Reproducibility of the Shanghai academic ranking of world universities results

Domingo Docampo

Abstract   This paper discusses and copes with the difficulties that arise when trying to reproduce the results of the Shanghai academic ranking of world universities. In spite of the ambiguity of the methodology of the ranking with regard to the computation of the scores on its six indicators, the paper presents a set of straightforward procedures to estimate raw results and final relative scores. Discrepancies between estimated scores and the results of the ranking are mostly associated with the difficulties encountered in the identification of institutional affiliations, and are not significant. We can safely state that the results of the Shanghai academic ranking of world universities are in fact reproducible.

Keywords   Shanghai · Ranking · Reproducibility · Results · ARWU

Introduction

Insofar as we live in a fully interconnected world, "global development is now largely a function of enabling countries and their people to function productively in the global economy and the network society" (Castells 2005, p. 19), hence the central role of innovation, ultimately rooted on the quality of the higher education system (Himanen 2005, p. 358), in the knowledge society. It is therefore widely recognized that a good education is of paramount importance for the success of individuals and countries in the network society.

The high value assigned by individuals to their education has extensively increased the demand for information about the quality of universities and higher education systems. University rankings constitute a logical response to that demand, and appear to help students, country officials and the public at large in making sense of a remarkably diverse higher education landscape. Among the truly worldwide higher education rankings, the Shanghai Jiao Tong University Institute of Higher Education Academic Ranking of World Universities (ARWU), also known as the Shanghai ranking, which ranks academic

D. Docampo (✉)
Universidad de Vigo, Atlantic Research Center for Information and Communication Technologies, Campus Universitario, 36310 Vigo, Spain
e-mail: ddocampo@uvigo.es

institutions on the basis of their research performance, has lately received a great deal of attention, because it is based on solid, transparent numerical data of research quality and quantity (Marginson 2005). The Shanghai ranking focuses exclusively on research and does not rely on subjective data. All its indicators are open to public scrutiny, since they measure either scientific production or individual excellence recognized by very prestigious awards or by a high number of citations. However, in spite of the conscious effort made to rank universities by comparing their research capacity and output, the ranking is increasingly being used as a stick to measure institutions, and not just in relation with research (Docampo 2011). According to the findings of Hazeltorn (2008), by and large Higher Education officials think that rankings, and particularly ARWU, force them to be more accountable, set strategic planning goals, and provide comparative information to students, parents and other stakeholders. They agree upon the fact that international rankings "provide a methodology—albeit the quality of the methodology is contested—by which institutions can benchmark their own performance and that of other institutions" (Hazeltorn 2008, p. 199).

The controversy around the Shanghai ranking has raised a number of criticisms to its methodologies and its choice of indicators. The ranking is clearly biased in favor of science and technology, favors English-speaking universities as English is the predominant language of academic publications, and uses overly simple bibliometric indicators (van Raan 2005a). This last remark concerning the misuse of bibliometric indicators was contested by the authors, but the dispute remains largely unsettled (Liu et al. 2005; van Raan 2005b).

Dehon, McCathie and Verardi (2010) warn us about one-dimensional measures that can be misleading due to oversimplification, while Zitt and Filliatreau (2007) point out that the Shanghai ranking essentially measures overall production, which favors large universities since the per capita indicator does not offset the strong size-dependency of the other measures, a shortcoming also pointed out by Kivinen and Hedman (2008) in their analysis of the Scandinavian universities in the ARWU ranking. The inherent defects that accompany any attempt to measure higher education performance raised a barrage of criticism from Billaut, Bouyssou, and Vincke (2009), who made use of multiple criteria decision making theory to conclude that the Shanghai ranking is not a pertinent tool to discuss the "quality" of academic institutions. Ioannidis et al. (2007) warn institutions not to focus just on the numbers being accounted for the ARWU indicators, since no measurement has perfect construct validity for the many faces of excellence.

It is therefore beyond any reasonable doubt that the Shanghai ranking, as any other attempt to capture in a few figures the essence of a highly complex institution, suffers from a number of shortcomings difficult to cope with. As reported by Van Parijs (2009), the website of Shanghai's ARWU, as displayed in November 2008, candidly recognized that its way of ranking universities suffered from "many methodological and technical problems". Nevertheless, it is also clear that "despite growing concerns about technical and methodological issues, there is a strong perception among university leaders that rankings help maintain and build institutional position and reputation" (Hazeltorn 2008, p. 195). Rankings are here to stay, and the Shanghai ranking in particular has not seen its popularity diminished in the past few years.

In scientific circles, the most damaging criticism to a classification that claims to be based upon objective and rigorous data is the one that states the irreproducibility of its results. Hence, after the serious concerns raised by Florian (2007) about the reproducibility of the Shanghai ranking results, in a paper that has not been refuted so far, it has become common currency among the published research that the results of ARWU can not be replicated, further igniting the case against the Shanghai ranking by diminishing its credibility among the recipients of its annual verdict. It is therefore of paramount

importance for institutions that worry about their reputation to settle the question for good: are or are not the ARWU results reproducible? If the answer is no, the loss of credibility of the ranking results would constitute an open invitation to disregard them as useless. If the answer is positive then university officials would be able to monitor the results of their institutions; not only the authorities of the institutions that are classified among the five hundred so called world class universities, but most importantly, also the policy makers of institutions that did not make it, who could then properly assess the gap between their universities and the ARWU list.

It is the aim of this paper to give a response beyond a reasonable doubt to the concern about the reproducibility of the Shanghai ranking results. In a previous contribution (Docampo 2008) I had found out that at least in the case of the numerically simpler indicator—the one concerning the number of highly cited (HiCi) authors—the way of dealing with the apparent inconsistency of the ranking scores noted by Florian was to change the way proportions are obtained to compute relative scores: once the highest scoring institution is identified, the one with the largest number of HiCi authors, relative scores of the other institutions are calculated not in direct proportionality of the number of HiCi authors, but in direct proportionality between the square roots of those numbers. That was the first successful attempt to uncover the ambiguous statement by Liu and Cheng (2005), "… the distribution of data for each indicator is examined for any significant distorting effect and standard statistical techniques are used to adjust the indicator if necessary". It was an empirical finding that provided a starting point to formulate the hypothesis that the same procedure would be used in the rest of the indicators.

This paper will show that we can indeed positively check the validity of the hypothesis that the procedure to compute relative scores on the HiCi indicator is also used to compute the scores of the remaining five indicators. Besides, other indicator-specific considerations shall be discussed. The paper will settle the issue of the reproducibility of the Shanghai ranking results by showing how the six indicators can be accurately replicated.

The rest of the paper goes as follows. First the background of the paper is covered by reproducing the methodology of the ranking. The five direct indicators (Alumni, Award, HiCi, N&S, and PUB) are then analyzed to show how the ARWU results can be accurately reproduced. Finally, the composed indicator (PCP) is dealt with and the paper is closed with the conclusions.

ARWU methodology

According to Liu and Cheng (2005), the six indicators that compose the Shanghai ranking are:

Alumni The number of alumni of an institution winning Nobel Prizes and Fields Medals. Alumni are defined as those who obtain a bachelor, master or doctoral degree from the institution. Different weights are set according to graduation times. The weight is 100 % for alumni obtaining a degree in 2001–2010, 90 % for the period 1991–2000, 80 % for the period 1981–1990, and so on, and finally 10 % for alumni obtaining a degree in 1911–1920. If a person obtains more than one degree from an institution, the institution is considered only once.

Award The number of staff of an institution winning Nobel Prizes in Physics, Chemistry, Medicine and Economics and Fields Medal in Mathematics. Staff is defined as those who work at an institution at the time of winning the prize. Different weights are set

according to the periods of winning the prizes. The weight is 100 % for winners in 2001–2010, 90 % for winners in 1991–2000, and so on, and finally 10 % for winners in 1911–1920. If a winner is affiliated with more than one institution, each institution is assigned the reciprocal of the number of institutions. For Nobel prizes, if a prize is shared by more than one person, weights are set for winners according to their proportion of the prize.

HiCi The number of highly cited researchers in 21 subject categories. The definition of categories and detailed procedures can be found at the website of Thomson ISI.

N&S The number of articles published in Nature and Science between 2006 and 2010. To distinguish the order of author's affiliations, a weight of 100 % is assigned for corresponding author, 50 % for first author (second author if the first and corresponding authors share the same affiliation), 25 % for the next author affiliation, and 10 % for other author affiliations.

PUB Total number of papers indexed in the Science Citation Index-Expanded and the Social Science Citation Index in 2010. Only "Articles" and "Proceedings Papers" are considered. When calculating the total number of papers of an institution, a special weight of two was introduced for papers indexed in the Social Science Citation Index.

PCP The weighted scores of the above five indicators divided by the number of full-time equivalent academic staff. If the number of academic staff for institutions of a country cannot be obtained, the weighted scores of the above five indicators are used. For ARWU 2011, the numbers of full-time equivalent academic staff are obtained for institutions in Australia, Austria, Belgium, Canada, China, Czech, France, Italy, Japan, Netherlands, New Zealand, Norway, Saudi Arabia, Slovenia, South Korea, Spain, Sweden, Switzerland, UK, USA etc.

Nobel prizes and Field's medals

To analyze the results on the indicators related to Nobel Prizes and Field's Medals, let us first point out the three major differences between the Award and Alumni indicators:

1. Only Laureates in the Sciences are taken into account in the Award indicator, while Literature and Peace prizes are accounted for the Alumni indicator as well.
2. The year of the award is what counts for the Award indicator, while the year of graduation is what counts for the Alumni indicator. In case of multiple graduation times from an institution—e.g. BS and PhD from the same university—it is the latest graduation time that counts.
3. In the computation of the Award indicator all the Field's Medalists get three points to their institution while Nobel Laureates get three points only when they do not share the prize. When a Nobel Prize is shared, the three points are distributed according to the partition of the prize. Points are shared in case of multiple affiliations. In the computation of the Alumni indicator, however, all the Nobel Laureates and Field Medalists get one point for all the institutions from which they graduated.

Harvard University achieves the maximum score on the two indicators in the 2011 edition of ARWU, with 37.88 points in the Award indicator and 28.90 points in the indicator Alumni. Let H be the number of points of Harvard, and X the number of points of any other institution. In both cases, Alumni and Award, estimated scores are computed through the same formula:

$$\text{EST} = 100\sqrt{\frac{X}{H}} \tag{1}$$

In the case of the Award indicator the results of the Shanghai ranking have been accurately matched, as Table 1 shows. Acronyms in the table stand for:

pts: Total number of points from each institution.
EST: Values of formula 1.
Award: Scores on the indicator Award from the ARWU web page.

Table 1 includes universities with different scores; in the case of universities with the same score, only one of the institutions is shown. Due to the difficulty in judging how the legacy of Paris-La Sorbonne should be distributed among its heirs, Universities Paris 5, 6, 7, 9 and 11 have not been included in the table; inasmuch as inaccuracies in those cases would be due to the incorrect identification of the institutions, their inclusion would not have been useful to test the estimation procedure.

In the case of the Alumni indicator almost all the results of the Shanghai ranking have been properly matched. Table 2 shows the results of the 25 highest scoring institutions. The acronyms 1st, 2nd, 3rd, used in Table 2 stand for the cases in which the Field's Medalist or Nobel Laureate was awarded his or her first, second or third degree, respectively. The rest of the acronyms in the table stand for:

pts: Addition of the previous three columns to compute the raw data.
EST: Values of formula 1.
Alumni: Scores on the indicator Alumni from the ARWU web page.

Table 2 shows only a few cases of clear inaccuracies, undoubtedly related to the demanding task of identifying all the graduates from those institutions, not to a shortcoming of the estimation procedure. Excel files containing the results of all the universities included in the Shanghai ranking on both the Alumni and Award indicators are available upon request.

Computing the HiCi indicator

The major difficulties in computing the scores on this indicator arise from the inaccuracies of the official information provided by Thomsom Reuters on the affiliations of the highly cited authors. When searching for the information about the highly cited authors from an institution the following problems have to be dealt with:

1. Outdated information. Researchers move and the new affiliation is not always registered in the official web page.
2. Some of the authors have unfortunately passed out in the past few years, but there names are still on the list.
3. Mistaken identities: a few of the authors have been mistakenly assigned to a different institution due to the difficulties in recognizing researchers with the same last name and initials.
4. The information about the institution is missing in a great deal of cases. We find the name of the research unit, hospital or institute, but not the institution to which they are affiliated.

Table 1 Scores on the Award Indicator

Institution	pts	EST	Award	Institution	pts	EST	Award
Harvard	37.93	100.0	100.0	Irvine	3.25	29.3	29.3
Cambridge	35.45	96.7	96.7	Jerusalem	3.00	28.1	28.1
Princeton	28.80	87.1	87.1	Karolinska	2.80	27.2	27.2
Chicago	26.70	83.9	83.9	Heidelberg	2.75	26.9	27.0
MIT	25.45	81.9	81.9	Wash. St. Louis	2.55	25.9	25.9
Berkeley	23.85	79.3	79.3	Buenos Aires	2.40	25.2	25.2
Stanford	23.33	78.4	78.4	ENS Paris	2.25	24.4	24.4
Caltech	17.98	68.8	68.8	Copenhagen	2.20	24.1	24.1
Columbia	17.21	67.4	67.4	Pisa	2.10	23.5	23.5
Rockefeller	12.93	58.4	58.4	Purdue	2.05	23.2	23.2
Oxford	12.60	57.6	57.6	Technion	2.00	23.0	23.0
Cornell	9.90	51.1	51.1	Munich	1.98	22.8	22.8
Yale	7.65	44.9	44.9	Sussex	1.95	22.7	22.7
Los Angeles	6.88	42.6	42.6	Strasbourg	1.90	22.4	22.4
San Francisco	6.10	40.1	40.1	Aarhus	1.85	22.1	22.1
Imperial College	5.25	37.2	37.2	Rice	1.80	21.8	21.8
Illinois	5.05	36.5	36.5	Freiburg	1.65	20.9	20.9
SFIT Zurich	4.95	36.1	36.1	Tsukuba	1.50	19.9	19.9
San Diego	4.85	35.8	35.8	Toronto	1.40	19.2	19.2
Wisconsin	4.75	35.4	35.4	Free Brussels	1.35	18.9	18.9
Santa Barbara	4.68	35.1	35.1	Helsinki	1.20	17.8	17.8
Kyoto	4.58	34.7	34.7	Tufts	1.05	16.6	16.6
Pennsylvania	4.45	34.3	34.3	Hokkaido	1.00	16.2	16.2
Moscow State	4.40	34.1	34.1	Roma	0.90	15.4	15.4
Manchester	4.35	33.9	33.9	Melbourne	0.75	14.1	14.1
Oslo	4.20	33.3	33.3	Louvain	0.70	13.6	13.6
TSMC at Dallas	4.15	33.1	33.1	Tech. Denmark	0.60	12.6	12.6
Carnegie Mellon	4.05	32.7	32.7	Utah	0.50	11.5	11.5
Uppsala	3.90	32.1	32.1	Innsbruck	0.45	10.9	10.9
Washington	3.80	31.7	31.7	Lisbon	0.30	8.9	8.9
Colorado	3.58	30.7	30.7	Mainz	0.25	8.1	8.1
Stockholm	3.30	29.5	29.5	Toulouse	0.15	6.3	6.3

5. When a reference to a hospital is made, it is not clear whether the author belongs to an academic institution as well: sometimes they do, sometimes they do not.
6. The trickiest problem is how to deal with fractional appointments, adjunct professorships, double affiliation for consulting purposes—particularly with universities in the Middle East with a generous budget to purchase in the new market of the highly cited authors—and positions in external units associated with academic institutions.

The best showcase for all these difficulties is Harvard University. It is of paramount importance to come up with the correct figure for the raw score of Harvard University, since the scores of all the other institutions will be related to it. A quick look at the

Table 2 Scores on the Alumni Indicator

Institution	1st	2nd	3rd	pts	EST	Alumni
Harvard	10.10	16.20	2.60	28.9	100.0	100.0
Cambridge	12.70	6.90	1.20	20.8	84.8	87.1
MIT	5.00	9.70	0.60	15.3	72.8	72.8
Columbia	6.10	7.00	0.90	14.0	69.6	69.6
Berkeley	4.70	7.50	1.30	13.5	68.3	68.3
Chicago	5.60	5.50	1.10	12.2	65.0	65.0
Princeton	2.30	7.00	0.00	9.3	56.7	56.7
Oxford	5.20	2.90	0.80	8.9	55.5	55.5
Caltech	2.60	5.10	0.30	8.0	52.6	52.6
Yale	3.80	2.80	0.50	7.1	49.6	49.6
Moscow State	6.40	0.00	0.00	6.4	47.1	47.4
Johns Hopkins	0.40	3.20	1.40	5.0	41.6	43.2
Cornell	1.90	3.20	0.00	5.1	42.0	42.0
Stanford	1.30	3.10	0.50	4.9	41.2	41.2
Tech University Munich	3.40	1.20	0.00	4.6	39.9	39.9
Jerusalem	3.90	0.00	0.00	3.9	36.7	36.7
Michigan	2.60	0.50	0.80	3.9	36.7	36.7
Pennsylvania	1.70	2.10	0.00	3.8	36.3	36.3
Carnegie Mellon	1.30	1.30	1.20	3.8	36.3	36.3
Wisconsin	1.20	2.40	0.00	3.6	35.3	35.3
Tokyo	3.30	0.30	0.00	3.6	35.3	35.3
Case Western Reserve	1.20	2.30	0.00	3.5	34.8	34.8
Frankfurt	2.50	0.90	0.00	3.4	34.3	34.8
Illinois	1.40	2.00	0.00	3.4	34.3	34.3
ETH Zurich	2.50	0.80	0.00	3.3	33.8	33.8

Thomson Reuters site results in a list of 225 possible highly cited authors related to Harvard University or to its associated teaching or research units. After a careful evaluation of all the cases the number of highly cited researchers currently affiliated with Harvard University in the year 2011 was set to 192. Therefore, to proceed further, the value of 100 was allocated by ARWU to Harvard in correspondence with the square root of 192, as the yardstick to measure the scores of the other institutions.

Computing the indicator for all the institutions would be a very tedious exercise, since we would have to prune the results from a database of highly cited authors that includes more than 7,000 researchers. Hence, a representative sample of institutions showing different numbers of highly cited authors has been selected to compute the estimated scores on the HiCi indicator.

The results are shown in Table 3, in which the acronyms stand for:

NHiCi: Number of highly cited authors from an institution (accounting for shared affiliations in some cases).
HiCi: Actual score on the indicator HiCi in ARWU 2011 of the institutions selected for the Table.
EST: Estimated score using the procedure explained in this section with $H = 192$:

$$\mathrm{EST} = 100\sqrt{\frac{\mathrm{NHiCi}}{H}} \qquad (2)$$

The table confirms that this indicator can be computed with absolute accuracy once the highly cited authors data are appropriately checked and corrected as necessary.

Computing the N&S indicator

In the case of the N&S indicator the methodology appears to be precise and clear; we can proceed by giving one point to the institution of the corresponding author, 0.5 points to the institution of the first author, 0.25 points to the institution of the next author, and finally 0.1 points to the remaining institutions. It is worth pointing out that an institution only scores once on each paper.

In order to test the square root hypothesis the first task is to evaluate the score of Harvard University. However, there is a number of papers during the period 2006–2010 coauthored by authors from that institution in excess of 700; hence, an attempt to compute the points awarded to Harvard by looking up all those papers is a doomed endeavor. Assuming the hypothesis holds, there is another way to arrive at the desired result by triangulating through the scores of universities with just a few papers in the period 2006-2010. To do so, we first compute the number of points for universities with a score below 7.0 in ARWU 2011. Since the scores are rounded up or down to the first decimal digit, the bounds for the number of points of Harvard university can be estimated as follows:

1. Suppose that an institution with a score y gets x points. Let H be the number of points of Harvard University. If the square root rule applies, $y = 100\sqrt{x/H}$, and H could then be estimated by reversing the formula:

$$H = x\left(\frac{100}{y}\right)^2 .$$

2. Now, since y has been rounded to the nearest first decimal digit, two bounds can be set for H, namely:

$$H_l = x\left(\frac{100}{y + 0.05}\right)^2$$

$$H_u = x\left(\frac{100}{y - 0.05}\right)^2$$

By computing the bounds of all the universities with a score below 7.0 in ARWU 2011 we arrive at an interval of values in which H must lie. The lower endpoint of the interval would be the upper value of all the H_l, L, and the upper one would be lower value of all the H_u, U. Provided that $L < U$, the center of the interval (L, U) would be an accurate estimation of the value of H. We can then test our hypothesis by computing the number of points of a number of universities in ARWU and evaluating the scores according to the square root formula. Depending on the results, we will know whether the rule remains correct.

Table 4 shows the number of points of institutions covering all the scores from 1.5 to 6.9 in ARWU. The acronyms in the table stand for:

Table 3 Scores on the HiCi Indicator

Institution	NHiCi	EST	HiCi
Harvard University	192	100	100
Princeton University	74	62.1	62.1
Yale University	68	59.5	59.5
University of Washington	58	55.0	55.0
University of Texas at Austin	40	45.6	45.6
University of Tokyo	34	42.1	42.1
University College London	30	39.5	39.5
University of Toronto	29	38.9	38.9
Karolinska Institute	20	32.3	32.3
Utrecht University	17	29.8	29.8
Leiden University	15	28.0	28.0
Osaka University	14	27.0	27.0
University of Zurich	12.5	25.5	25.5
Lund University	12	25.0	25.0
Tel Aviv University	11	23.9	23.9
University of Basel	10.5	23.4	23.4
University of Delaware	10	22.8	22.8
University of Wuerzburg	9.5	22.2	22.2
Tohoku University	9	21.7	21.7
University of Sheffield	8.5	21.0	21.0
University of Queensland	8	20.4	20.4
University of Frankfurt	7	19.1	19.1
University of Milan	6.5	18.4	18.4
Ghent University	6	17.7	17.7
Paris Sud University	5	16.1	16.1
University of Freiburg	4.5	15.3	15.3
Indian Institute of Science	4	14.4	14.4
City University of Hong Kong	3.5	13.5	13.5
Autonomous University of Madrid	3	12.5	12.5
University of Lausanne	2.5	11.4	11.4
University of Cape Town	2	10.2	10.2
University of Western Ontario	1.5	8.8	8.8
Seoul National University	1	7.2	7.2

pts: Results of the computation once all the papers from those institutions have been looked up.

N&S: True scores on the N&S indicator in ARWU 2011.

LOWB: Lower bound for the estimation of H, the number of points of Harvard University, as discussed above.

UPRB: Upper bound for the estimation of H.

The final row of the table shows the tightest bounds for H (L and U), maximum value of column LOWB and minimum value of column UPRB, respectively.

Table 4 Bounds for the number of points of Harvard University

Institution	Points	N&S	LOWB	UPRB
Univ Zaragoza	0.10	1.5	416.2	475.6
King Saud University	0.20	2.1	432.7	475.9
Cairo University	0.25	2.4	416.5	452.7
Sichuan University	0.30	2.6	427.2	461.4
Yamaguchi University	0.35	2.8	430.9	462.8
Federal University of Sao Paulo	0.45	3.2	426.0	453.5
University of Malaya	0.50	3.4	420.1	445.5
University of Greifswald	0.55	3.5	436.4	462.1
Massey University	0.70	4.0	426.8	448.6
Kyungpook Natl University	0.80	4.3	422.8	442.9
Shandong University	0.90	4.5	434.7	454.5
Texas Tech University	0.95	4.7	421.1	439.4
Huazhong University of Science and Technology	1.00	4.8	425.1	443.2
Xian Jiao Tong University	1.10	5.0	431.3	448.9
University of Surrey	1.20	5.2	435.4	452.4
University of Parma	1.25	5.3	436.7	453.5
Jilin University	1.30	5.5	422.0	437.7
The Medical University of South Carolina	1.35	5.6	422.9	438.3
University of canterbury	1.40	5.7	423.4	438.6
University of Lisbon	1.45	5.8	423.7	438.6
University of Vigo	1.50	5.9	423.7	438.3
Erasmus University	1.55	6.0	423.5	437.8
Georgetown University	1.60	6.1	423.0	437.1
Indian Institute of Science	1.70	6.2	435.2	449.5
University of Ljubljana	1.80	6.4	432.7	446.4
Lanzhou University	1.85	6.5	431.2	444.7
Griffith University	1.90	6.6	429.6	442.9
Hahyan University	1.95	6.7	430.4	443.5
Szeged University	2.00	6.8	430.5	443.4
University of Twente	2.10	6.9	430.7	443.3
			L	U
Final Bounds for Harvard University			436.7	437.1

Given that $L < U$, we can proceed further and check the validity of the square root hypothesis. Let first H be the value of the center of the interval (L, U), $H = 436.9$. We shall use that value to estimate the scores of a number of institutions and check whether the estimates match the true values. The validity of the square root rule will be checked for institutions scoring between 7.0 and 9.5 in ARWU 2011 along with universities from Spain.

Let "pts" be the number of points obtained by an institution. To compute scores on the N&S indicator, the formula will again be

$$\text{EST} = 100\sqrt{\frac{\text{pts}}{H}} \qquad (3)$$

A total of 23 institutions have been analyzed to test the estimation methodology. All the articles in Science or Nature with at least one author from any of those institutions have been identified through the Web of Knowledge.

Table 5 shows the number of papers of the universities under analysis, and the position of the institution affiliation within the list of authors.The acronyms of the columns of Table 5 stand for:

NP: Number of articles in Science or Nature between 2006 and 2010.
CA: Number of articles as corresponding author.
FA: Number of articles as first author.
NA: Number of articles as next author.
OA: Number of remaining articles.
pts: Total number of points of each institution on the indicator N&S.
EST: Estimated score on N&S of each institution.
N&S: True value of indicator N&S of each institution in ARWU 2011.

As Table 5 shows, the estimated score coincides with the one assigned by ARWU in all the cases. We can safely state that the N&S indicator is indeed reproducible using the procedure described in this section.

Computing the PUB indicator

It is obvious, by just taking a quick look at the number of articles published every year by each institution, that the method used for assigning scores to the PUB indicator follows the square root rule as well. Using a rough approach we could then get a first estimation of the points awarded by ARWU. There are, however, other problems associated with the computation of the PUB indicator. The problems mainly arise when searching the Web of Knowledge, since a great deal of universities show a variety of affiliation names, and taking them all into account is a very demanding task.

Besides the difficulties that arise when trying to properly identify affiliations in the web of knowledge, which constitute a disturbing source of noise in the process of assigning papers to institutions, it is not easy to interpret the meaning of the "special" weight of two introduced by the authors of the ranking for papers indexed in Social Science Citation Index (SSCI), since we encounter papers indexed in the SSCI that are indexed in the Science Citation Index Expanded (SCIE) as well, and papers listed only in the SSCI.

To overcome the first hurdle a large sample of easily identifiable institutions has been analyzed. To come up with an answer for the meaning of the "special" weight a linear regression analysis was performed, using the number of papers indexed in different ways (as explained below) as predictors of the final score (dependent variable).

Let's first split the papers from an institution (articles and proceeding papers only) into three different sets, namely:

oc: Papers that are listed only in the SCIE.
cs: Papers that are listed in both the SCIE and the SSCI.
os: Papers that are listed only in the SSCI.

There are two extreme approaches to assign a special weight to the papers listed in the Social Science Citation Index: either a weight or 2 for just the os papers, or a weight of

Table 5 Estimation of N&S scores of selected institutions

Institution	NP	CA	FA	NA	OA	pts	EST	N&S
Korea University	4	2	0	0	2	2.20	7.1	7.1
University of Jyvaskyla	7	0	4	0	3	2.30	7.3	7.3
University of Turin	7	1	2	1	3	2.55	7.6	7.6
University of Buenos Aires	4	2	1	0	1	2.60	7.7	7.7
University of Bari	16	1	0	1	14	2.65	7.8	7.8
University of KwaZulu-Natal	7	1	2	2	2	2.70	7.9	7.9
University of Vermont	6	2	1	0	3	2.80	8.0	8.0
Lehigh University	8	1	2	3	2	2.95	8.2	8.2
University of Chile	11	1	1	4	5	3.00	8.3	8.3
University of Ferrara	9	2	1	0	6	3.10	8.4	8.4
University of Granada	9	2	1	1	5	3.25	8.6	8.6
Ben-Gurion University	5	3	0	1	1	3.35	8.8	8.8
University of Tokushima	8	2	1	3	2	3.45	8.9	8.9
Complutense Madrid	14	1	2	3	8	3.55	9.0	9.0
Rensselaer Polytechnic Institute	8	3	0	1	4	3.65	9.1	9.1
University of Manitoba	12	1	3	3	5	3.75	9.3	9.3
Yonsei University	10	2	2	2	4	3.90	9.4	9.4
University of Valencia	14	2	4	0	8	4.80	10.5	10.5
Autonomous Barcelona	21	2	2	5	12	5.45	11.2	11.2
University of Politecn Valencia	11	6	1	1	3	7.05	12.7	12.7
University of Pompeu Fabra	24	5	1	6	12	8.20	13.4	13.4
University of Barcelona	30	4	3	4	19	8.40	13.8	13.8
Autonomous Madrid	24	4	6	4	10	9.00	14.4	14.4

2 for all the SSCI listed papers, (os plus cs). There is a third possibility though, one that results in a different weight for the papers included in os and the papers included in cs. To explore the likelihood of the different hypotheses, a regression analysis was carried out in the case of the two extreme approaches mentioned before.

According to Tabachnick and Fidell (2007, p. 123), the minimum sample size requirement for a linear regression analysis depends on the number of independent variables. Those authors recommend a sample size in excess of $50 + 8N$ (where N stands for the number of independent variables). Since we have three variables (oc, cs, os), at least 74 institutions will have to be included in the analysis. All those institutions are listed below.

1: Aarhus; 2: Beijing Normal; 3: Brandeis; 4: Caltech; 5: Leuven; 6: Columbia; 7: Dalian Tech; 8: Dartmouth; 9: Duke; 10: Erasmus; 11: Jilin; 12: Kanazawa; 13: King Fahd; 14: Kobe; 15: Kyushu; 16: Med Vienna; 17: Michigan State; 18: MIT; 19: Nankai; 20: Nanyang Tech; 21: Natl Cheng Kung; 22: Natl Tsing Hua; 23: Natl Singapore ; 24: New York; 25: Oxford; 26: Peking; 27: Princeton; 28: Shandong ; 29: Sichuan; 30: Stanford; 31: Aberdeen; 32: Antwerp ; 33: Birmingham; 34: Bologna; 35: British Columbia; 36: Buenos Aires; 37: Calgary; 38: Calif Davis; 39: Calif Irvine; 40: Calif Los Angeles; 41: Calif Santa Barbara; 42: Cambridge; 43: Geneva; 44: Helsinki; 45: Kiel; 46: Koln; 47: Leeds; 48: Liverpool; 49: Manchester; 50: Melbourne; 51: Michigan; 52: Milan; 53: Munster; 54: New Mexico; 55: Nottingham; 56: Oslo; 57: Padua; 58: Pisa; 59: Queensland; 60:

Rochester; 61: Sao Paulo; 62: Siena; 63: Southern California; 64: Stockholm; 65: Texas Austin; 66: Tubingen; 67: Warwick; 68: Washington; 69: Wuerzburg; 70: Zurich; 71: Uppsala; 72: Vanderbilt; 73: Washington St Louis; 74: Xiamen.

Procedure to test the two extreme cases:

1. Compute the points awarded to Harvard University according to the selected case; let again H be the result of that operation.
2. Evaluate, according to their ARWU score, the points that should accrue to all the universities in the sample were the selected case the true one. Hence, and assuming the square root rule is in place,

$$PUB = 100\sqrt{\frac{\text{pts}}{H}} \Rightarrow \text{pts} = H\left(\frac{\text{PUB}}{100}\right)^2$$

3. Perform a linear regression analysis using the three variables (oc, cs, os) to predict the points computed in step 2.
4. Test the validity of the procedure by looking up the confidence intervals for the regression coefficients.

Table 6 shows the data gathered from the Web of Knowledge on the three variables (oc, cs, os) corresponding to the institutions included in the regression analysis. Column I points to the number assigned to the university in the list of institutions provided before.

A preliminary analysis was conducted to ensure no violation of the assumptions of normality, linearity and homoscedasticity. In the two extreme approaches already mentioned, the Normal P–P plots show points closely aligned to a straight diagonal line, suggesting no major violations from normality. Besides, no outliers were found, the residuals obey to no systematic pattern and conform to a rectangular distribution. We are now in a position to explore the results of the multiple regression analysis in both cases.

Let's begin by assigning the weight of 2 to all the papers included in the SSCI index to compute the raw score of universities (including Harvard University). In that case, the predictor coefficients for the three variables (oc, cs, os) should be (1, 2, 2).

The three variables explain a 100 % of the variance of the sample, $F(3, 70) = 300.56$, $p \ll 0.001$. The values of the predictor coefficients were (1.050, 1.556, 2.106), and the confidence intervals were:

$$\text{oc} : (1.045, \ 1.055); \text{cs} : (1.500, \ 1.612); \text{os} : (2.064, \ 2.148).$$

These results are not in line with the assumed (1, 2, 2) weighting scheme. In fact, they are very close to (1, 1.5, 2).

Let's now assign the special weight of 2 just to the papers listed only in the SSCI index. In that case, the predictor coefficients for the three variables (oc, cs, os) should be (1, 1, 2). The three variables again explain a 100 % of the variance of the sample, $F(3, 70) = 300.56$, $p \ll 0.001$. The values of the predictor coefficients were (0.956, 1.417, 1.917), and the confidence intervals were:

$$\text{oc} : (0.951, \ 0.960); \text{cs} : (1.365, \ 1.468); \text{os} : (1.879, \ 1.955).$$

Again, these results are not in line with the assumed (1, 1, 2) weighting scheme. The predictors remain very close to (1, 1.5, 2).

We are then entitled to check the possibility of a very special weighting scheme, in which the papers listed only in the SSCI would receive a weight of 2, while the papers listed in both the SSCI and the SCIE would just get a weight of 1.5. Using those weights to

Table 6 Scientific Production in 2010 of the institutions included in the regression analysis

I	oc	cs	os	I	oc	cs	os	I	oc	cs	os
1 :	2,481	205	220	26 :	4,018	153	137	51 :	5,323	776	762
2 :	1,118	72	44	27 :	2,390	103	271	52 :	2,692	85	48
3 :	259	39	74	28 :	2,694	39	8	53 :	1,411	60	87
4 :	2,760	40	43	29 :	1,187	48	7	54 :	1,145	107	134
5 :	2,669	255	312	30 :	1,558	441	568	55 :	1,968	241	377
6 :	3,784	665	658	31 :	2,535	147	152	56 :	2,047	311	257
7 :	1,862	20	6	32 :	1,118	88	114	57 :	2,731	129	137
8 :	751	112	142	33 :	1,103	211	290	58 :	1,669	54	37
9 :	3,658	513	403	34 :	1,669	144	142	59 :	2,859	423	497
10 :	499	208	389	35 :	4,918	498	548	60 :	1,664	233	161
11 :	2,239	9	5	36 :	2,731	82	62	61 :	6,166	445	194
12 :	872	22	8	37 :	2,512	237	189	62 :	866	31	53
13 :	454	3	10	38 :	872	278	310	63 :	2,476	320	399
14 :	1,134	20	44	39 :	2,335	200	255	64 :	1,326	122	216
15 :	2,769	49	9	40 :	1,862	665	690	65 :	2,468	292	568
16 :	1,239	48	15	41 :	2,476	81	210	66 :	1,933	87	91
17 :	2,227	242	486	42 :	2,468	372	413	67 :	1,103	109	244
18 :	4,110	187	254	43 :	2,692	125	109	68 :	4,918	684	528
19 :	1,597	27	13	44 :	1,968	254	201	69 :	1,404	60	58
20 :	2,535	62	207	45 :	4,240	45	67	70 :	2,547	230	275
21 :	2,298	171	94	46 :	1,239	98	135	71 :	2,512	202	172
22 :	1,409	35	37	47 :	1,404	184	270	72 :	2,337	232	361
23 :	3,651	207	273	48 :	2,547	132	136	73 :	2,840	292	319
24 :	2,326	305	642	49 :	4,018	323	459	74 :	1,352	18	28
25 :	4,621	364	637	50 :	1,411	548	476				

recalculate the value of H and the points awarded to all the universities, the three variables again explain a 100 % of the variance of the sample, $F(3, 70) = 300.56, p \ll 0.001$. The values of the predictor coefficients were $(1.003, \ 1.486, \ 2.012)$, and the 95.0 % confidence intervals were:

$$\texttt{oc} : (0.998, \ 1.007); \texttt{cs} : (1.433, \ 1.540); \texttt{os} : (1.972, \ 2.052).$$

This time, the results are in line with the assumed $(1, \ 1.5, \ 2)$ weighting scheme. The solution to our problem is then to assign a weight of 2 to papers listed only in the SSCI and a weight of 1.5 to papers listed in both the SCIE and the SSCI. Let now H be the number of points of Harvard University using those weights, and $(\texttt{oc}, \ \texttt{cs}, \ \texttt{os})$ the sizes of the three different sets of papers from an institution. To compute scores follow the already well known path:

$$\text{pts} = \text{oc} + 1.5\text{cs} + 2\text{os}; \qquad \text{EST} = 100\sqrt{\frac{\text{pts}}{H}} \qquad (4)$$

Table 7 shows the results of the estimation procedure for the sample of universities under analysis. Acronyms in the table stand for:

pts: Total number of points from Eq. 4.
EST: Estimation of the PUB indicator Eq. 4.
PUB: Value of the PUB indicator taken from the ARWU web site.

The results obtained are very accurate, with a mean square error of just 0.014, although small and non systematic errors in the computation may arise as Table 7 shows. Those errors can be caused by a number of reasons, ranging from the already mentioned difficulty in checking all the possible affiliations linked to an institution, to the fact that results in the WOK do not remain constant in time but depend on the day the searching takes place.

Computing the PCP indicator

There are a number of issues that make the computation of the PCP indicator very difficult. First of all, there are two ways of computing the indicator: "the weighted scores of the above five indicators divided by the number of full-time equivalent academic staff. If the number of academic staff for institutions of a country cannot be obtained, the weighted scores of the above five indicators are used" (Liu and Cheng 2005). A number of caveats are in place. First, the coefficients chosen for "weighting" the scores. Second, the number of full-time academic staff from an institution. Third, whether the square root rule applies or not in this case.

As of the weights used to compute the scores, it is not difficult to produce them in the case of the countries in which the authors do not make use of the full-time equivalent academic staff data. Multiple regression analysis was carried out to assess the weights applied to the first five indicators to produce the scores on the PCP indicator. The universities available for this regression analysis were the ones from Argentina, Brazil (all universities but Universidade de Sao Paulo), Chile, Croatia, Germany, Egypt, Finland, Greece, Hungary, India, Ireland, Israel, Mexico, Malaysia, Russia, Poland, Portugal, Singapore, and Turkey. They conform a sample of 90 universities. Again, a preliminary analysis was conducted to ensure no violation of the assumptions of normality, linearity and homoscedasticity. The Normal P-P plots show points closely aligned to a straight diagonal line, suggesting no major violations from normality. Besides, no outliers were found, the residuals obey to no systematic pattern and conform to a rectangular distribution.

The minimum sample size requirement for a linear regression analysis in this case is $50 + 8 \times 5 = 90$, since we have five predictors. So we have the appropriate sample size to carry out the regression analysis. The hypothesis is again that the square root rule does apply; however, a first attempt to use it directly resulted in an unlikely set of weights, very different from the expected (0.1, 0.2, 0.2, 0.2, 0.2). Since the indicators had been computed after taking the square root of the raw scores, the rule was tried with the weights applied to the squares of the scores, not to the scores themselves. This time results do conform with the expected weighting scheme. Multiple regression analysis shows that the five variables explain a 100 % of the variance in the PCP indicator, $F(5, 84) = 535.84$, $p \ll 0.001$. The values of the predictor coefficients were (0.105, 0.213, 0.213, 0.213, 0.213), and the 95.0 % confidence intervals were:

$$\texttt{Alumni}:(0.104,\ 1.006);\ \texttt{Award}:(0.211,\ 0.215);\ \texttt{HiCi}:(0.211,\ 0,216);$$
$$\texttt{N\&S}:(0.211,\ 0.216);\ \texttt{PUB}:(0.212,\ 0.214)$$

The results are in line with the assumptions, although an extra factor must be introduced to correct for the slight deviation from the nominal 0.1 and 0.2 weights; that enables us to

Table 7 Estimated and true PUB scores in ARWU 2011

Institution	pts	EST	PUB	Institution	pts	EST	PUB
Aarhus	3,228.5	49.23	49.3	Davis	4,834.0	60.24	60.2
Beijing N	1,314.0	31.41	31.5	Irvine	3,145.0	48.59	48.7
Brandeis	465.5	18.69	18.5	Los Angeles	7,198.5	73.52	73.6
Caltech	2,906.0	46.71	46.5	Santa Barbara	2,175.5	40.41	40.6
Leuven	3,675.5	52.53	52.5	Cambridge	5,624.0	64.98	65.1
Columbia	6,097.5	67.66	67.5	Geneva	1,777.5	36.53	36.6
Dalian T	1,904.0	37.81	37.7	Helsinki	3,467.0	51.02	51.2
Dartmouth	1,203.0	30.05	29.9	Kiel	1,388.5	32.29	32.2
Duke	5,233.5	62.68	62.5	Koln	1,945.0	38.21	38.4
Erasmus	1,589.0	34.54	34.6	Leeds	2,592.0	44.11	44.3
Jilin	2,262.5	41.21	41.0	Liverpool	2,146.0	40.14	40.1
Kanazawa	921.0	26.30	26.4	Manchester	4,258.5	56.54	56.6
King Fahd	478.5	18.95	18.9	Melbourne	5,127.0	62.04	62.1
Kobe	1,252.0	30.66	30.7	Michigan	8,011.0	77.55	77.7
Kyushu	2,860.5	46.34	46.5	Milan	2,915.5	46.79	46.9
Med Vienna	1,341.0	31.73	31.6	Munster	1,675.0	35.46	35.6
Michigan S	3,562.0	51.71	51.5	New Mexico	1,573.5	34.37	34.3
MIT	4,898.5	60.64	60.6	Nottingham	3,083.5	48.11	48.0
Nankai	1,663.5	35.34	35.3	Oslo	3,027.5	47.68	47.5
Nanyang T	3,042.0	47.79	47.7	Padua	3,198.5	49.00	49.1
N Cheng Kung	2,742.5	45.38	45.2	Pisa	1,824.0	37.01	37.1
N Tsing Hua	1,535.5	33.95	33.8	Queensland	4,487.5	58.04	58.1
N Singapore	4,507.5	58.17	58.1	Rochester	2,335.5	41.87	41.7
New York	4,067.5	55.26	55.4	Sao Paulo	7,221.5	73.63	73.7
Oxford	6,441.0	69.54	69.5	Siena	1,018.5	27.65	27.6
Peking	4,521.5	58.26	58.4	South Calif	3,754.0	53.09	53.2
Princeton	2,528.5	43.57	43.4	Stockholm	1,941.0	38.17	38.1
Shandong	2,464.5	43.02	43.1	Texas Austin	4,042.0	55.09	55.2
Sichuan	2,780.0	45.69	45.6	Tubingen	2,245.5	41.06	41.0
Stanford	6,567.5	70.22	70.3	Warwick	1,754.5	36.29	36.2
Aberdeen	1,554.5	34.16	34.0	Washington	7,000.0	72.49	72.4
Antwerp	1,512.0	33.69	33.5	Wuerzburg	1,610.0	34.77	34.9
Birmingham	2,801.5	45.86	45.8	Zurich	3,442.0	50.83	50.7
Bologna	3,039.0	47.77	47.7	Uppsala	3,159.0	48.70	48.8
Br Columbia	5,758.0	65.75	65.9	Vanderbilt	3,407.0	50.58	50.5
Buenos Aires	1,805.0	36.81	36.9	Washington	3,916.0	54.22	54.4
Calgary	2,823.5	46.04	46.2	Xiamen	1,435.0	32.82	32.9

estimate the results in the PCP indicator with a high degree of accuracy, in the case of universities from countries for which the authors do not make use of the full-time equivalent academic staff data, using the following formula:

$$\text{EST}^2 = \frac{1}{0.94}\left(0.1\text{Alumni}^2 + 0.2(\text{Award}^2 + \text{Hici}^2 + \text{N\&S}^2 + \text{PUB}^2)\right) \tag{5}$$

In the sample, the PCP indicator was predicted correctly in 80 out of the 90 cases. The square error in the other 10 cases was less than 0.005, and the average square error was 0.001. Undoubtedly those errors are caused by the rounding operation to the first decimal unit in the ARWU web site. The authors of the ranking use the true values but only publish the rounded ones.

By and large, there is no easy access to the number of equivalent full-time faculty of all the institutions listed in ARWU. Australian universities have been chosen as a showcase, since there are public and reliable data published by the Department of Education, Employment and Workplace Relations of the Australian Government (2011).

In the case of universities from Australia, the number of Full Time Equivalent Academic Staff used in ARWU seems to be the number of Full Time Equivalent Faculty at the levels above Senior Lecturer and Lecturer (Level C). Data from Australian universities are shown in Table 8, where the acronyms of the three columns stand for:

ASL: Full Time Equivalent number of Faculty above Senior Lecturer.
LLC: Full Time Equivalent number of Faculty at Lecturer Level C.
FTE: Full Time Academic Staff for the calculations of the ARWU data.

To get the actual PCP values we first compute the weighting sum of the squares of the first five indicators; let's call that value WSS.

Table 8 Full time equivalent staff of Australian universities in ARWU 2011

Institution	FTE faculty		
	ASL	LLC	FTE
Australian National University	521	303	824
Curtin University of Technology	371	272	643
Flinders University of South Australia	213	190	403
Griffith University	346	332	678
James Cook University	185	169	354
La Trobe University	225	264	489
Macquarie University	288	213	501
Monash University	676	684	1,360
Swinburne University of Technology	142	140	282
University of Adelaide	373	279	652
University of Melbourne	799	497	1,296
University of New South Wales	712	675	1,387
University of Newcastle	221	228	449
University of Queensland	685	481	1,166
University of Sydney	792	628	1,420
University of Tasmania	192	227	419
University of Technology, Sydney	235	299	534
University of Western Australia	433	312	745
University of Wollongong	265	212	477

Source: Australian Government, Department of Education, Employment and Workplace Relations

$$\text{WSS} = 0.1\text{Alumni}^2 + 0.2(\text{Award}^2 + \text{Hici}^2 + \text{N\&S}^2 + \text{PUB}^2) \tag{6}$$

Let's call WSSCT the value of WSS for Caltech, the university with the highest score in the indicator PCP. Let FTECT be the Full Time Equivalent Staff of Caltech. To estimate the PCP indicator of the university X the following operation will be carried out:

$$\text{EST} = 100\sqrt{\frac{\frac{\text{WSSX}}{\text{FTEX}}}{\frac{\text{WSSCT}}{\text{FTECT}}}} = 100\sqrt{\frac{\text{FTECT}}{\text{WSSCT}}}\sqrt{\frac{\text{WSSX}}{\text{FTEX}}} \tag{7}$$

To compute formula 7 the value of 276 has been assigned to the Full Time Equivalent Academic Staff of Caltech. Since it follows from the 2011 ranking data that WSSCT = 3,157.47, we have that

$$\sqrt{\frac{\text{FTECT}}{\text{WSSCT}}} = 0.29565 \Rightarrow \text{EST} = 29.565\sqrt{\frac{\text{WSSX}}{\text{FTEX}}} \tag{8}$$

Table 9 shows, for all the Australian universities in ARWU 2011, the values on the five indicators (Alumni, Award, HiCi, N&S, and PUB). The rest of the acronyms in the table stand for:

FTE: Full Time Academic Staff for the calculations of the ARWU data.
WSS: WSS sums from formula 6.
PCP: Values of the PCP indicator taken from the ARWU website.
EST: Estimates of the PCP indicator according to formula 8.

Table 9 Values of the PCP indicator (actual and predicted) of Australian Universities in ARWU 2011

Institution	Alu	Awd	HiCi	N&S	PUB	FTE	WSS	PCP	EST
Australian	15.6	12.6	33.9	26.3	43.5	824	802.7	29.2	29.2
Curtin Tech	0	0.0	0.0	10.5	29.4	643	194.9	16.3	16.3
Flinders	17.6	0.0	10.2	5.2	25.9	403	191.4	20.4	20.4
Griffith	0	0.0	0.0	6.6	30.3	678	192.3	15.7	15.7
James Cook	0	0.0	10.2	13.4	24.8	354	179.7	21.1	21.0
La Trobe	0	0.0	7.2	4.3	26.5	489	154.5	16.6	16.6
Macquarie	0	0.0	17.7	13.6	28.8	501	265.5	21.5	21.5
Monash	0	0.0	14.4	18.6	51.7	1,360	645.2	20.4	20.3
Swinburne Tech	0	0.0	10.2	10.5	18.6	282	112.1	18.7	18.6
Adelaide	16.6	0.0	10.2	12.2	38.3	652	371.5	22.3	22.3
Melbourne	19.5	14.1	25.0	21.1	62.1	1,296	1,063.1	26.8	26.8
New South Wales	0	0.0	20.4	13.3	53.1	1,387	682.5	20.7	20.7
Newcastle	0	0.0	12.5	6.1	29.3	449	210.4	20.3	20.2
Queensland	14.4	0.0	20.4	24.4	58.1	1,166	898.2	25.9	25.9
Sydney	16.6	0.0	19.1	19	60.5	1,420	904.8	23.6	23.6
Tasmania	0	0.0	7.2	11.8	26.0	419	173.4	19.0	19.0
Sydney Tech	0	0.0	12.5	2.4	22.5	534	133.7	14.8	14.8
Western Australia	15.6	14.1	22.8	13.9	42.2	745	562.9	25.7	25.7
Wollongong	0	0.0	7.2	8	27.5	477	174.4	17.9	17.9

Results from Table 9 show that the PCP indicator was computed correctly in 17 out of 19 cases. The square error in the other 2 cases was less than 0.005, and the average square error was 0.001. It is apparent again that those errors are caused by the rounding operation to the first decimal unit in the ARWU web site. The accuracy of the computation has been proved, so by reversing formula 7 one can get the value of the full-time academic staff of an institution that has been used by the authors of the ranking.

Conclusions

In spite of the barrage of criticism that the Shanghai ranking has raised due to the multiple shortcomings of its methodologies, it would not have attracted so much interest if the ranking had not been used so extensively as a rough guide to gauge the quality of research universities.

A painful feeling arises from having to acknowledge that the reputation of an institution is not good enough according to this or that classification, while, by sheer lack of information, there is no clear path to assessing why and in which way things should be improved. Rankings not based solely on hard data, that incorporate a certain degree of perceived reputation in the form of opinions, are doomed to replicate the status quo, and offer little help to new higher education institutions; meanwhile, rankings based on hard evidences enable universities to better assess their strengths and shortcomings, provided that the results of the ranking are reproducible.

Given that the Shanghai academic ranking fulfils the condition of being based upon hard data, it was of paramount importance to clarify a matter that has been the subject of strong criticism since the first edition of the ranking was released: the reproducibility of its results. As discussed in the paper, it has been taken at face value the statement that the ARWU results were in fact not reproducible. That helped in fueling the debate on the shortcomings of the ranking, while at the same time diminished its credibility among the scientific community at large. It is therefore important for the academic world to settle this issue for good so that we can move forward and discuss ways in which the Shanghai ranking could become a useful tool for academic policy makers.

To settle the issue a complete methodology to compute scores of universities on all the indicators that compose the ARWU ranking has been presented. The main finding of this paper is that the accuracy of the computed scores attests to the reproducibility of the results of the Shanghai academic ranking of world universities in all its indicators.

In searching for the solution to the problems posed in reproducing the results of the ranking some possible sources of error have also been identified, particularly in the case of the indicators related to the scientific output of the universities. In (Docampo 2012) the case of universities from Spain has been extensively analyzed. After carefully reviewing the publication databases, a number of examples of incorrect assignment of affiliations to institutions were identified; among those examples, the most striking one was the case of a very well known university, that clearly qualified to be among the 300–400 in the 2011 edition of ARWU, but did not even make it to the final list due to those identification problems. The acknowledgement of those issues, and the subsequent dialogue established between university authorities and ranking makers will hopefully result in the reparation of the damage made to the reputation of that university.

The presentation of the findings of this paper have been documented in order to be used as a guide to replicate the results of any Higher Education institution. All the necessary computations have been well described, including the final equations to render the scores

on each indicator. Those guidelines could then help university officials around the world to monitor the results in the six indicators from the Shanghai ranking, regardless of whether their institutions are listed among the five hundred world universities.

One of the best contributions that can be made to further the academic and political debate about university rankings and league tables is to focus, among other aspects, in reducing the uncertainty and clarifying the possible sources of errors when interpreting their results (Leyersdoff 2012). As of the Shanghai ranking, the proof of the reproducibility of its results clearly results in a significant reduction of uncertainty, which could help in making better informed decisions. The identification of particular sources of error (particularly those related to the use of all the affiliations corresponding to the same institution) could also help to fostering the necessary communication between institutions and ranking makers.

Methods summary

ARWU data on academic institutions were gathered directly from the Shanghai Jiao Tong University ARWU website, http://www.shanghairanking.com; data on the scientific production of the institutions analyzed in the paper were taken from the Web of Knowledge in October 2011. Estimation of scores were computed using different Excel files containing all the data from the ARWU and WOK websites. The multiple regression analyses were performed using SPSS Statistics 18.0. All Excel and SPSS files are available upon request.

References

Australian Government (2011) Australian Government, Department of Education, Employment and Workplace Relations: Staff 2010 Full Time Equivalence. Web address: http://www.deewr.gov.au/HigherEducation/Publications/HEStatistics/Publications/Pages/Staff.aspx.

Billaut, J. C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shangai ranking: an MCDM view. *Scientometrics, 84*(1), 237–263.

Castells, M. (2005). *The Network Society: From Knowledge to Policy*, chap The Network Society: From Knowledge to Policy (pp. 3–22). Washington: Johns Hopkins Center for Transatlantic Relations.

Dehon, C., McCathie, A., & Verardi, V. (2010). Uncovering excellence in academic rankings: a closer look at the Shanghai ranking. *Scientometrics, 83*(2), 515–524.

Docampo, D. (2008). International rankings and quality of the university systems. *Revista de Educación, Special Issue*, 149–176.

Docampo, D. (2011). On using the Shanghai ranking to assess the research performance of university systems. *Scientometrics, 86*(1), 77–92.

Docampo, D. (2012). Universities from Spain in the 2011 edition of the Shanghai ranking. Submitted to Revista de Educación. Working paper available at http://www.gts.tsc.uvigo.es/~ddocampo/ONLINE/docamporevistaeducacion.pdf.

Florian, R. V. (2007). Irreproducibility of the results of the Shangai academic ranking of world universities. *Scientometrics, 72*(1), 25–32.

Hazeltorn, E. (2008). Learning to live with league tables and ranking: the experience of institutional leaders. *Higher Education Policy, 21*, 193–215.

Himanen, P. (2005). *The Network Society: From Knowledge to Policy*, chap Challenges of the Global Information Society (pp. 337–372). Washington, DC: Johns Hopkins Center for Transatlantic Relations.

Ioannidis, J. P. A., Patsopoulos, N. A., Kavvoura, F. K., Tatsioni, A., Evangelou, E., Kouri, I., Contapoulos Ioannidis, D. G., & Liberopoulos, G. (2007). International ranking systems for universities and institutions: a critical appraisal. *BMC Medicine, 5*:30.

Kivinen, O., & Hedman, J. (2008). World-wide university rankings: a Scandinavian approach. *Sciento-metrics, 74*(3), 391–408.

Leyersdoff, L. (2012). World shares of publications of the USA EU-27 and China compared and predicted using the new Web of Science interface versus Scopus. *El profesional de la información, 21*(1), 43–49.

Liu, N. C., & Cheng, Y. (2005). Academic ranking of world universities: Methodologies and problems. *Higher Education in Europe, 30*(2), 127–136.

Liu, N. C., Cheng, Y., & Liu, L. (2005). Academic ranking of world universities using scientometrics: A comment to the "fatal attraction". *Scientometrics, 64*(1), 101–109.

Marginson, S. (2005) *There must be some way out of here. Tertiary Educ. Management Conference*. Perth: Keynote address.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics, 5th edn)*. Boston: Pearson Education, Inc., Allyn and Bacon.

Van Parijs, P. (2009). European higher education under the spell of rankings. *Ethical Perspectives, 16*(2), 189–206.

Van Raan, A. F. J. (2005a). Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 133–143.

Van Raan, A. F. J. (2005b). Reply to the comments of Liu et al. *Scientometrics, 64*(1), 111–112.

Zitt, M., & Filliatreau, G. (2007). *The World Class Universities and Ranking: Siming Beyond Status*, chap Big is (made) beautiful: Some comments about the Shangai ranking of world-class universities, Part Two, 141–160. Romania: UNESCO-CEPES, Cluj University Press.

Additional Data sources

ARWU Scores http://www.shanghairanking.com/ARWU2011.html.

Nobel laureates http://nobelprize.org/.

Fields Medals http://www.mathunion.org/index.php?id=prizewinners.

HiCi http://www.isihighlycited.com.

N&S http://www.webofknowledge.com.

PUB http://www.webofknowledge.com.