Contribution to the Theme Section 'Global university rankings uncovered'

The Times Higher Education World University Rankings, 2004–2012

Phil Baty*

Times Higher Education Rankings, 26 Red Lion Square, London WC1R 4HQ, UK

ABSTRACT: In the present essay, I briefly describe the transition from the original Times Higher Education Supplement World University Rankings, which were developed together with Quaquarelli Symonds (QS), to the Times Higher Education world rankings powered by Thomson Reuters. In addition, I describe the 'sample' characteristics (i.e. the distribution of respondents by geographic area and scientific discipline) of the Thomson Reuters' annual academic reputational surveys, upon which 2 key indicators for the categories of teaching and research are based, during 2010 to 2012. Finally, I briefly discuss the criticisms raised concerning these 2 ranking systems.

KEY WORDS: Global university rankings · Times Higher Education Supplement · Thomson Reuters · Academic Reputational survey · Quaquarelli Symonds · Indicators

THE EARLY TIMES HIGHER EDUCATION SUPPLEMENT RANKINGS

Global university rankings have been a part of the higher education landscape for 10 yr now. The first world ranking was produced in 2003 by Shanghai Jiao Tong University in China — it was designed initially as a simple exercise, internal to the university, to monitor the university's research performance against competitors, but it rapidly evolved into a major public phenomenon (e.g. Rauhvargers 2011). The Shanghai ranking was quickly followed in 2004 by the first World University Rankings produced by the Times Higher Education (THE) magazine (which was at the time known as the Times Higher Education Supplement [THES], before a name change in 2008) (Baty 2013).

The original THES global ranking used data collected by a company called Quaquarelli Symonds (QS) and was published as the THES-QS World University Rankings during 2004 and 2009. THES-QS was designed to take a broader look at the world-

class university: looking at research performance as the Shanghai rankings did, but also adding indicators that were supposed to capture something of the 'teaching capacity' as well as the international outlook of an institution. It was the first global ranking to use a global survey of academics' opinion as well. The THES-QS initial methodology was exceptionally simple, based on 5 performance indicators (i.e. staff-student ratio, reputation of academics [based on a survey], research paper citations, proportion of international staff on campus and proportion of international students, and the responses to a survey of employers, introduced later, who were asked which institutions they like to recruit from), with weights of 20, 20, 50 (40 % after the employer survey was introduced), 10 and 10 %, respectively.

While the THES-QS approach was pioneering at the time, the higher education landscape has changed dramatically since 2004 (e.g. www.oecd.org/edu/eag.htm): the globalization of higher education has continued apace with greater student and faculty mobility and more cross-border research col-

laboration; the internal information management systems of universities have improved; and the demand for ever more sophisticated comparative performance data has intensified. In retrospect, the old THES-QS system now looks hopelessly crude by today's standards.

THE CRITICISMS

The rankings proved remarkably popular, gaining worldwide media interest and a global reach and influence (e.g. Butler 2010). However, as the prominence and power of the rankings grew, so did the criticism of the old THES-QS methodology. Thus, for instance, in November 2007, Ian Diamond, who was at the time chief executive of the Economic and Social Research Council (and who is now vice-chancellor of the University of Aberdeen), wrote to THE to criticize the THES-QS rankings' misleading use of citations data to judge research quality (Diamond 2007):

> We would argue that the nature of one of the components of the rankings undervalues… institutions strong in social science. The use of a citation database must have an impact because such databases do not have as wide a cover of the social sciences (or arts and humanities) as the natural sciences. …. In addition, we know that in the social sciences the databases tend to have a more comprehensive coverage of US journals and, given the context- specific nature of much social science research, this may affect the frequency of citation of UK institutions strong in social sciences relative to their US counterparts.

The fact that the THES-QS rankings examined citations per faculty member, while failing to take into account dramatically different volumes of citations between different subject areas, was particularly problematic. Other criticisms focused on the small and unrepresentative sample size used for the academic reputation survey. All these concerns were subsequently highlighted by Rauhvargers (2011), who noted that the reputation scores in the THES-QS ranking system (used between 2004 and 2009) were based on 'a rather small number of responses: 9,386 in 2009 and 6,534 in 2008; in actual fact, the 3,000 or so answers from 2009 were simply added to those of 2008'. Rauhvargers (2011) also maintained, 'The number of answers is pitifully small compared to the 18,000 email addresses used' and that the lists from which survey respondents were asked to select from were restricted: 'What are the criteria for leaving out a great number of universities or whole countries?'

Bookstein et al. (2010) also strongly criticized the instability of the THES-QS results, writing:

> Several individual indicators from the Times Higher Education Survey (THES) [sic] database — the overall score, the reported staff-to-student ratio, and the peer ratings — demonstrate unacceptably high fluctuation from year to year. The inappropriateness of the summary tabulations for assessing the majority of the 'top 200' universities would be apparent purely for reason of this obvious statistical instability regardless of other grounds of criticism. There are far too many anomalies in the change scores of the various indices for them to be of use in the course of university management.

One of the most damning and powerful criticisms came from Andrew Oswald, Professor of Economics at the University of Warwick. Oswald (2007), commenting on the 2007 annual rankings results in the *Independent* newspaper on 13 December 2007, said:

> The organisations who promote such ideas should be unhappy themselves, and so should any supine UK universities who endorse results they view as untruthful. Using these league table results on your websites, universities, if in private you deride the quality of the findings, is unprincipled and will ultimately be destructive of yourselves, because if you are not in the truth business what business are you in, exactly?

THE TIMES HIGHER EDUCATION RANKINGS POWERED BY THOMSON REUTERS

Despite such strong criticisms, THE continued to publish the rankings with QS until 2009. However, during 2009, a new senior editorial team at THE magazine, conscious of the mounting criticism, carried out a comprehensive internal review of the rankings methodology. The review concluded that the global rankings that the magazine had been publishing since 2004 were no longer fit for the purposes being assigned to them. Mroz (2009), then editor of THE, explained in an editorial that THE must improve the way that data are compiled and that universities deserve 'a rigorous, robust and transparent set of rankings'.

In November 2009, THE ended its 6 yr relationship with its previous rankings data supplier, QS, and set up a new partnership with Thomson Reuters to develop a new, more sophisticated ranking system, the THE World University Rankings, powered by Thomson Reuters (Baty 2013). In this partnership, THE had the responsibility for the rankings methodology and institutions, and Thomson Reuters was responsible for collecting, analysing and supplying the data, without itself publishing a ranking (Baty 2013). QS continued to publish the heavily criticized

former THES-QS rankings and continues to do so today, under the name the QS World University Rankings.

In a key step toward developing a new and improved ranking system, Thomson Reuters carried out a global opinion survey to find out what higher education professionals and student consumers of rankings thought of existing ranking systems (Adams & Baker 2010). The survey raised a series of concerns about the existing ranking systems (Adams & Baker 2010):

> The data indicators and methodology currently utilized were perceived unfavorably by many and there was widespread concern about data quality in North America and Europe. A concern found in the survey, and echoed in discussions with representative groups, was that published ranking tables could have more insidious effects. They changed the behavior, even the strategy, of institutions, not to become more effective, but to perform well against arbitrary ranking criteria. Some would even manipulate their data to move up in the rankings. This is of great concern and warns against any reliance on indicators that could be manipulated without creating a real underlying improvement.

The Thomson Reuter survey found, in particular, that 74% of respondents agreed with the statement that 'Some institutions manipulate their data to move up in the ranking,' and 70% maintained that methodologies and data used in rankings are neither transparent nor reproducible. In any case, the survey found there was strong support for the utility of global university rankings:

> The overriding feeling was that a need existed to use more information, not only on research, but also on broader institutional characteristics… Respondents generally felt that the current analytic comparison systems had recognizable utility. About 40% globally said they were 'extremely/very useful' and a further 45% said they were 'somewhat useful.' (Adams & Baker 2010).

This survey report also provided some clear information on what sort of indicators the users of rankings valued and what indicators they wanted to see being employed (see Baty 2013 for details). THE sought to directly address the concerns outlined in the survey, and also those highlighted by the growing body of critics throughout the years, and to deliver a new set of metrics that were valued by the users of rankings and a methodology that addressed the concerns about data quality and manipulation.

In developing a new methodology, it was important to establish that THE's World University Rankings examine only a globally competitive, research-led elite group of institutions. Such institutions are research-led but of course have teaching as a funda-

mental role. Our metrics reflect the Humboltian ideal of an institution which combines teaching and research, where undergraduate and graduate learning takes place in close proximity to new knowledge creation, in a shared undertaking between student and lecturer. There is no 'one size fits all' methodology, and no single group of metrics should be used to judge all institutions: the diversity in global higher education is one of its great strengths, and metrics for globally-focused research institutions are not appropriate for examining the performance, for example, of local or national teaching-led institutions. It is simply not appropriate to judge every university on the same scale against the model set by the universities such as Harvard, Stanford, Oxford and Cambridge. Different missions and roles — for example, social inclusion, or local skills development — require different metrics or different combination of metrics. Thus, THE's official rankings list comprises only the first 200 placed universities — representing only ~1% of higher education institutions in the world. We seek to provide an insight into the full range of each institution's activities, across the teaching environment, research, international outlook and knowledge transfer, but our metrics are weighted toward research activity. The underlying database that fuels the rankings, owned by Thomson Reuters, currently contains ~650 institutions in total, selected for deeper data collection and analysis on the basis of their research publication output and impact. The final rankings list is restricted specifically to undermine the notion that everyone should aspire to the same model.

The new rankings — first published on 16 September 2010 and again on 6 October 2011, 4 October 2012 and 3 October 2013 — recognize a wider range of what global universities do. While the QS rankings are based substantially on subjective opinion surveys and are dominated by research, and ARWU really focus only on research performance, the THE rankings seek to capture the full range of a global university's activities: research, teaching, knowledge transfer and internationalization.

THE METHODOLOGY BEHIND THE NEW THE RANKINGS AND THE ACADEMIC REPUTATION SURVEY

The new THE rankings use 13 separate indicators — more than any other global system, with the most important innovation being the set of 5 indicators designed to give proper credit to the role of teaching in universities (total weighting of 30%). The

full methodology is described at www.timeshigher education.co.uk/world-university-rankings/2012-13/world-ranking/methodology. The reasoning behind these indicators is presented by Baty (2013). Thus, the only methodological issue that is presented here concerns the annual academic reputational survey carried out by Thomson Reuters, which brought in a third-party professional polling company, Ipsos, to conduct the survey (the full survey instrument and more information is available at http://ip-science.thomsonreuters.com/globalprofilesproject/gpp-reputational/methodology/). In the interests of transparency, since March 2011, THE has made the results of the reputation survey public, in isolation from the other rankings indicators. The results of each year's reputation survey are published as the Times Higher Education World Reputation Rankings, each March, at www.timeshighereducation.co.uk/world-university-rankings/2013/reputation-ranking. This survey is actually used for 2 indicators: the data collected on teaching reputation are used for the 5 'teaching environment' indicators (worth 15% overall), and the data collected on research reputation are used within the category 'research: volume, income and reputation' (18% overall).

The Academic Reputation Survey is distributed worldwide each spring. It examines the perceived prestige of institutions in both research and teaching. Respondents are asked only to pass judgment based on direct, personal experience within their specific area of expertise. The respondents are asked 'action-based' questions, such as 'Where would you send your best graduates for the most stimulating postgraduate learning environment?' to elicit more meaningful responses.

The number of respondents increased from 13 388 in 2010 to 17 529 in 2011 and 16 696 in 2012 (Table 1), derived from 131, 137 and 144 countries, respectively. Respondents were from 5 different geographic areas, with Africa and Oceania being underrepresented (Table 1). Respondents were from 6 general academic disciplines, with the percentages of the 6 disciplines differing by geographic area and year (Table 2). Arts and Humanities were clearly underrepresented in all geographic areas and years compared to the remaining 5 general disciplines (Table 2). No one who completed the survey in 2010 was invited to take part again in 2011. Nearly 75% of the respondents were academic staff, with an even higher per-

centage among those who are full-time. The survey itself was 20 min in length, and respondents could choose to take the survey in 7 languages in 2010 (English, French, German, Spanish, Portuguese, Chinese, and Japanese) and in 10 languages in 2011 and 2012 (including Latin American Spanish, Brazilian, and Arabic). The average academic respondent had been working at an institution for over 16 yr (in 2010 and 2011) and 17 yr (in 2012). For the 2010 survey, the average academic had published over 50 scientific papers (median = 30). Academics involved in Arts and Humanities and Social Sciences published less frequently in journals than academics in hard sciences. This is the main reason why these 2 general disciplines are less represented in the data, as the sample was pulled based on journal publishing records.

THE RESPONSE TO THE NEW TABLES

The response to these new tables was rich and encouraging. For instance, in 2010, David Naylor, president of the University of Toronto, recognized that THE consulted widely to pinpoint weaknesses in other ranking systems and in the previous THE approach (see Beck & Morrow 2010). Ferdinand von Prondynsky, vice chancellor of Robert Gordon University in Scotland, said that the THE rankings are 'now increasingly seen as the gold standard' (http://universitydiary.wordpress.com/2011/09/05/the-rankings-season/). David Willetts, the UK Minister of Universities and Science, said in October 2012, 'we broadly accept the criteria used by THE, which is why our policies are focused on the same areas' (www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking/analysis/david-willetts). More recently, in May 2013, Shashi Tharoor, Minister of State for Human Resource Development in India, said, 'Times Higher Education

Table 1. Number of respondents (and percentages) by geographic area and year of the Thomson Reuters' annual academic reputational surveys in 2010 to 2012

Geographic area	No. of respondents by year			Percentages by year		
	2010	2011	2012	2010	2011	2012
Americas	5088	7699	6520	38	44	39
Africa	468	718	555	3	4	3
Asia/Middle east	2932	3069	3677	22	18	22
Oceania	1110	1230	1633	8	7	10
Europe	3790	4813	4311	28	27	26
Total	13 388	17 529	16 696	100	100	100

Table 2. Percentages of respondents by academic discipline and year of the Thomson Reuters' annual academic reputational surveys in 2010 to 2012

Discipline/year	Americas	Africa	Asia, Middle East	Oceania	Europe
2010					
Social Sciences	20	15	12	33	14
Engineering and Technology	21	15	33	12	22
Physical Sciences	19	22	19	10	29
Clinical, Preclinical and Health	16	21	15	22	12
Life Sciences	16	22	19	15	19
Arts and Humanities	8	5	2	8	4
2011					
Social Sciences	20	13	18	35	15
Engineering and Technology	19	16	28	9	20
Physical Sciences	19	20	18	12	27
Clinical, Preclinical and Health	18	21	17	22	15
Life Sciences	15	26	15	12	18
Arts and Humanities	9	5	5	11	6
2012					
Social Sciences	22	22	18	37	21
Engineering and Technology	21	13	29	11	21
Physical Sciences	17	18	18	11	23
Clinical, Preclinical and Health	16	21	16	18	13
Life Sciences	12	16	14	12	13
Arts and Humanities	13	11	6	12	10

is widely seen as the principal yardstick we should look to' (www.timeshighereducation.co.uk/world-university-rankings/news/india-embrace). But perhaps the most gratifying praise of all came from Sir Ian Diamond, by this time vice-chancellor of the University of Aberdeen. Chairing a panel debate on 'Measuring success: League tables in the public sector' in March 2012 at the launch of a British Academy report, Diamond noted that THE had listened to the criticisms of its rankings, and learned from its early mistakes:

> For many years I was 'Outraged of Swindon [the location of the Economic and Social Research Council's headquarters]' every time that the rankings came out, on the grounds that in the initial years they did not benefit the arts and humanities and social sciences… I would like to say that [THE rankings editor] Phil [Baty] and his colleagues didn't just file the 'Outraged of Swindon' letter, they actually invited me up and we had some very sensible conversations which in my view… certainly changed positively the rankings (www.timeshighereducation.co.uk/news/the-damaging-culture-of-league-tables/419500.article).

Nevertheless, critical voices still exist, such as Rauhvargers (2013), who concludes,

> Regarding the impact of reputation-based rankings, as the reputation-based score dwindles rapidly from the first top-ranked university down to the 50th, the reputa-

tion indicators have a very limited impact on the THE World University Ranking. This implies that rankings based entirely on reputation are of little value. The very steep reputation-based curve means that the influence of reputation is substantial for the first few most highly ranked institutions but quickly decreases in significance thereafter.

Yet, the purpose of the present article was to describe the evolution of the THE World university rankings and to better describe its methodology, in particular the reputational survey, so that readers will have a better understanding of the ranking system, rather than examining the arguments on the validity of each and every indicator on which THE is based. In addition, the criticisms of the THE rankings mentioned above were raised specifically in the context of how THE sought to address them when developing an improved system and not in the context of engaging in a discussion of criticisms, which are generally raised on all types of global university rankings.

To conclude, THE accommodated, and will keep doing so in the future, its critics, sought the advice of experts and will keep seeking their advice on further methodological modifications and innovations (i.e. iPhone and iPad applications) that will eventually lead to more transparent, user-driven and multi-faceted global university rankings (Baty 2013).

LITERATURE CITED

Adams J, Baker K (2010) New outlooks on institutional profiles, February 2010. Available at http://ip-science. thomsonreuters.com/m/pdfs/Global_Opinion_Survey.pdf

Baty P (2013) An evolving methodology: the Times Higher Education World University Rankings. In: Marope PTM, Wells PJ, Hazelkorn E (eds) Rankings and accountability in higher education - uses and misuses. UNESCO Pub-lishing, Paris, p 41–54

Beck S, Morrow A (2010) Canadian universities ranked among world's best. The Globe and Mail, GlobeCampus supplement, 16 September 2010, available at www. globecampus.com

Bookstein FL, Seidler H, Fieder M, Winckler G (2010) Too much noise in the Times Higher Education rankings.

Scientometrics 85:295–299

Butler D (2010) University rankings smarten up. Nature 464: 16–17

Diamond I (2007) Social Sciences lose 1. Times Higher Edu-cation, 16 November 2007, available at www.timeshigh-ereducation.co.uk/311132.article (assessed 10 July 2013)

Mroz A (2009) Leader: only the best for the best. Times Higher Education, 5 November 2009, available at www. timeshighereducation.co.uk/408968.article

Oswald A (2007) There's nothing Nobel in deceiving our-selves. The Independent, 13 December 2007, available at www.independent.co.uk/news/education/higher/andre w-oswald-theres-nothing-nobel-in-deceiving-ourselves-764880.html

Rauhvargers A (2011) Global university rankings and their impact. European University Association, Brussels, avail-able at www.eua.be

Rauhvargers A (2013) Global university rankings and their impact — Report II. European University Association, Brussels, available at www.eua.be