



On impact factors and university rankings: from birth to boycott

Konstantinos I. Stergiou^{1,2,*}, Stephan Lessenich³

¹Laboratory of Ichthyology, Department of Zoology, School of Biology, Aristotle University of Thessaloniki, UP Box 134, 541 24 Thessaloniki, Greece

²Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research, Aghios Kosmas, 16777 Athens, Greece

³Department of Sociology, Friedrich Schiller University Jena, 07737 Jena, Germany

ABSTRACT: In this essay we explore parallels in the birth, evolution and final 'banning' of journal impact factors (IFs) and university rankings (URs). IFs and what has become popularized as global URs (GURs) were born in 1975 and 2003, respectively, and the obsession with both 'tools' has gone global. They have become important instruments for a diverse range of academic and higher education issues (IFs: e.g. for hiring and promoting faculty, giving and denying faculty tenure, distributing research funding, or administering institutional evaluations; URs: e.g. for reforming university/department curricula, faculty recruitment, promotion and wages, funding, student admissions and tuition fees). As a result, both IFs and GURs are being heavily advertised — IFs in publishers' webpages and GURs in the media as soon as they are released. However, both IFs and GURs have been heavily criticized by the scientific community in recent years. As a result, IFs (which, while originally intended to evaluate journals, were later misapplied in the evaluation of scientific performance) were recently 'banned' by different academic stakeholders for use in 'evaluations' of individual scientists, individual articles, hiring/promotion and funding proposals. Similarly, URs and GURs have also led to many boycotts throughout the world, probably the most recent being the boycott of the German 'Centrum fuer Hochschulentwicklung' (CHE) rankings by German sociologists. Maybe (and hopefully), the recent banning of IFs and URs/GURs are the first steps in a process of academic self-reflection leading to the insight that higher education must urgently take control of its own metrics.

KEY WORDS: Impact factors · Global university rankings · Boycott · Higher education · Scientific performance

— Resale or republication not permitted without written consent of the publisher —

INTRODUCTION

Managers, administrators, policy makers, journalists and the public at large all like the simple numerical ordering of people and products because it is readily accessible. Thus, it comes as no surprise that both journal impact factors (IFs) and university rankings (URs), either global¹ (GURs) or not, were met with both a sense of relief and greed by those who primarily use them (Table 1). Decisions about the fate of something are made easier, and can be more easily

¹Global university rankings are not really global. It is the companies (or institutions) that promote rankings and the universities that are highly ranked who make this claim. However, as the great majority of the world's universities are not ranked in any of the available schemes (see Table 1), using the term 'global' is granting an authenticity and credibility to rankings that they actually do not merit. The only exception is the 'Webometrics Ranking of World Universities' that ranks all existing universities. Having stated this, however, in what follows we use the term 'GUR' for rankings comparing universities from different countries and 'UR' for national/regional rankings and when summarily referring to all categories.

Table 1. Comparison of various aspects related to journal impact factors (IFs) and global university rankings (for references see text). CHE: Centrum fuer Hochschulentwicklung (Centre for Higher Education Development)

	Journal impact factors	Global university rankings
Annual revenues of implicated activity	English language academic and scientific publishing industry 9.4 billion US\$ (and 4 billion US\$ from books)	Higher education: tens of billions of US\$; for-profit universities are among the 10 most fast-growing industries in the USA
Date of inception	1975	2003
Global coverage	About 46 % of peer-reviewed journals in 2012	About 6 % of existing universities/colleges ^a
Who pays attention	Publishing companies	Newspapers, magazines, radio, TV, internet media and blogs
	Journal editors/editorial boards	Governments
	Professors	Political parties
	Graduate students, post-docs	Policy makers
	University administrators	University managers and administrators
	Libraries	Faculty
	Promotion and evaluation committees	Students and their families
		Public
Who is affected	Publishing companies/journals	University/department curricula
	Journal editors/editorial boards	Faculty (recruitment, promotion, wages)
	Faculty (promotion/hiring/tenure)	Research funding
	Young scientists (job prospects)	Students (admissions, fees)
	Research funding	Students' future job prospects
	Institute evaluations	
Frequency of calculation	Annual	Annual ^b
Method of calculation	Simple and transparent: from the number of citable items published in a journal and the number of citations these articles receive	Complex, not transparent: differing between companies
Importance	Increases with time	Increases with time
Motto	They are here to stay	They are here to stay
Diversity	Thomson Reuters' monopoly	High; >12 international institutions with many producing >1 product; many national and regional ones
Manipulation	Yes	Yes
Critics (examples)	Many	Many
	Small coverage of published items (i.e. journals, conference proceedings, books)	Methodological concerns with respect to indicators and weightings
	IFs are estimated over a very short time period, 2 yr, which does not allow enough time to really capture the impact of a publication	English speaking countries dominate rankings
	English language dominance	Teaching quality hard to be measured
	IFs must not be used to evaluate scientists and research activities	Arts, humanities and social sciences are relatively under-represented
	IFs are not comparable across disciplines	Symbolically violent character 'as a form of social categorization and hierarchization'
Response to critics	Yes	Yes
Boycott	17 May 2009 meeting of the International Respiratory Journal Editors' Roundtable: IFs ' <i>should not be used as a basis for evaluating the significance of an individual scientist's past performance or scientific potential</i> ' ^c	Many examples of universities in Asia, Pacific region, USA, Canada, Australia refusing to participate in the rankings
	December 2012 meeting of the American Society for Cell Biology—San Francisco Declaration on Research Assessment (DORA): IFs must not be used ' <i>as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions</i> ' ^d	2013 German Sociological Association: 'Scientific Evaluation, Yes—CHE Ranking, No'
	As of 20 August 2013, DORA has been signed by 9008 individual scientists and 367 organizations (46.8% from Europe, 36.8% from North and Central America, 8.9% from South America, 5.1% from Asia and the Middle East, 1.8% from Australia and New Zealand and 0.5% from Africa)	The boycott of the CHE ranking by sociologists has so far been followed by the scientific associations of historians, communication scientists, educational scientists, political scientists, Anglicists and chemists

^aThe 'Webometrics Ranking of World Universities' ranks all existing universities; ^bThe 'Webometrics Ranking of World Universities' publishes rankings every 6 mo; ^cRussel & Singh (2009, p. 265); ^dDORA (<http://am.ascb.org/dora/>, p. 1)

(albeit superficially) justified, when this ‘something’ can be expressed in numbers and ranked from (supposedly) best to worst. The historical similarities in the birth, evolution and fate of these 2 instruments of academic ‘numerology’ are striking (and summarized in Table 1). In the following sections, these analogies are explored and made transparent.

IMPACT FACTORS AND JOURNAL RANKINGS

Although the origin of IFs goes back to the late 1880s (Smith 2007), the idea of IF was first introduced by Garfield (1955) who in 1958 founded the Institute for Scientific Information (ISI), now part of Thomson Reuters. The term IF appeared for the first time in 1963 in the context of the Science Citation Index published by the ISI (Smith 2007). The estimation of IFs is very simple (resulting from the number of citations to a journal divided by the number of articles published in the journal over a period of time; Garfield 1999). Since 1975, IFs are produced annually by Thomson Reuters (Church 2011), which virtually monopolizes the arena of journal rankings. Journal rankings are also produced by a few other companies, e.g. SCImago journal rankings, with small impact, but IFs can be estimated for any journal of the world using Google Scholar and Harzing’s (2007) ‘Publish or Perish’². Thomson Reuters published IFs for about 13 000 peer-reviewed journals out of >28 000 existing ones in 2012 (Ware & Mabe 2012), reaching a coverage of about 46.4% of existing journals. Within a few decades, the IF became an advertising tool for publishing companies— attracting also the attention of journal editors and editorial boards, professors, graduate students, post-docs, university administrators, promotion and evaluation committees, and libraries (e.g. Opthof 1997, Seglen 1997, Garfield 1999, Cameron 2005, Monastersky 2005, Polderman 2007, Cheung 2008, Tsikliras 2008).

The obsession with IFs soon went global, especially in the last 2 decades, and for quite diverse academic issues such as for hiring and promoting faculty, giving and denying faculty tenure, distributing research funding, or administering institutional evaluations, affecting not least the future job prospects of young scientists (e.g. Opthof 1997, Cameron 2005, Mona-

stersky 2005, Fersht 2009, Church 2011). They are relevant as well for journals, journal editors and editorial boards (e.g. Polderman 2007).

IFs have become ‘the disease of our times’, as Sir John Sulston (joint winner of the 2002 Nobel prize in the physiology or medicine category) stated to Zoë Corbyn (2009). The paranoia of using IFs for evaluations is best described by Fersht (2009, p. 6883):

An extreme example of such behavior is an institute in the heart of the European Union that evaluates papers from its staff by having a weighting factor of 0 for all papers published in journals with $IF < 5$ and just a small one for $5 < IF < 10$. So, publishing in the Journal of Molecular Biology counts for naught, despite its being at the top for areas such as protein folding.

Although IFs do not get any media coverage and are of no concern whatsoever to the public at large, they are heavily advertised, especially in the last decade, on publishers’ and journals’ webpages as soon as they are released by Thompson Reuters. Journal editors, editorial board members and scientists get mass emails from scientific publishing companies such as: ‘*The Impact Factors have been announced. Don’t delay; find out where your favourite journal features ... The moment you’ve all been waiting for ...*’— informing them about the latest IFs of ‘their’ journals. To be sure, IFs are part of the huge publishing industry, which generates a revenue of about 9.4 billion US\$ per year (Ware & Mabe 2012) and is effectively being subsidized by the voluntary work of scientists all over the world (Tsikliras & Stergiou 2013).

UNIVERSITY RANKINGS

Just like IFs, the idea of university rankings also dates back to the 1880s, in the form of classifications of universities in the United States (Salmi & Saroyan 2007, Lynch 2013 this Theme Section). Yet, what has become popularized as ‘GURs’ was actually born in 2003 with the release of the Shanghai league table (now known as the Academic Ranking of World Universities) (e.g. Rauhvargers 2011)— thus, GURs are about 30 yr younger than IFs. When launched a decade ago, they were immediately embraced by journalists, governments, political parties and policy makers, and attracted the strong interest of faculty, students and their families as well (e.g. Clarke 2007, Salmi & Saroyan 2007, Rauhvargers 2011, 2013, Robinson 2013 this Theme Section). University managers and administrators, however, often fear them— rankings are on ‘*a thin line between love and hate*’ (Salmi & Saroyan 2007, p. 40). Obsession with

²‘Publish or Perish’ is a software program that retrieves and analyzes academic citations. It uses Google Scholar and Microsoft Academic Search (since Release 4.1) to obtain the raw citations, then analyzes them and presents a large number of statistics (see www.harzing.com/pop.htm).

rankings was soon globalized (Labi 2008). As with IFs, rankings also support a huge business: the higher education complex has an annual turnover rate of 10s of billions of US\$ (Gürüz 2011), whereas the public expenditure on education was >1.3 trillion US\$ in 1997 (UNESCO 2000) and for-profit universities are among the 10 fastest growing industries in the United States (Setar & MacFarland 2012). Their impact is constantly increasing and, contrary to the monopoly of Thomson Reuters' IF, nowadays there are >12 different GUR institutions, with many of them having several products, and several UR systems (Rauhvargers 2013). Like IFs, rankings are generally produced annually³, although their product — usually a league table — involves more complex and less transparent calculations and more variables than IFs (Rauhvargers 2013). The different ranking systems generally cover 1200 to 1500 universities (Rauhvargers 2013) out of 21 067 universities/colleges in the world (www.webometrics.info), reaching a coverage of about 6%, which is much smaller than that of IFs. Within less than a decade, rankings have become important instruments for various aspects of higher education (e.g. reforming university/department curricula, faculty recruitment, promotion and wages, research funding, student admissions and tuition fees, a student's future job prospects; Clarke 2007, Salmi & Saroyan 2007, Rauhvargers 2011, 2013). As a result, they are being heavily advertised and covered by the media (e.g. international and national magazines and newspapers, TV, radio, internet media and blogs) as soon as they are released by the competing companies. Their publication is also accompanied by press releases and public gloating from universities or countries ranked at the top of the lists (e.g. www.nytimes.com/2010/11/15/education/15iht-educLede15.html?pagewanted=all). Not least, they trigger reactions at different governmental levels (e.g. with the release of the 2012 rankings, Putin announced \$2.4 billion for the innovation of the Russian higher education system over the next 5 yr: www.nytimes.com/2012/03/26/world/europe/russia-moves-to-improve-its-university-rankings.html?pagewanted=all&_r=0; see also Salmi & Saroyan 2007).

REACTION OF ACADEMICS TO IFS AND RANKINGS

Academics — including scientists, philosophers and even theorists — are humans, and as humans they

like numbers too. However, academics are pretty strange human beings: they like to criticize debate, comment, evaluate, reject and eventually propose alternatives to whatever becomes orthodoxy (e.g. Pimm 2001). In fact, it is these characteristic traits of scientists that lay at the very heart of scientific progress. In addition, most of them certainly know how to read numbers better than managers, administrators, politicians and journalists, and are aware of the dangers of reducing value to what can be counted numerically. Finally, they are especially trained in reading what lies behind those numbers, and in identifying patterns and propensities in them (e.g. Cury & Pauly 2000).

Thus, it is not surprising that academics received IFs and rankings with great skepticism, questioning both their estimation and their performance. The critical literature on IFs and rankings rapidly increased in the years following their emergence. For instance, a quick search in Scopus (24 June 2013) for articles with 'journal impact factor' and 'university rankings' in their title produced 657 scientific articles, with a total of 7129 citations ($h = 34$, i.e. 34 articles have >34 citations; Hirsch 2005) and 200 scientific articles that overall received 1057 citations ($h = 16$), respectively (i.e. an average IF of about 11 and 5). The number of the above-mentioned articles on IFs increased from <20 yr⁻¹ from 1985 to 2001 to a maximum of about 75 articles yr⁻¹ in 2010 to 2012. Similarly, the number of articles on URs/GURs increased from <3 yr⁻¹ from 1978 to 2004 to a maximum of about 30 articles yr⁻¹ in 2010 to 2012.

Among other things, scientists questioned (1) the estimation of IFs over a very short time period (2 yr), which does not allow enough time to really capture the impact of a publication, (2) the limited coverage of existing peer-reviewed journals and the practically non-coverage of conference proceedings and books, which are extremely important for disciplines such as mathematics, computer sciences, social sciences and the humanities, (3) the English language dominance, and (4) the practice of using IFs as a measure to evaluate scientists and their research, as well as for comparing between disciplines (e.g. Seglen 1997, Garfield 1999, Dong et al. 2005, Church 2011; see also various contributions in Browman & Stergiou 2008). Scientists also noted that IFs can quite easily be manipulated by the editors who can make decisions that increase the perceived IF of their journal: (1) deciding to publish more reviews, which are generally cited more often than 'research' articles; (2) increasing the number of self-citations to the journal, i.e. asking authors to cite more papers from

³The 'Webometrics Ranking of World Universities' publishes rankings every 6 months.

their journal; and (3) extending the type of citable material (e.g. Dong et al. 2005, Alberts 2013, Misteli 2013). When IF becomes the panacea in academia, as the gold medal is for the Olympic Games, then undoubtedly and inevitably doping will become part of the game. Indeed, the percentage of articles retracted because of fraud has increased 10-fold since 1975 (Fang et al. 2012). In addition, Fang & Casadevall (2011) examined the retraction rate for 17 medical journals, ranging in IF from 2.00 to 53.48, and found that the journal's retraction index (i.e. the number of retractions from 2001 to 2010, multiplied by 1000, and divided by the number of published articles with abstracts) was highly ($p < 0.0001$) correlated with the journal's IF. Liu (2006) and Steen (2011) provide more examples of positive relations between retracted papers and journals' IFs.

Similarly, rankings have also been heavily criticized for (1) many methodological issues related to the indicators used and their weightings, (2) English speaking countries dominating the rankings, (3) teaching quality being hard, if at all, to measure, and (4) arts, humanities and social sciences being relatively under-represented (e.g. Enserink 2007, Salmi & Saroyan 2007, Harvey 2008, Rauhvargers 2011, 2013, Shin & Toutkoushian 2011). As Usher & Savino (2007, p. 13) aptly state: '*In fact, most indicators are probably epiphenomena of an underlying feature that is not being measured.*' In addition, rankings have been also criticized for their '*symbolically violent character as a form of social categorization and hierarchization*' (Amsler 2013 this Theme Section). And just as for IFs, they can effectively be 'manipulated' (1) by favoring specific science and bio-science disciplines, (2) by discontinuing programs and activities that negatively affect performance, (3) by identifying weak performers and rewarding faculty for publishing in high IF journals (see Hazelkorn 2009; Table 1) and (4) by not admitting more low-income students from urban public schools who might lower the retention and completion rates (McGuire 2007).

It is true that both Thomson Reuters producing IFs and the companies/institutions producing rankings respond to criticisms. Thus, Thomson Reuters started to release the 5 yr IF, and their database was expanded to cover more journals as well as conference proceedings and books (<http://thomsonreuters.com/web-of-science/>). Similarly, companies and institutions producing rankings change their methodology almost annually, partially in response to critics (e.g. Enserink 2007, Rauhvargers 2013, Baty 2014 this Theme Section).

Last, but not least, IFs, rankings (Abbott 1999, 2011, Bornmann 2011) and not least anonymous peer reviewing (Espeland & Sauder 2009, Sauder & Espeland 2009) can breed academic/intellectual conservatism and, indeed, populism as they provide incentives to write or do what is assumed to please (or at least not put off) reviewers, especially reviewers of high impact factor journals with high rejection rates and hence of high reputation. At least in the social sciences, part of the reviewing is less concerned with academic quality than with the 'fit' of what an author says with current academic conventions, fashions, paradigms, etc. From a scientific viewpoint, this is the last thing academia would want to encourage.

RESISTING AND BOYCOTTING IFS AND RANKINGS

Eventually, after >30 yr since their inception, IFs—a simplified numeric expression meant to evaluate journals but misapplied in the evaluation of scientific performance (Polderman 2007)—were recently banned as 'evaluations' of individual scientists, individual articles, in hiring/ promotion and in the distribution of funding. Thus, at the 17 May 2009 meeting of the International Respiratory Journal Editors' Roundtable it was decided that IFs '*should not be used as a basis for evaluating the significance of an individual scientist's past performance or scientific potential*' (Russell & Singh 2009, p. 265). Three years later, scientists at the December 2012 meeting of the American Society for Cell Biology released the San Francisco Declaration on Research Assessment (DORA) (<http://am.ascb.org/dora/>, p. 1) in which it is again stated that '*the impact factor must not be used as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions*'. DORA also provides detailed recommendations to funding agencies, institutions, publishers and the organizations that supply metrics for improving assessment of scientific publications (see <http://am.ascb.org/dora/>). As Alberts (2013, p. 787), the editor of the journal *Science*, puts it:

The DORA recommendations are critical for keeping science healthy. As a bottom line, the leaders of the scientific enterprise must accept full responsibility for thoughtfully analyzing the scientific contributions of other researchers. To do so in a meaningful way requires the actual reading of a small selected set of each researcher's publications, a task that must not be passed by default to journal editors.

In addition, DORA (p. 3) calls individual scientists to be actively engaged in such a boycott:

When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics. Wherever appropriate, cite primary literature in which observations are first-reported rather than reviews in order to give credit where credit is due. Use a range of article metrics and indicators on personal/supporting statements, as evidence of the impact of individual published articles and other research outputs. Challenge research assessment practices that rely inappropriately on Journal Impact Factors and promote and teach best practice that focuses on the value and influence of specific research outputs.

The DORA recommendations were originally signed by 155 scientists and 78 scientific organizations, including the Academy of Sciences of the Czech Republic, the European Association of Science Editors, many scientific societies and journals, the Higher Education Funding Council for England and the American Association for the Advancement of Science. As of 20 August 2013, DORA has been signed by 9008 individual scientists and 367 organizations. The analysis of the data on those who signed DORA as of 24 June 2013, showed that '6% were in the humanities and 94% in scientific disciplines; 46.8% were from Europe, 36.8% from North and Central America, 8.9% from South America, 5.1% from Asia and the Middle East, 1.8% from Australia and New Zealand, and 0.5% from Africa' (<http://am.ascb.org/dora/>).

This ban, which was expressed in a common voice by journal editors, representatives from funding agencies, research institutions, associations and individual scientists, appeared in many common editorials (see e.g. Alberts 2013, Misteli 2013). In the end, as Tsikliras (2008) puts it, the rhetorical question of whether or not an article in *Nature* is better than 30 articles in the *Journal of the Marine Biological Association of UK* will never be answered objectively.

University rankings, global or not, like IFs (but much sooner, possibly because of their larger impact on higher education and society at large), have also led to several boycotts throughout the world. Thus, after the publication of the 1997 and 1998 rankings of universities in the Asian and Pacific region, 35 universities refused to participate in the 1999 survey and as a result the initiative was terminated (Salmi & Saroyan 2007). Similarly, 11 universities decided to not participate in the Maclean's 2006 rankings (Salmi & Saroyan 2007). Patricia McGuire, the president of Trinity University (Washington DC), boycotted U.S. News & World Report rankings: '*Rip it up and throw it away. That's the advice I'm giving my fellow col-*

lege and university presidents this month as the 'reputation survey' from U.S. News & World Report lands on our desks. I am one of 12 presidents who wrote a letter urging colleagues to take a stand for greater integrity in college rankings — starting by boycotting the magazine's equivalent of the 'American Idol' voting process.' (McGuire 2007). Similarly, the dean of St. Thomas University School of Law in Miami Gardens, Florida, Alfredo Garcia, also boycotted the U.S. News & World Report rankings by refusing to fill out the survey. Garcia said, '*I have personally stood in front of The Florida Bar's standing committee on professionalism and attacked U.S. News & World Report because it does a disservice to groups like us that represent minorities ... Everybody decries the survey, but everyone participates in the survey. Boycotting is not going to solve matters, but I figured I would put my money where my mouth is.'* (Kay 2010). James Cook University in Townsville, Australia, one of the most influential institutions in marine and climate sciences (placed second in the world on climate change science, behind the Smithsonian Institute and ahead of NASA), also refused to take part in the World University Rankings because of bias against small specialist universities (Hare 2012). Its vice-chancellor, Sandra Harding, wrote '*highly focused research endeavours in marine and environmental sciences worked against it, as did its location in Townsville ... As individual institutions we are deeply complicit in this nonsense. I say: enough.'* (Hare 2012).

Publications of rankings have even led to lawsuits. Thus, '*In March 2004, two universities in New Zealand successfully sued the government to prevent the publication of an international ranking that found them poorly placed in comparison with their Australian and British competitors. The vice-chancellors were concerned that the rankings would negatively affect their ability to attract fee-paying international students. In the end, the government was allowed to publish only the rankings of the national tertiary education institutions without comparing them to their peer institutions overseas'* (Salmi & Saroyan 2007, p. 42).

Probably the most recent rejection of rankings is evident from the boycott of the German Centrum fuer Hochschulentwicklung (CHE — translation: Centre for Higher Education Development)⁴ rankings by German sociologists (Dörre et al. 2013; see the German Sociological Association statement in Appendix 1). By suggesting to be able to measure the relative quality of academic teaching at German universities by way of ranking the subjective satisfaction scores of a small sample of students (fre-

quently not more than 10% of the main unit) in different disciplines, the CHE ranking has been very effective during the last decade in contributing to the political construction of a landscape of 'good' and 'bad' universities. However, rather than being a reliable instrument in advising students of which university department to go to if they want to fare well, the CHE ranking has proved to be welcomed by politics and bureaucrats as a seemingly self-evident measure of 'excellence' and 'non-excellence' in academic teaching. In a system of higher education which, as in the German one, is ever more influenced by the power of numbers, teaching rankings are a further instance of producing an academic 'reality' of differences in quality which, by way of a self-fulfilling prophecy, eventually results in a cemented division of winners and losers.

THE WAY FORWARD

The consequences of such individual boycotting of rankings might be either favorable or harmful to the individual institution(s) (Salmi & Saroyan 2007). Many maintain that boycotting is not going to solve matters because 'rankings are here to stay' (see Amsler 2013). Yet, the same was true of IFs—but the wide global acceptance of the DORA declaration shows that boycotting can really 'solve matters'. As Amsler (2013) claims, ranking is not a professionally necessary or inevitable activity, and we should turn

away from the ranking business, not only for scientific, but also for ethico-political reasons. Thus, rankings are not 'here to stay' if we do not want them to. This will be realized if, and only if, an international declaration similar to DORA is signed by universities, faculty associations, scientific associations and individual scientists throughout the world, with the leading universities being among the first signers.

As Peter Murray-Rust (Cambridge) stated to Zoë Corbyn (2009)—regarding journal metrics, yet equally applicable to URs—'*Higher education has to take control of academic metrics if it is to control its own destiny ... it should determine what is a metric and what isn't*'. Probably (and hopefully), DORA and a potential DORA counterpart for university rankings, which could be triggered by the recent German Sociological Association statement (see Appendix 1), are the first steps on the road to realizing Murray-Rust's appeal.⁵ Yet, even if academics take control of metrics, the problem of measuring scientific quality remains. Simplified ranking and counting, even if organized by academics themselves, will still have serious limitations, and thus will not be the solution if the same type of power struggles and reputation games remain—and attention is restricted to what 'counts' in numerical terms.

Acknowledgements. The authors thank Athanassios Tsikliras, Volker Schmidt and 3 anonymous reviewers for useful comments and suggestions as well as the (former) board members of DGS/GSA, especially Uwe Schimank, for their contribution to the DGS/GSA statement.

LITERATURE CITED

- ⁴The CHE university ranking (CHE-Hochschulranking) provides rankings of higher education institutions in German-speaking countries for 35 subjects. It primarily addresses the needs of first-year students. It was published for the first time in 1998 in co-operation with Stiftung Warentest. From 1999 until 2004, the ranking was issued with the German magazine Stern. Since 2005 the rankings have been published by the German weekly newspaper *DIE ZEIT*. CHE is responsible for conception, data collection and analysis, whereas *DIE ZEIT* is in charge of publication, sales and marketing. In its public self-description, the CHE university ranking (1) is strictly subject-related (i.e. does not compare entire Higher Education Institutes); (2) is multi-dimensional, i.e. for each subject, no overall value is derived from predetermined weighted individual indicators; (3) takes into account facts about departments and study programs, the assessments of students on the study conditions and evaluation of the reputation of the departments by professors of the individual subjects; and (4) does not give an individual ranking position but provides 3 ranking groups, i.e. top, middle and end group. (www.che-ranking.de/cms/?getObject=644&getLang=en, accessed 26 August 2013).
- ⁵In the German case, the boycott of the CHE Ranking by sociologists has so far been followed by the scientific associations of historians, communication scientists, educational scientists, political scientists, Anglicists, and chemists.
- Abbott A (1999) Department and discipline: Chicago Sociology at one hundred. University of Chicago Press, Chicago, IL
- Abbott A (2011) Library research infrastructure for humanistic and social scientific scholarship in America in the twentieth century. In: Camic C, Gross N, Lamont M (eds) *Social knowledge in the making*. University of Chicago Press, Chicago, IL, p 43–87
- Alberts B (2013) Impact factor distortions. *Science* 340:787
- Amsler S (2013) University ranking: a dialogue on turning towards alternatives. *Ethics Sci Environ Polit* 13:155–166
- Baty P (2014) The Times Higher Education World University Rankings, 2004–2012. *Ethics Sci Environ Polit* 13: 125–130
- Bornmann L (2011) Scientific peer review. *Annu Rev Inform Sci Tech* 45:199–245
- Browman HI, Stergiou KI (2008) (eds) The use and misuse of bibliometric indices in evaluating scholarly performance. *Ethics Sci Environ Polit* 8:1–107
- Cameron BD (2005) Trends in the usage of ISI bibliometric data: uses, abuses, and implication. *Libraries and the Academy* 5:105–125

- Cheung WWL (2008) The economics of post-doc publishing. *Ethics Sci Environ Polit* 8:41–44
- Church S (2011) Journal impact factor. In: Mitchell GR (ed) *Measuring scholarly metrics*. Oldfather Press, University of Nebraska, Lincoln, NE, p 9–16
- Clarke M (2007) The impact of higher education rankings on student access, choice, and opportunity. *High Educ Eur* 32:59–70
- Corbyn Z (2009) Do academic journals pose a threat to the advancement of science? *Times higher education*. Available at: www.timeshighereducation.co.uk/407705.article (accessed 20 July 2013)
- Cury P, Pauly D (2000) Patterns and propensities in reproduction and growth of marine fishes. *Ecol Res* 15:101–106
- Dong P, Loh M, Mondry A (2005) The 'impact factor' revisited. *Biomed Digit Libr* 2:7, doi:10.1186/1742-5581-2-7
- Dörre K, Lessenich S, Singe I (2013) German sociologists boycott academic ranking. *Glob Dialogue* 3. Available at www.isa-sociology.org/global-dialogue/ (accessed 23 July 2013)
- Enserink M (2007) Who ranks the university rankers? *Science* 317:1026–1028
- Espeland WN, Sauder M (2009) Rankings and diversity. *South Calif Rev Law Soc Justice* 18:587–608
- Fang FC, Casadevall A (2011) Retracted science and the retraction index. *Infect Immun* 79:3855–3859
- Fang FC, Steen RG, Casadevall A (2012) Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci USA* 109:17028–17033
- Fersht A (2009) The most influential journals: impact factor and eigenfactor. *Proc Natl Acad Sci USA* 106:6883–6884
- Garfield E (1955) Citation indexes to science: a new dimension in documentation through association of ideas. *Science* 122:108–111
- Garfield E (1999) Journal impact factor: a brief review. *CMAJ* 161:979–980
- Gürüz K (2011) Higher education and international student mobility in the global knowledge economy (revised and updated 2nd edn). State University of New York Press, Albany, NY
- Hare J (2012) Uni boycotts rankings system. *The Australian*, May 23, 2012. Available at www.theaustralian.com.au/higher-education/uni-boycotts-rankings-system/story-e6frgcjx-1226363939248 (accessed 22 August 2013)
- Harvey L (2008) Rankings of higher education institutions: a critical review. *Qual High Educ* 14:187–207
- Harzing AW (2007) Publish or perish. Available at www.harzing.com/pop.htm (accessed 21 August 2013)
- Hazelkorn E (2009) Rankings and the battle for world-class excellence: institutional strategies and policy choices. *Higher Educ Manage Policy* 21:1–22
- Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 102:16569–16572
- Kay J (2010) Florida law school dean boycotts 'U.S. News' rankings survey. Available at <http://www.lawjobs.com/newsandviews/LawArticle.jsp?id=1202457535997&slreturn=20131014015418> (accessed 11 November 2013)
- Labi A (2008) Obsession with rankings goes global. *Chron Higher Educ*, October 17, 2008. Available at <http://chronicle.com/weekly/v55/i08/08a02701.htm> (accessed 23 July 2013)
- Liu SV (2006) Top journals' top retraction rates. *Scientific Ethics* 1:91–93
- Lynch K (2013) New managerialism, neoliberalism and ranking. *Ethics Sci Environ Polit* 13:141–153
- McGuire P (2007) Rank this, U.S. *News*. *Los Angeles Times*, May 14, 2007. Available at www.latimes.com/news/opinion/commentary/la-oe-mcguire14may14,1,7255278.story?ctrack=1&cset=true (accessed 21 August 2013)
- Misteli T (2013) Eliminating the impact of the impact factor. *J Cell Biol* 201:651–652
- Monastersky R (2005) The number that's devouring science. The impact factor, once a simple way to rank scientific journals, has become an unyielding yardstick for hiring, tenure, and grants. *Chron High Educ* 52:A12
- Ophof T (1997) Sense and nonsense about the impact factor. *Cardiovasc Res* 33:1–7
- Pimm SL (2001) The world according to Pimm: a scientist audits the earth. McGraw-Hill, New York, NY
- Polderman KSA (2007) Keep your hands off our impact factor. *Eur Sci Ed* 33:98–99
- Rauhvargers A (2011) Global university rankings and their impact. European University Association, Brussels (electronic version available at: www.eua.be)
- Rauhvargers A (2013) Global university rankings and their impact—Report II. European University Association, Brussels (electronic version available at: www.eua.be)
- Robinson D (2013) The mismeasure of higher education? The corrosive effect of university rankings. *Ethics Sci Environ Polit* 13:65–71
- Russell R, Singh D (2009) Impact factor and its role in academic promotion. *Int J Chron Obstruct Pulmon Dis* 4: 265–266
- Salmi J, Saroyan A (2007) League tables as policy instruments: uses and misuses. *Higher Educ Manag Policy* 19: 31–68
- Sauder M, Espeland WN (2009) The discipline of rankings: tight coupling and organizational change. *Am Sociol Rev* 74:63–82
- Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *BMJ* 314:497–502
- Setar L, MacFarland M (2012) Top 10 fastest-growing industries. Special Report April 2012. Available at: www.ibis-world.com (accessed 13 September 2013)
- Shin JC, Toutkoushian RK (2011) The past, present, and future of university rankings. In: Shin JC, Toutkoushian RK, Techler U (eds) *University rankings, the changing academy—the changing academic profession in international comparative perspective*. 3. Springer Science and Business Media BV, Heidelberg, p 1–16
- Smith DR (2007) Historical development of the journal impact factor and its relevance for occupational health. *Ind Health* 45:730–742
- Steen RG (2011) Retractions in the scientific literature: Do authors deliberately commit research fraud? *J Med Ethics* 37:113–117
- Tsikliras AC (2008) Chasing after the high impact. *Ethics Sci Environ Polit* 8:45–47
- Tsikliras AC, Stergiou KI (2013) What's on the (publication fee) menu, who pays the bill and what should be the venue? *Mediterr Mar Sci* 14: 363–364
- UNESCO (2000). *World Education Report 2000. The right to education—towards education for all throughout life*. UNESCO Publications, Rome
- Usher A, Savino M (2007) A global survey of university ranking and league tables. *High Educ Eur* 32:5–15
- Ware M, Mabe M (2012) *The STM report—An overview of scientific and scholarly journal publishing*, 3rd edn. International Association of Scientific, Technical and Medical Publishers, The Hague

Appendix 1. Deutsche Gesellschaft für Soziologie (DGS)/German Sociological Association (GSA) statement from June 2012 (long version)

Scientific Evaluation, Yes—CHE Ranking, No Methodological Problems and Political Implications of the CHE University Ranking

The results of the CHE (Centre for Higher Education Development) University Ranking, a subject-level classification covering a range of academic disciplines, have been published each spring since 1998. The ranking has acquired high public visibility by virtue of the fact that it has been published in the weekly newspaper *DIE ZEIT* and in the annual *ZEIT Studienführer* (Study Guide) since 2005.

Doubts about the professional quality of the CHE Ranking have been voiced repeatedly within the field of sociology since it was first implemented. However, in view of the informational needs of prospective students of sociology, sociological institutes have participated in the data collection for the ranking. Rather than neglecting to mention it here, we self-critically acknowledge that sociology and the social sciences have been officially represented on the CHE Advisory Board in the past and that they may not have exercised, and availed of, their influence and their supervisory responsibilities—or at least may not have done so effectively enough.

However, since the middle of last year, mounting professional and science-policy-related misgivings on the part of a number of sociological institutes have led to a rethink. In June 2011, the Institute of Sociology at the University of Jena—which had consistently received very good ratings from the CHE—decided that it no longer wished to participate in the CHE Ranking. This prompted the Board of the German Sociological Association (GSA) to undertake a thorough analysis of the CHE Ranking. After studying the available documentation and conducting a lengthy discussion with the representatives of the Centre for Higher Education Development responsible for the ranking, the GSA Board arrived at the appraisal and the recommendations documented below. At its meeting on 20 April 2012, the GSA Council endorsed this appraisal and unanimously adopted the recommendations ensuing therefrom.

Professional and Science-Policy-Related Appraisal of the CHE Ranking

Firstly, the CHE Ranking has a number of serious methodological weaknesses and empirical gaps. Secondly, the summary assessment practice and the specific publication formats of the ranking systematically invite misinterpretations. Both aspects will be discussed in greater detail here.

Professional Appraisal: Research Indicators

For a number of years, at least, the quality of the research conducted at the individual faculties was measured on the basis of publication databases that not only the German Council of Science and Humanities (*Wissenschaftsrat*), but, meanwhile, also the CHE itself,

deems to be an unsuitable, or—in the case of sociology, at least—an insufficiently meaningful indicator. As an alternative, the CHE now measures research performance on the basis of external research funding per (budgeted) academic staff member. When doing so—and without any further differentiation—Higher Education Pact positions, for example, which were created expressly not for research purposes but rather to cope with teaching loads, are also included in the divisor of the external funding values. In effect, this means that—in purely arithmetical terms—as the teaching load of an institute increases (in the area of teacher training, for example), its per capita research performance, which the CHE claims to ‘measure’, deteriorates. It is obvious that the universities particularly affected are those that, because of the region in which they are located, have taken in a large number of students within the framework of the Higher Education Pact. Thus, the ‘burden of proof’ of the quality of research of an individual institute is borne almost entirely by the subjective criterion of that institute’s research reputation among fellow academics at other—in the logic of the ranking, rival—institutions. Anyone who has ever participated in the CHE survey of professors will be aware of its lack of methodological sophistication and the undifferentiated nature of its contents. The informational value of such sweeping faculty-specific judgements for prospective students, as the intended target audience of the ranking, is definitely questionable.

Professional Appraisal: Teaching Indicators

For this specific target audience the central criterion for the choice of a possible study location is obviously the quality of teaching at the various sociological institutes. However, this indicator is measured by the CHE largely on the basis of a student survey characterised by: (a) low response rates (19.3% in sociology in the last round); (b) a small number of participants (at every third university, less than 30 students from the subject area in question); and (c) completely unexplained survey selectivity, with the result that the danger of responses biased by careless or inattentive response behavior is correspondingly high. The CHE is well aware of the fact that by no means all universities draw a genuinely random sample with a calculable probability of selection. Moreover, a self-administered questionnaire survey with no systematic reminders and no nonresponse study can claim practically no validity. By the end of his or her basic training in methodology at the latest, any student of sociology would recognize that the survey is simply absurd. Therefore, massive doubts must be expressed with regard to the results of the CHE student survey—which is often described in discussions about the ranking as an opportunity for student participation qua evaluation.

Moreover, important, if not decisive, parameters for the assessment of the study situation—parameters that cannot be influenced by the teaching staff—are not included in the analysis (or the evaluation) at all. These parameters include, for example, (a) the respective fac-

Appendix 1 (continued)

ulty-student ratio (the ratio of the teaching load of faculty employed in budgeted positions to the number of students), (b) the associated arithmetical (and actual) class sizes and (c) the efficiency of examination offices. Furthermore, the CHE forgoes the collection of qualitative data that are, or would be, extremely relevant for the assessment of the quality of teaching at the individual locations and for prospective students' choice of study programmes, for example, the areas of focus and specialization offered by the various sociological study programmes and the systematic linking of teaching with the research conducted at the institute in question—irrespective of the external-funding intensity or reputational standing of that research. Such an inadequate, extremely selective and factually misleading data situation renders absolutely untenable the construction of a ranking of institutes with regard to their teaching performance.

Science-Policy-Related Appraisal: Evaluation Practice and Publication Formats

The basic problem with the university ranking is that the Centre for Higher Education Development aims to construct a ranking of institutes with regard to their teaching performance, and actually 'succeeds' in doing so, namely by dividing sociological institutes on the basis of extremely doubtful data into 'good' and 'bad'—or 'better' and 'worse'—institutes, and listing them hierarchically with spurious accuracy. Because of the sweeping evaluation practice and simplistic modes of presentation, the publication formats of the ranking invite systematic misconceptions about the situation in sociology.

The CHE collects data for a total of approximately eighteen indicators of research and teaching quality in the field of sociology, and these indicators are also published in the online version of the ranking. However, for a description of the individual indicators and their derivation, readers are referred to the small print, which most people are unlikely to understand. In the print version published in *DIE ZEIT* and in the *ZEIT Studienführer* (Study Guide), however, these eighteen indicators are not combined to form indices. Rather, only 5 or 6 indicators are selectively presented. This fact is neither discernible from a cursory reading, nor is any explanation given for the selection that has been made. Moreover, for both the quality of research and the quality of teaching, only the subjective evaluations from what we have shown to be methodologically extremely questionable surveys are presented. In particular, the simplistic ranking by means of traffic-light symbols (recently modified to green, yellow and blue) obscures the remarkable paucity of the database; in some cases, a single binary-coded response to a questionnaire item can yield a traffic light symbol signalling 'good' or 'bad' performance. The CHE Ranking—willingly bowing to the presentational demands of the mass media—gives the impression of unequivocal, reliable assessments, which are by no means covered by the available data. Here, systematic differentiations and thick descriptions would clearly be indicated and appropriate.

It is indeed disturbing in itself that the CHE Ranking thus misleads the very group whose interests, according to its authors, it is primarily supposed to serve, namely prospective students of sociology, who could, indeed, benefit from having accurate information about individual study locations when choosing a university and a study programme. It is perhaps a blessing in disguise, therefore, that—as far as teachers of sociology can ascertain—hardly any of the students who are now studying sociology at German universities, at any rate, allowed themselves to be decisively influenced by the CHE Ranking. Obviously, only a small minority of prospective students take serious note of the ranking—and that is a good thing.

On the other hand—and quite apart from its lack of informational value—the CHE Ranking has a very problematic effect on science policy. Therefore, if we are to believe the declared intentions of its authors, the ranking serves *de facto* a purpose for which it was not 'actually' intended. However, in higher-education-policy reality, the CHE Ranking invites—or, indeed, practically demands—extremely simplistic interpretations on the part of faculty and university management and ministerial bureaucracies. This may lead to structural decisions that have grave consequences for sociology, as an academic discipline, and its study programmes at individual locations—decisions that may well be objectively unfounded.

In view of the danger of such political uses of the CHE Ranking, it appears all the more remarkable that the persons responsible for the ranking at the CHE are unwilling to limit themselves to an informational function—however incomplete and unsatisfactory its implementation may be. They maintain that they cannot do without the construction of a ranking of the sociological institutes in Germany. At the preliminary meeting with those responsible for the ranking at the CHE, the German Sociological Association representatives were told quite openly that it would not be possible for the discipline to satisfy its own informational intentions within the framework of the procedure organized by the CHE, while at the same time avoiding the obligatory assessment and ranking. Thus, it became quite clear to the GSA that the CHE at least accepts the possibility that the university ranking will be politicized. The authors of the ranking claim that it merely depicts existing differences in quality between the sociological faculties. However, in the opinion of the German Sociological Association, there are strong grounds for assuming that the CHE Ranking contributes significantly to the construction of 'difference' and, thus, to creating divisions in the university landscape in the field of sociology.

In the worst case, therefore, the ranking will act as a self-fulfilling prophecy in the long term. Faculties labelled on a supposedly sound empirical basis as 'good' or 'bad' may actually become so in the long run because of the structural policy decisions and—perhaps one day, after all—changing student flows prompted by their rankings. More than any other academic discipline, sociology is aware of the way in which such social definitions of situations influence action. It therefore feels both a scientific obligation to draw attention to the far-reaching

Appendix 1 (continued)

consequences of actions based on incorrect definitions of situations, and a scientific responsibility not to contribute to such consequences any longer.

Recommendations Concerning the Handling of the CHE Ranking

Firstly, because the CHE Ranking has serious methodological and empirical deficiencies, secondly, because it withholds vital information from prospective students, as its declared target audience, and, thirdly, because it gives rise to wrong decisions on the part of science-policy decision-makers, sociology must take a stand against this presentation of its teaching and research performance in the public sphere constructed by the media. On the basis of this appraisal and the justifications thereof outlined above, the Board and the Council of the German Sociological Association have arrived at the following recommendations:

1. Because our analyses and the discussion of the considerable methodological deficiencies with the CHE representatives responsible for the ranking yielded no prospect of significant improvements in the CHE Ranking in the future, we hereby declare that this evaluation does not meet the basic quality requirements of empirical social research. As a professional sociological society, we call on the sociological institutes at German universities not to give the impression any longer that they support an empirical procedure that sociology must reject on professional grounds. In concrete terms, this means that the sociological institutes should defend and explain this resolution and its professional justifications vis-à-vis their faculty and university managers and their students, and, in particular, that they should not take part in the collection of data for the next CHE Ranking of sociology.

2. The GSA calls on science-policy decision-makers at university and ministerial levels not to rely any longer on appraisals and information derived from the CHE Rank-

ing when deliberating on, and undertaking interventions for, the development of sociology at the discipline's various university locations. More reliable information than that provided by the ranking already exists; in individual cases, occasion-specific evaluations should be conducted, for which both suitable concepts and unbiased institutions are available.

3. As an empirically oriented social science discipline, sociology claims to be particularly competent in the assessment of all kinds of empirical social research—including evaluations such as the CHE Ranking. In the present case, this competency implies a responsibility to recommend other disciplines, which are perhaps less sensitive in this regard, not to participate in the CHE Ranking any longer. After all, the grave deficiencies and misuses of this ranking that have been observed in the case of sociology are equally characteristic of its application to other disciplines.

4. Sociology is a discipline that is proficient in evaluation in every sense of the word. For this reason, it made itself available in 2006 for a pilot study on the rating (and precisely not the ranking) of research performance conducted by the Council of Science and Humanities (Wissenschaftsrat). In a process characterised by considerable social and technical complexity, this scientific rating demonstrated in an exemplary way the minimum requirements that a reliable and valid scientific evaluation must fulfill. To further meet the specific and justified desire on the part of prospective students of sociology for assistance in choosing a course of study and a study location, the GSA will develop a publicly accessible information package, which will also feature descriptions of the sociology programmes offered by German universities.

This statement, a summary thereof, and the latest information on the GSA's science-policy initiative launched herewith are available online at www.soziologie.de/che. This statement is reproduced here with permission from the German Sociological Association.