

International university rankings: For good or ill?

Bahram Bekhradnia



Higher Education Policy Institute

HEPI Report 89

About the Author

Bahram Bekhradnia established the Higher Education Policy Institute in 2002, and was its Director until January 2014 following which he became HEPI's first President.

Before establishing HEPI he had been Director of Policy for the Higher Education Funding Council for England (HEFCE) since its formation in 1992.

Contents

Introduction	3
How rankings work	5
Rankings – positive or negative	9
Lessons	23

Introduction

One remarkable feature of higher education in the past decade or so has been the prominence achieved by rankings. Rankings – both national and international – have not just captured the imagination of higher education but in some ways have captured higher education itself. In part, this is because of the increasing spirit of competition between institutions as they strive to recruit students, to secure commercial and other research funding and to enhance their prestige. In part also, this is because there has been demand for information about comparative standings of institutions from applicants, parents and society. International rankings claim to provide a more objective account of relative quality than perceptions based simply on history and reputation; and indeed, to some extent, they claim to enable not only institutions but systems as a whole to be compared between countries.¹

There are four main international rankings:

1. The **THE World University Rankings**, produced by *Times Higher Education (THE)*;
2. The **QS World University Rankings**, produced by Quacquarelli Symonds Ltd (QS);
3. The **Academic Ranking of World Universities (ARWU)**, produced by Shanghai Jiao Tong University; and
4. **U-Multirank**, produced by a consortium led by the Centre for Higher Education in Germany (CHE), the Center for

¹ A ranking of university systems provided by Universitas 21 claims to do this.

Higher Education Policy Studies (CHEPS) at the University of Twente and the Centre for Science and Technology Studies (CWTS) from Leiden University, and initially funded by the EU.

Despite their increasing sophistication, the first three are essentially unidimensional: they provide rankings based on research criteria. Although they claim to take account of other dimensions of universities' performance, it is essentially research that they measure. U-Multirank allows comparisons on different dimensions but is beset by other problems – for example, it is very partial as many universities refuse to provide it with their data.

Almost all rankings publish ordinal lists of universities from best to worst. U-Multirank, however, does not produce a single list but allows users to create their own rankings based on criteria that they select rather than those selected by the rankings compilers.

Beyond the international rankings, national rankings now exist in an increasing number of countries, generally compiled by commercial entities. In the United Kingdom, many newspapers publish rankings of what they claim to be the 'best' universities. These are of varying validity but, because they are generally based on more robust data, are more highly regarded than international rankings.

How rankings work

Fundamentally rankings all work in a similar way and to similar principles:

- The compiler of the rankings identifies the broad dimensions of quality (research, teaching and so on) and then within each dimension identifies the indicators of quality for that dimension that are available and which fit their objectives.
- The compiler also has to decide which dimensions of quality are more important and which less, and similarly which indicators within each dimension are more important and which less, and attach weights to these accordingly.

The basic algorithm that is created has these two features at its core, and the results obtained by each institution are weighted and summed to calculate their relative positions. There may be variations on this. For example, the institution with the highest score on an indicator may be given a value of 100 and others awarded marks depending on their scores relative to the top scoring institution – but the fundamental principles of dimensions, indicators and weights remains the same.

Although the basic workings of rankings are common between the compilers of rankings, the specific dimensions, indicators and weights vary. In large part the variations are pragmatic, constrained by the data that are available; but in part the dimensions and indicators themselves, and the weights in particular, follow from value judgements on the part of the compilers about what is important in judging the quality of

universities and, more critically, on the data that are available or that can be constructed.

Among the international rankings the criteria and weights used by QS, *THE* and ARWU are as set out in Table 1.

Table 1: International Rankings: dimensions, indicators and weights

Rankings	Criteria of Excellence	Weights
Times Higher Education (<i>THE</i>)	• Teaching (the learning environment) (30%)	
	• Reputation survey	15%
	• Staff-to-student ratio	4.5%
	• Doctorate-to-bachelor's ratio	2.25%
	• Doctorates-awarded- to-academic-staff ratio	6%
	• Institutional income	2.25%
	• Research (30%)	
	• Reputation survey	18%
	• Research income	6%
	• Research productivity	6%
	• Citations (research influence) (30%)	30%
	• International outlook (staff, students & research) (7.5%)	
	• International-to-domestic-student ratio:	2.5%
• International-to-domestic-staff ratio:	2.5%	
• International collaboration	2.5%	
• Industry income (knowledge transfer) (2.5%)		
Quacquarelli Symonds (QS)	• Academic reputation (by survey)	40%
	• Employer reputation (by survey)	10%
	• Student-to-faculty ratio	20%
	• Citations per faculty	20%
	• International faculty ratio	5%
	• International student ratio	5%

ARWU (Shanghai Jiao Tong)	• Education Quality (30%)	
	• Alumni of an institution winning Nobel Prizes and Fields Medals	10%
	• Staff of an institution winning Nobel Prizes and Fields Medals	20%
	• Faculty quality (40%)	
	• Highly cited researchers in 21 broad subject categories	20%
	• Papers published in Nature and Science	20%
	• Research output (20%)	
	• Papers indexed in Science Citation Index & Social Science Citation Index	20%
• Per capita academic performance of an institution (10%)	10%	

Among national rankings there exists a greater variety of indicators than with international rankings, in part reflecting the greater range of data that are available. Examples are given in Table 2.

Table 2: Examples of domestic rankings, dimensions of excellence, indicators and weights²

Rankings	Criteria of Excellence	Weights
<i>Maclean's</i> University Rankings (Canada)	• Students and classes	20%
	• Faculty	20%
	• Resources	12%
	• Student support	13%
	• Library	15%
	• Reputation	20%

² Weights converted to percentages where appropriate, to ease comparisons, and subject to rounding.

<i>US News & World Report Best College Rankings (USA)</i>	• Undergraduate academic reputation	22.5%
	• Retention	22.5%
	• Faculty resources	20%
	• Student selectivity	12.5%
	• Financial resources	10%
	• Graduation rate performance	7.5%
	• Alumni giving rate	5%
<i>Guardian (UK)</i>	• National Student Survey Teaching	10%
	• National Student Survey Assessment & Feedback	10%
	• National Student Survey Overall score	5%
	• Value Added	16.5%
	• Student Staff Ratio	16.5%
	• Expenditure per student	10%
	• Entry Scores	16.5%
• Career Prospects	16.5%	
<i>The Complete Universities Guide (CUG) - published in the Independent (UK)</i>	• Entry standards	11%
	• Student satisfaction (from National Student Survey)	17%
	• Research assessment	11%
	• Research intensity	6%
	• Graduate prospects	11%
	• Student-staff ratio	11%
	• Academic services spend	6%
	• Facilities spend	6%
	• Good honours	11%
• Degree completion	11%	

Rankings – positive or negative?

Although rankings attract considerable criticism, they also have perceived benefits.

Among the benefits claimed for rankings are that they provide information to potential students, particularly those from backgrounds where there is little knowledge of higher education or of individual providers, and they similarly provide information to policymakers and other stakeholders. A further claim made for rankings is that they provide benchmarking information for universities themselves and are a stimulus to improvement both as a result of such benchmark information and because of the incentives to improvement that competition provides. For example, Jan Konvalinka, Vice-Rector of Charles University, Prague, has said that the availability of international comparisons might enable him to shake his staff out of their complacency by showing that, although they are first in their own country, they are way down in international terms – a sentiment often repeated by institutional leaders.

Against these positives, four major negative factors principally affect the international rankings:

1. The impact of rankings – the extent to which they influence the decisions of governments and universities themselves;
2. The data on which the rankings are based – if the data are not robust, then the rankings that follow will not be robust either;

3. The way rankings are presented – the results are generally presented as an ordinal list from best to worst, according to the number of points scored in the rankings scheme; and
4. choice of appropriate weights.

These issues are less acute with regard to rankings produced within a country.

Impact of rankings and distortion of behaviour

One of the most acute problems with regard to international rankings is that they only measure research performance. Although they claim to measure other things, fundamentally the measures used are measures of research performance and activity.

- Specifically, reputation surveys, which account for half of the inputs to the QS ranking and one-third of those to the *THE*, can be nothing much more than surveys of research reputation. How, other than through knowledge of research articles, conference presentations, historical prestige and so on, is an academic in, say, Belgium likely to be aware of a university in Australia? They are certainly most unlikely to know anything about the quality of the teaching or outreach, which may be outstanding. And the survey of employers is even less likely to provide meaningful information about the comparative quality of universities in different countries. Again, it seems far-fetched to expect an employer in Belgium to provide a view about the quality of graduates of an Australian university.

- Academic staff to student ratio is a measure claimed by both *THE* and QS to be an indicator of teaching quality. But there is no attempt to separate out the research effort of the staff concerned: and the more research a university does, the more staff it will have. Staff to student ratios are an almost direct measure of research activity.
- Similarly, commercial income is in large part a function of research, albeit applied research.
- Included under the heading 'internationalism' is the ratio between international staff and local staff. Again, the recruitment of international staff is much more likely to be related to research expertise than teaching ability.
- The proportion of PhD students is also to a large extent an indication of research activity. It says little about the quality of the education.
- The ARWU's use of the number of Nobel Prize winners among a university's staff as a measure of teaching as distinct from research excellence seems extraordinary.

Only the ratio of international to domestic students can reasonably be claimed to be a factor independent of research but that partly arises from a country's migration policies. Moreover, it accounts for only a tiny proportion of the value in any ranking. Calculated properly over 85 per cent of the measures attached to the *THE* and the QS rankings – and 100 per cent of those of ARWU – are in one way or another research-related.

So the only way of improving performance in the international rankings is to improve research performance. This drives universities around the world, at the expense of a focus on teaching, on widening participation and on outreach. Such a focus on research is appropriate only for a small number of universities. One important – perhaps the most important – function of universities is to develop the human capital of a country, and enable individuals to achieve their potential. Most institutions should be focusing on their students, and a ranking scheme that takes no account of that cannot, as all do, claim to identify the ‘best’ universities.

In their anxiety to rise up the international rankings, universities have prioritised research over other activities. Governments too have changed their policies explicitly in response to the performance of their universities in international rankings. The governments of France, Germany, Russia and China, among others, have put in place policies aimed explicitly at improving the position of their universities in international rankings.³ The result is that very large amounts of money are being provided to a small number of selected universities in order to enable them to improve their research – money that could have been used elsewhere in the system to improve other aspects of university performance. In other words, there is an opportunity cost as well as a cash cost.

³ One of the consequences of the fact that international rankings are based almost exclusively on research criteria is that German universities perform relatively poorly, because much of the research carried out in that country is conducted in the Max Planck and other research institutes.

Data issues

Valid comparisons require confidence that the data on which the rankings are based have been gathered to comparable standards and using the same definitions. No such confidence exists, other than in relation to research publication data where the arrangements for journal publication, peer review and citations are well established and internationally used. But for no other data are there such internationally comparable arrangements. There is not even a universally used definition of a student: the notion of a full-time equivalent student does not exist everywhere, and in some systems a Master's student is not distinguishable from an undergraduate.

The definition of a full-time member of academic staff also varies from country to country, in some including PhD students. The *THE* and QS do produce their own definitions, but other than for research publications they cannot rely on international databases for the data. Universities supply their own data and the compilers of the rankings accept the data as supplied: there can be no confidence that the thousands of universities around the world that are assessed in the rankings are using these definitions in a consistent way.

Worse, there is no effective attempt by the compilers of the rankings to audit or assure the quality of the data that are submitted. While *THE* and QS rankings have had aspects of their rankings audited – their internal processes and calculations – these audits do not extend to the accuracy of the data submitted by institutions. Data integrity for the most part is a matter left to the institutions themselves. A case in point is provided by

the most recent *THE* rankings, where before publication Trinity College Dublin concerned that its position had deteriorated, investigated and discovered that on a key measure it had misplaced the decimal point in its returns - an error of a factor of 10! There was no effective data verification process other than that conducted subsequently by the University.

Trinity College Dublin is a respectable institution and, although what it uncovered was a data error that worked against its interests, we can be confident that it would have drawn attention to errors that worked in its favour. But there may be other universities that submit erroneous data – whether deliberately or accidentally – and there can be no confidence in the accuracy of their data nor whether the data are produced to common definitions and standards. Although the ranking bodies do have automated checks which are intended to ensure, for example, that the data returned are within credible bounds, these fall far short of an audit of the data returns themselves.

While relying on universities to supply their own data gives rise to problems, those problems are compounded by the practice engaged in by QS (though not apparently by *THE*) of 'data scraping'. This involves the seeking out of data from a variety of data sources (institutions' websites, for example), where a university does not itself provide data to the ranking body, and where there is absolutely no control over the data that are being gleaned.

An example of the sort of problem that this can give rise to is provided by the fact that in 2013 Sultan Qaboos University in

Oman found that it had dropped 150 or so places in the QS tables. When it queried this, it was told that the data scraping in previous years had wrongly identified non-teaching staff as teaching staff thereby greatly enhancing the academic staff to student ratio that had been used. *THE* on the other hand, which provides definitions and obtains all its data direct from the universities, does not suffer from these further problems.

The 'reputation surveys' provide particular grounds for concern. These are especially significant in the QS ranking, accounting for 50 per cent of the total weight, and where the quality control over the data is especially inadequate. First, in the QS survey of academics universities are invited to make suggestions about academics that should be surveyed – and it can be imagined that universities are likely to nominate well-disposed academics. No attempt is made to ensure that those surveyed comprise a statistically representative sample (this is not a problem with the *THE* survey, where the ranking body itself selects the academics to be surveyed and ensures that the sample is representative).

Second, each year the list of those surveyed by QS is extended by adding to the previous year's list – with responses recorded up to five years previously being counted, without regard to whether the respondents are still active in their current positions and role, nor even whether they are still alive. (The *THE* survey counts responses from the previous year only.)

Third, although there is some attempt at normalisation, there is no satisfactory attempt to allow for biases in subject discipline, country of respondent or previous institutional affiliation.

Finally, the response rate is less than 10 per cent. For a measure that accounts for half of the QS ranking (taking both the survey of academics and employers into account) and one third of the *THE*, these issues further undermine the rankings produced - quite apart from the question of whether opinion surveys provide meaningful evidence of quality.

ARWU presents a further data issue. Whereas in the case of the other rankings the results are adjusted to take account of the size of institutions, hardly any such adjustment is made by ARWU. So there is a distortion in favour of large institutions. If two institutions were to merge, the very fact of merger would mean that the merged institution would do nearly twice as well as either of the individual institutions prior to merger, although nothing else had changed.

So the issues with regard to data are twofold.

- First, there can be no confidence in the quality of the data used in the rankings. Universities supply their own data which are not subject to proper audit – worse, in the absence of university-provided data QS engages in data scraping, and uses data over whose origins and definitions there is no control whatever.
- Second, in the absence of meaningful data, reputation surveys are used to bulk up the basis for the rankings, accounting for half of the rankings produced by QS and one third by *THE*. There are severe doubts over the samples used, quite apart from the question of whether such surveys provide meaningful indicators of quality.

Presentation issues and choice of weights

Other than U-Multirank, all the rankings considered present conclusions in the same way: an ordinal list from first to last in order of the scores achieved – although there is also some bunching into groups at the lower numbers (for example, 201-250, 251-300 and 301-350). Such a presentation can be seriously misleading:

- First, it can exaggerate the significance of very small differences. In the latest QS survey, the Humboldt University of Berlin achieved a score of 58.9 per cent and was ranked 121st in the world. With just four percentage points more, it would have been ranked in the top 100, above the Karlsruhe Institute of Technology, its compatriot ranked at 98th. It could be that the Karlsruhe Institute of Technology is a better university than the Humboldt University of Berlin, but the way this difference of four points is presented gives the misleading impression that it is very much better.
- Secondly, there is nothing objective about the criteria chosen for the rankings – the dimensions and indicators on the one hand and the weights attached to them on the other. The fact that the different ranking bodies use different indicators and weights bears witness to this. It could be that with slightly different weights attached to different indicators or indeed with different indicators, very different results would have been achieved.

Table 3 below shows also how sensitive ordinal lists are to the weights placed upon different indicators. It shows that

if Indicator 1 is weighted 0.5 and Indicator 3 is weighted 0.2, institution B is top and institution C is last. If the weights are reversed so are the positions of the institutions. There is nothing objective or scientific about the indicators or weights – or therefore the rankings themselves.

Table 3: Impact on weight changes on Ordinal Lists

Institution	Score on Indicator 1	Score on Indicator 2	Score on Indicator 3
A	91	60	75
B	85	77	67
C	65	90	80

				Ranking
weights	0.5	0.3	0.2	B,A,C
weights	0.4	0.2	0.4	A,B,C
weights	0.2	0.3	0.5	C,A,B

A comparison of the criteria and weights used in the Complete University Guide (CUG) and in the *Guardian* rankings of UK universities illustrates this point. Table 2 shows the dimensions used in each and the weights attached to these; and Table 4 shows the top 10 universities in each.

Table 4: Top 10 universities in the *Guardian* and CUG rankings

Rank	CUG	Guardian
1	Cambridge	Cambridge
2	Oxford	Oxford
3	LSE	St Andrews
4	Imperial College	Surrey
5	St Andrews	Bath
6	Durham	Durham
7	Loughborough	Warwick
8	Warwick	Imperial College London
9	Lancaster	Exeter
10	UCL	Lancaster

Both purport to identify the ‘best’ 10 universities, and produce quite different lists – indeed three institutions in each are completely absent in the other. Neither is wrong, but neither is right. It is misleading to pretend that one university is better than another based on the subjective judgements of a compiler. Rankings are not only subjective but misleading.

A less misleading way of presenting the results would be to present bands rather than ordinal lists – very much like the results of the Research Assessment Exercise in the past. This would still have unsatisfactory features, but would avoid some of the most egregious problems of the ordinal list type presentation. So all universities above a certain score would be in band A, and bands would be created for those with lower scores. Tables 5 and 6 demonstrate such an arrangement, first as an ordinal list and then using illustrative data and bands.

Table 5 Ordinal List			Table 6 Banding	
1	HEI A	78%	HEI A	Level I
2	HEI B	76%	HEI B	
3	HEI C	66%	HEI C	Level II
4	HEI D	62%	HEI D	
5	HEI E	55%	HEI E	
6	HEI F	51%	HEI F	
7	HEI G	45%	HEI G	Level III
8	HEI H	25%	HEI H	Level IV

Such a presentation would still have issues of cut-off points and steps – though fewer of them – and it might avoid some of the misleading impressions given by the precise numbers provided in ordinal presentation.

Both types of presentation combine the different dimensions of an institution’s performance and present a single judgement. A more sophisticated presentation that would show the performance of each institution on each dimension rather than as a single summary would be to present results as a Radar diagram.⁴ Putting aside concerns about whether the rankings methodologies meaningfully capture the dimensions of institutions’ performance as they claim, such a presentation would enable the claimed strengths and weaknesses of each to be identified at a glance. It could show, for example, that an institution might be strong on research but weaker on resources and internationalisation – and arguably may be more useful to a potential student than a simple ordinal list or

⁴ Such a presentation is used in the comparative study of research performance conducted for the Business, Innovation and Skills Department by Elsevier.

banding which simply says where the institution as a whole stands. Such a presentation is reproduced in Figures 1 and 2 for two institutions – A and B.

Figure 1 shows Institution A performs much better than the average, especially for ‘Societal Impact’ and ‘Research’, while Figure 2 shows Institution B performs slightly better than average on three measures, and much worse than Institution A on most. Such a picture provides a more rounded view of the strengths of a university, and facilitates direct comparison between a small number of universities. However, it is not an effective way of comparing large numbers of institutions, and therefore may be of less interest to politicians, newspapers and the general public.

Figure 1 Radar Diagram for Institution A

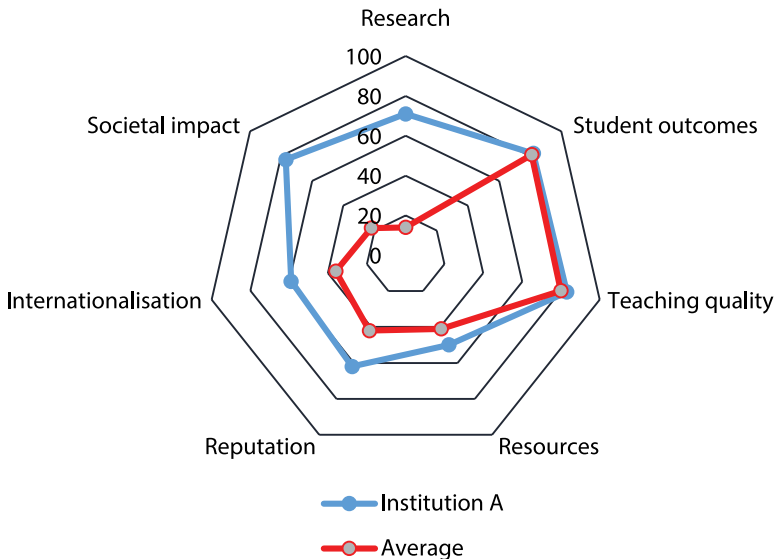
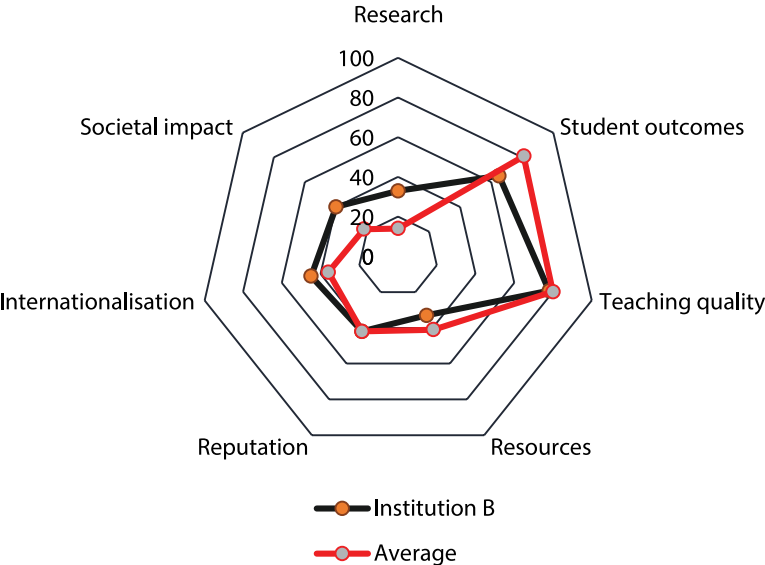


Figure 2 Radar Diagram for Institution B



Lessons

There is commercial interest in maintaining rankings, and it may appear Canute-like to try to halt their rise. Nor does there seem much hope of correcting their most serious faults. Comparable data simply do not exist internationally other than for research.

Among the problems is that universities provide their own data – though the fact that the rankings bodies provide definitions is helpful: they should audit and validate the data provided by universities to ensure that these common definitions are adhered to. On the other hand, where institutionally provided data do not exist, they should certainly avoid data scraping where there is no control over the data that are collected.

The difficulty within England of identifying robust and valid measures for teaching excellence shows how hard it will be to devise anything internationally, and attempts to create these are still some way from fruition (as shown in HEPI's 2015 Annual Lecture by Andreas Schleicher, Director for Education and Skills at the Organisation for Economic Co-operation and Development). Nevertheless, it should be a key aim of ranking bodies to broaden the criteria from their almost 100 per cent reliance on research-related measures.

The international surveys of reputation should be dropped – methodologically they are flawed, effectively they only measure research performance and they skew the results in favour of a small number of institutions.

Universities – particularly the top-ranked – already in

considerable numbers refuse to provide data to U-Multirank, so impacting its credibility. However, it is unlikely that they will do so with regard to the major international rankings. But if some of the top-ranked universities were to join together and refuse to provide data unless the ranking methodologies were broadened and improved, then that might be effective. Unfortunately it is those very same universities that are likely to suffer with the broadening of criteria beyond research, and so that may be an aspiration too far. Similarly, universities cannot forbid their staff from participating in the surveys – which are perhaps the most shocking element of the rankings but are at the heart of two of them. Universities might nevertheless discourage their staff from doing so.

Ranking providers should acknowledge that their rankings effectively only measure research performance; and the sector more widely should acknowledge and publicise this fact. If governments, students and the public recognised this then some of the negative impacts of the rankings would be reduced.

The lack of success of U-Multirank provides some lessons. Although apparently fairer and more robust in concept than the unidimensional rankings of the others, it does not provide a simple answer to the question 'is one university better than another?'. So it fails to provide politicians, the media and the general public with a simple easily-absorbed answer.

Finally, universities and governments should discount the rankings when deciding their priorities, policies and actions.

In particular, governing bodies should resist holding senior management to account for performance in flawed rankings. Institutions and governments should do what they do because it is right, not because it will improve their position in the rankings.

Trustees

Professor Sir Ivor Crewe (Chair)
Dame Sandra Burslem
Professor Sir Peter Scott
Professor Sir Nigel Thrift

Advisory Board

Professor Janet Beer
Sir David Bell
Professor Sir David Eastwood
Professor Dame Julia Goodfellow
Professor David Maguire
Professor Dame Helen Wallace

Partners

BPP University
Ellucian
Elsevier
Higher Education Academy
Jisc
Kaplan
Mills & Reeve LLP
Pearson
UPP Group Limited
Wiley

President

Bahram Bekhradnia

Rankings of universities have become increasingly influential in recent years. This report demonstrates the dangers that this represents. It shows that international rankings are almost entirely based on research-related criteria, and if universities are to move up the rankings the only way of doing so is to focus on their research performance – at the expense of teaching, widening participation and other activities.

The report analyses the data on which the rankings are based and demonstrates that these are of doubtful quality. It concludes by urging governments, university management and governing bodies to ignore rankings when making decisions and to do what they do because it is right, not because it will impact performance in rankings.

HEPI was established in 2002 to influence the higher education debate with evidence.

We are UK-wide, independent and non-partisan.

December 2016 ISBN: 978-1-908240-20-0

Higher Education Policy Institute
99 Banbury Road, Oxford OX2 6JX
Tel: 01865 284450 www.hepi.ac.uk

Printed in the UK by Oxuniprint, Oxford

Typesetting: Steve Billington, www.jarmanassociates.co.uk