CHE&/Stern/Die Zeit* in Germany. Ranking of national institutions and international comparisons have in the past few years spread to many countries including those with relatively weak higher education sectors (see Van Dyke (2005), and Usher and Savino (2007) for details). Much of the expansion in country rankings has been undertaken by independent research groups and government agencies. The expansion has been driven by two motives: a desire to see how institutions in a country rank internationally, especially in research, and an interest in methodology. In table 2 below I reproduce a list of countries where rankings of institutions are publicly available. Asian countries are well represented.

The nature of the rankings reflects the interests of the suppliers. Commercial newspapers and magazines concentrate on measuring the quality of teaching and learning at the undergraduate level because of the large market for this information. Governments on the other hand tend to be most interested in university research performance as they see this feeding into national economic performance.

Table 2: Ranking Systems 2007

Region	National and international ranking systems
Europe	Germany (B&C, C), Italy (C), Netherlands (A), Portugal (C), Spain (B, C, IC), Sweden (C), Switzerland (B&C), United Kingdom (A, B, IC)
Eastern Europe and Central Asia	Kazakhstan (A, B), Poland (C), Slovakia (B), Romania (B&C) Russia (B), Ukraine (B&C)
Asia and Pacific	Australia (B), China (B, C, IB), Hong Kong (C), India (C&D), Japan (B, C), Korea (A), Malaysia (A), Pakistan (A), New Zealand (A), Thailand (A). Taiwan (IA, B)
Latin America and Caribbean	Argentina (D), Brazil (A), Chile (C, D), Mexico (B), Peru (B)
Africa	Nigeria (A), Tunisia (A)
North America	Canada (B, C, B&C), United States (C, IC)

Key: A= government agency; B= independent organization/university; C=newspaper/magazine/media; D= accrediting agency; I = international ranking;
Source: Updated version of table 2 in J. Salmi and A. Saroyan (2007).

Researchers in European universities and research institutes have provided much of the intellectual basis for methodological advances. Professor Van Raan and his team at Leiden University undertake very careful detailed measures of research performance. In Germany, there is the independent Centre for Higher Education (CHE). The International Ranking Expert Group meets annually, with support from UNESCO; at the 2006 meetings they drew up the 'Berlin Principles' of good ranking procedures.

International Rankers

SJTU

The first world ranking of universities by Professor Lui at the Institute for Higher Education at Shanghai Jiao Tong University (SJTU) in 2003 was designed to benchmark Chinese universities. The emphasis was on research performance in science and technology because it is an area in which China wishes to be strong; this emphasis also reflects the characteristics of the available data bases. The bias towards English language publications in the data base was not a concern – such publications are recognized as providing a pathway to influence. The SJTU annual rankings, supplemented by

discipline rankings since 2007, remain the most quoted and respected international rankings. They have weaknesses, but as with the QWERTY keyboard it is hard to replace first movers even if the developers had a specific purpose in mind.

The criteria used all relate to research: Nobel prizes in sciences and economics and Field medal winners (20% weight if on staff of institution when awarded, 10% if alumni), high citation researchers (20%); articles in Thomson Scientific journals in science and social science (20%), articles in *Science* and *Nature* (20%), research performance per head of academic staff (10%).

The SJTU index performs well against the Berlin principles. The index measures outputs, it is transparent, and data are verifiable. There have been limited changes in the attributes and weights used: compared with the original 2003 index, the Field medal has been added, the weight on performance per head has been reduced from 20 to 10 per cent with the weight transferred to a new category of number of alumni who have been awarded the Nobel prize or Field medal. In addition, publications in the social sciences are now given a double weight to reflect the lower publication rates in these disciplines.

The SJTU discipline rankings are in six areas: natural sciences and mathematics; engineering/technologies and computer science; engineering and IT; biomedicine, life and agricultural sciences; clinical medicine and pharmacy; and social sciences. The attributes included are similar to those used in the whole-of-institution rankings except that there is no measure of size adjusted performance and an additional quality measure of publications is included (publication in top 20 per cent of journals as measured by citations per paper).

THES-QS

The other main international ranking is published by *The Times Higher Education Supplement* in associated with QS career and education consultants (THES-QS). In this index 50 per cent of the weight is given to surveys of academics (40 per cent) and employers (10 per cent). Internationalisation is measured by the proportion of students and staff that are foreign (each with a weight in the index of 5 per cent). Staff-student ratios are used as a proxy for teaching quality (20 per cent) and research citations per head are given a weight of 20 per cent.

The THES-QS methodology is less transparent than SJTU although it is improving. By surveying academics and employers, the THES-QS World University Rankings cover more than research. But the surveys suffer from a number of limitations: the response rates are low at around 1 per cent (Sowter, 2007) and the respondents not representative. For example, in 2007 peer respondents from New Zealand were the fourth highest in number and together with Australia made up 7.5 per cent of the total compared with 16.5 per cent from the USA (www.topuniversities.com).

In the quantitative data there is no control for the quality of international students or staff; there is also scope for game playing by institutions when providing data on numbers of international staff and students.

The THES-QS disciplines rankings, based on peer review, are in five areas: Arts & Humanities, Engineering & IT, Life Sciences and Biomedicine, Natural Sciences and Social Sciences.

The THES-QS rankings show great fluctuations from year to year. This is not unexpected when 50 per cent of the weight is for survey results based on very low response rates. There is just too much noise in these data. In addition, in 2007 two important changes were introduced. First, the source of the citations data was changed from Thomson ESI to Scopus, the data bank developed by Elsevier Publishing. It is unclear how this changes the rankings, but it probably works to the advantage of European universities.

Second, and more significantly, instead of measuring performance relative to the best institution it is now measured by looking at standard deviations about the mean value (z-transforms are used on all variables). The disadvantage of this measure is that individual scores depend on the number and nature of the universities included. For the peer surveys the responses will be bunched, with top institutions receiving high scores but with many lesser institutions scoring quite poorly. The effect when standardized is to reduce the range within the top institutions – in 2007 twenty-one institutions scored 100 whereas in 2006 there were only four universities that scored above 90. In effect the peer ranking, that has a weight of 40 per cent, is much less important for the top institutions in the 2007 rankings than in early years.

Where do Australian Institutions Rank?

In the 2007 SJTU rankings no Australian university is in the top 50 and only two, ANU and Melbourne, are in the top 100. A similar result occurs in the new rankings produced by the Higher Education Evaluation & Accreditation Council of Taiwan (www.heeact.edu.tw), except that Sydney replaces ANU in the 51 to 100 group. (The Taiwan methodology is similar to that of SJTU but Nobel prize winners are excluded.) If we look only at publications (other than *Nature* and *Science*) in the SJTU rankings, two Australian universities are in the top 50 and four in the top 100.

The SJTU rankings essentially measure research standing in the sciences and social sciences as measured by journal articles. The journal coverage in these areas is adequate, especially following the inclusion of more Australian journals in the last two years. However, the SJTU rankings ignore most Australian output in law and the humanities.

At the Melbourne Institute we have ranked Australian universities using not only research performance but measures of teaching and research training. We also try to capture more

general measures of standing through attributes such as membership of learned academies. Our rankings are very similar to those produced by SJTU. The congruence of the SJTU and Melbourne Institute results has two main causes: (i) different quantitative measures of research performance are highly correlated and (ii) the variability in research performance across institutions is much greater than the variability in available measures of teaching performance so that research performance tends to dominate.

The SJTU discipline results released in February 2008 show three entries for Australian universities in the top 50: ANU in Science, and ANU and UWA in Life and Agricultural Sciences. In twelve cases, covering six universities, Australian disciplines are ranked in the top 100 in the world. Selected SJTU country rankings are given in table 3. The Melbourne Institute findings (Williams and Van Dyke, 2006) are broadly similarly in areas that can be compared and place three Australian universities in the top 100 in the humanities.

Table 3: SJTU country rankings 2007-08, number of institutions in top 100

Country	Science	Engineering	Life Sciences	Medicine	Social Science	Overall
USA	59	49	62	61	77	54
UK	9	7	11	12	11	11
Germany	7	1	6	6	0	6
Japan	7	7	3	2	0	6
Canada	2	6	5	6	7	4
France	5	2	1	1	0	4
Sweden	2	3	2	2	0	4
Switzerland	3	2	4	2	0	3
Netherlands	1	3	2	5	4	2
Australia	1	3	4	3	1	2

Source: http://ed.sjtu.edu.cn/ranking.htm

The positions of Australian universities in the THES-QS rankings are biased upwards owing to the sample bias in the surveys and the inclusion of international student and staff numbers without quality control. To illustrate, on the transformed peer survey results, Harvard scores 100.0, ANU 99.8 and Melbourne 99.6. Five Australian universities are listed in the top 50 in the world, but none appear in the top 100 on the only quantitative research criterion used, namely, citations per academic staff member.

How Should Australian Universities Respond to Rankings?

Rankings are here to stay and will continue to gain in importance. Australian universities need to respond in two sets of ways: work to improve outcomes in the existing rankings and encourage new types of rankings.

Existing rankings are biased towards publications and citations in journals. It is therefore important that Australian journals are well represented in the data bases; this is a responsibility of academic editors in conjunction with publishing houses. Where other forms of publication are important, such as books in the humanities and refereed conference papers in engineering, the disciplines need to come up with robust measures that can be suggested to rankers. Electronic downloads of published and working papers are being included in some rankings, but the methodology needs improvement.

The downside of the current international rankings is that they tend to enshrine the existing homogeneity of mission statements amongst Australian universities. Realistically, only a handful of Australian universities can aspire to be in the top 100 as ranked by SJTU although a larger number can aspire to be in the top 100 in selected disciplines. Discipline rankings should be supported as they encourage vertical specialization within institutions.

Australian universities should support the development of rankings which first classify universities into groups by characteristics, especially by income. For most countries other than Australia classification by horizontal specialization, such as liberal arts colleges, research-intensive comprehensive universities etc, is very useful. When these rankings become international, it will be interesting to see the effect on the mission statements of those Australian universities that are unlikely to reach the Shanghai top 500.

There is a need in Australia for an ongoing appropriately funded ratings research group, at arms length from the universities and government. Such a group could develop methodologies that governments and universities could call upon when they wished to introduce financial incentives or gauge performance, whether in monitoring and fostering research, good teaching, evaluation of disciplines and so on. The group could also influence the methodologies used in international rankings.

What is a World Class University?

A related strand of research to ranking is: What is a World Class University (WCU) and how do you get one? To quote from the Tertiary Education Coordinator at the World Bank, Jamil Salmi (2007):

> *In the past decade, the term world-class university has become a catch phrase for not simply improving the quality of learning and research in tertiary education but more importantly for developing the capacity to compete in the global tertiary education marketplace through the acquisition and creation of advanced knowledge.*

In defining the attributes of a WCU, Salmi collapses the range of performance measures into a set of three factors:

> *.. the superior results of these [world class] institutions (highly sought graduates, leading edge research, technology transfer) can essentially be attributed to three complementary sets of factors that can be found at play among most top universities, namely (i) a* ***high concentration of talent*** *(faculty and students), (ii)* ***abundant resources*** *to offer a rich learning environment and conduct advanced research, and (iii)* ***favorable governance*** *features that encourage strategic vision, innovation and flexibility, and enable institutions to make decisions and manage resources without being encumbered by bureaucracy.*

How well would Australian universities fare on these criteria? On criterion (i) some Australian universities do well overall, and there are stand out discipline performers that are in the top 50 in the world. But the scores on criteria (ii) and (iii) would not be high. (For example, two North American public universities ranked by SJTU in 2007 in the top 25 in the world, Wisconsin and Toronto, had income in 2006 that was much above that of the Australian university with the highest revenue (Melbourne) —Wisconsin by 90 per cent, Toronto by 60 per cent.)

Salmi also notes that the very best universities are modest in size (often less than 20,000 students) and have a large percentage of students at the graduate level with very selective entry. Many Australian PhD programs lack the critical mass that promotes peer discussion and contributes so much to the strength of US programs. These attributes are usually missing from ranking measures.

Australia should be striving for a world class *system* of higher education. This would include some world class universities in research, other universities with pockets of world class disciplines, and some institutions opting to specialize in the provision of world class undergraduate programs.

With the additional resources now being directed towards higher education in Asia and Europe it is hard to see Australian universities maintaining their relative positions internationally without an improvement in funding – from government, students and private benefactors. Does this matter? Might it be more economic, at least in some disciplines, for research students to train overseas and to buy in overseas research findings? The returns to research suggest not, and it would lead to a reduced ability to solve peculiarly Australian problems. Crucially, a slide in the international research rankings would reduce international connectedness and the quick access to new research findings that this brings.

It may be politically difficult in Australia to follow the Chinese model of heavy additional government funding for selected institutions. An alternative is to encourage greater specialization across institutions, either though government funding or fiat.

References

Hazelkorn, E. (2007), "The Impact of League tables and Ranking Systems on Higher Education Decision Making", *Higher Education Management and Policy*, 19 (2), pp. 81-105.

Marginson, S. (2007) "Global University Rankings". Chapter 3 in S. Marginson (ed.), *Prospects of Higher Education: Globalization, market competition, public goods and* ***the*** *future of the university*, Sense Publishers, Rotterdam, pp 79-100.

Salmi, J. (2007), "The Challenge of Establishing World-Class Universities". Paper presented at 2nd International Conference on World-Class Universities (WCU-2), Shanghai Jiao Tong University, October 31-November 3, 2007. (Forthcoming in *Higher Education in Europe*.)

Salmi J. and Saroyan, A (2007), "League Tables as Policy Instruments: Uses and Misuses", *Higher Education Management and Policy*, 19 (2), pp. 24-62.

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/

Usher, A. and Savino, M. (2007), "A Global Survey of University Ranking and League Tables, *Higher Education in Europe*, 32 (1), pp. 5-15.

Van Dyke, N. (2005), "Twenty Years of University Report Cards", *Higher Education in Europe*, 30 (2), pp. 103-125.

Williams, R. (2007), "Ranking Australian Universities: Controlling for Scope", Melbourne Institute *Research Report, October 2007*. Forthcoming in *Higher Education in Europe*.

Williams, R. and Van Dyke, N. (2006), "Rating Major Disciplines in Australian Universities: Perceptions and Reality", Melbourne Institute Research Report, November. Forthcoming in *Higher Education*.

Williams R. and Van Dyke N. (2007), "Measuring the International Standing of Universities with an Application to Australian Universities", *Higher Education*, 53 (6), pp. 819-841.

Appendix: The Berlin Principles

A. Purposes and Goals of Rankings

1. Be one of a number of diverse approaches to the assessment of higher education inputs, processes, and outputs.

2. Be clear about their purpose and their target groups.

3. Recognise the diversity of institutions and take the different missions and goals of institutions into account.

4. Provide clarity about the range of information sources for rankings and the messages each source generates.

5. Specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked.

B. Design and Weighting of Indicators

1. Be transparent regarding the methodology used for creating the rankings.

2. Choose indicators according to their relevance and validity.

3. Measure outcomes in preference to inputs whenever possible.

4. Make the weights assigned to different indicators (if used) prominent and limit changes to them.

C. Collection and Processing of Data

1. Pay due attention to ethical standards and the good practice recommendations articulated in these Principles.

2. Use audited and verifiable data whenever possible.

3. Include data that are collected with proper procedures for scientific data collection.

4. Apply measures of quality assurance to ranking processes themselves.

5. Apply organizational measures that enhance the credibility of rankings.

D. Presentation of Ranking Results

1. Provide consumers with a clear understanding of all the factors used to develop a ranking, and offer them a choice in how rankings are displayed.

2. Be compiled in a way that eliminates or reduces errors in original data, and be organized and published in a way that errors and faults can be corrected.