Journal of Higher Education Policy and Management, 2013
Vol. 35, No. 6, 574–585, http://dx.doi.org/10.1080/1360080X.2013.844670

Routledge
Taylor & Francis Group

Rectifying an honest error in world university rankings: a solution to the problem of indicator weight discrepancies

Kaycheng Soh*

50 Lorong 40 Geylang #07-29, The Sunny Spring, Singapore 398074

Discrepancies between the nominal and attained indicator weights misinform rank consumers as to the relative importance of the indicators. This may lead to unwarranted institutional judgements and misdirected actions, causing resources being wasted unnecessarily. As a follow-up to two earlier studies, data from the Academic Ranking of World Universities (ARWU), Quacquarelli Symonds World University Ranking (QSWUR) and Times Higher Education World University Rankings (THEWUR) for the top 100 universities were re-analysed. Discrepancies have again been found in the three ranking results, and the problem was rectified by applying score standardisation via *T*-scaling. The universities were then ranked on the new *T*-score, which accurately reflect the systems' indicator weights. Implications for future use of ranking results and refinement of the ranking procedure are discussed.

Keywords: academic excellence; discrepancy; regression; scaling

> It is the highest form of self-respect to admit our errors and mistakes and make amends for them. To make a mistake is only an error in judgment, but to adhere to it when it is discovered shows infirmity of character.
>
> – Dale Turner

When a world university ranking system propagates its indicator weighting scheme, it makes a promise to rank consumers that the total or overall score will be made up of the indicator scores in the announced proportions. When the indicator scores do not adhere to the intended weighting scheme, allowing for minor deviations that may not matter, there is a discrepancy between the promised and produced overall score. Given the benefit of the doubt, such errors may be honest errors which may not be known to the ranker for unknown reasons. When it is found, it needs to be admitted and amended, for the *highest form of self-respect* as suggested by Dale Turner.

In an article recently published in this *Journal* (Soh, 2013a), the ill effects of discrepancies between the *nominal* and *attained* weights in world university rankings were discussed, using the example from the *Times Higher Education World University Rankings* (*THEWUR) 100 under 50* (Times Higher Education, 2012). Another article published in the *Studies in Higher Education* (Soh, 2013b), further discussed the issue using the example from the *Universitas Ranking of National Higher Education Systems* (Williams, Rassenfosse, Jensen, & Marginson, 2012).

*Email: sohkaycheng@hotmail.com

Ill effects of discrepancies

It was pointed out in the two cited articles that discrepancies mislead rank consumers who are unaware of the problem into believing that the individual indictors are as powerful or predictive as the nominal weights assigned by the ranking agencies. For instance, in *THEWUR 100 under 50*, Teaching and Citation are assigned the same weighting of 30, thus having a ratio of 1:1 but, their corresponding beta-weights are 0.333 and 0.617, a ratio of 1:2, indicating that Citation is doubly powerful or predictive. In such cases, believing the nominal weights, a university trying to improve ranking by improving *Teaching* (which may be easier than improving *Citation*), will find its effort unrewarded in the next round of ranking. It should instead attempt to improve *citation* to be doubly rewarded for its effort*, if it is aware of the actual weighting attained.* As demonstrated in the two cited articles and the present one, gross discrepancies are far from sporadic.

When a university judges and is judged for its academic excellence, based on its ranking which in turn is based on the overall score, it is natural that those indicators given heavier weights by the ranker will be seen as more critical, and they suggest where future effort is needed to improve the ranking in the next round. Basically, this is a diagnosis of the strengths and weaknesses of the institution. What if an indicator given a heavy weight, turns out to have only half of its intended weight, as in the example above? Then, there is an erroneous diagnosis. In such a case, the university will direct its effort and resources to rectify the wrongly identified problems – literally barking up the wrong tree.

Misleading discrepancies between the nominal and attained indicator weights were identified here as being caused by the naïve practice of simply weighting the raw indicator scores and thereafter summing them to form the overall scores for ranking. This simplistic procedure satisfies simple minds but it does not take into account a critical statistical fact: when raw scores from distributions with different means and standard deviations are aggregated, scores from more widely spreading distributions (as indicated by greater standard deviations) are automatically weighted to be more powerful or predictive of the resultant total scores.

As an aside, an analogy may be appropriate here to bring out the essence of the problem. It is a truism that different currencies have different purchasing power parities (PPP). When two or more amounts are summed for a total taking them at their face values, the result is erroneous and misleading. For example, it does not make good sense to say USD50 + SGD50 = \$100. The two amounts need to be standardised to the same currency (metric), before being summed so as to avoid nonsensical outcomes, thus, USD50 + USD37.40 = USD87.40, alternatively, SGD63.45 + SGD50 = SGD113.45 (at the exchange rate for 27 June 2013). In the context of the present study, the indicator scores are *not based on the same metric* as they have different standard deviations and therefore need to be scaled to be the same. Incidentally, the *Economist*'s Big Mac Index is commonly used in social science research to get around the PPP problem as it makes the exchange rate more 'digestible'.

***T*-Scaling as a solution**

When this problem is identified and understood, the solution is obvious – ensure the scores have the same mean and standard deviation before summation. This can be achieved by standardising the raw scores. There are two steps to do this. First, all indicator scores should be transformed to the *z*-score distribution which has a mean of zero (0.00) and a standard deviation of unity (1.00). This, however, may cause interpretation difficulties, especially to the statistically uninitiated, because a mean of zero is not easy to imagine and the many

resultant negative scores are even more puzzling. Next, for ease of interpretation and some other purposes (e.g., to have all scores falling within a preferred range), the *z*-scores are scaled, arbitrarily, to have a chosen mean and standard deviation.

The arbitrariness of scaling scores to a preferred mean and standard deviation is illustrated by two well-known examples. First, the American College Testing scales all subject raw scores to a distribution with a mean of 8 and a standard deviation of 6, so that all scaled scores fall between 1 and 36. The next example, the Educational Testing Service scales the raw scores for the Standard Achievement Test to have a mean of 500 with a standard deviation of 100, thus having scores fall within 200 and 800 each for its language and mathematics subtests, yielding a distribution of combined scores varying from 400 to 1600. (Note that even the worst student gets 400, which substantively means zero.) With these examples, the arbitrariness of scaling is obvious, and the purpose of scaling is to ensure that scores for one subtest (analogous to an indicator in world university ranking) do not unduly outweigh those for another subtest (another indicator).

Present study

In the context of the present article, a *T*-scale with a mean of 50 and a standard deviation of 10 was chosen. The choice is arbitrary, of course, but it serves the purpose of equalising the raw indicator scores before they are weighted and summed to form the overall scores for ranking the universities. This is demonstrated by using the data gleaned from the three currently most popular world university ranking systems, that is, the Academic Ranking of World Universities 2011 (ARWU), the Quacquarelli Symonds World University Ranking 2012 (QSWUR) and the THEWUR 2012–2013 (TSL Education Ltd. 2012).

For each of the systems, the relative contributions of each indicator in terms of raw scores and *T*-scores will be compared, with reference to the nominal weights assigned by the system. Then, the top 100 universities will be given new rankings based on their *T*-scores, with indication of positive or negative gain in position. Finally, the influence on ranking is summarised.

Academic Ranking of World Universities (ARWU) 2011

Table 1 shows for ARWU 2011 (Shanghai Ranking Consultancy 2012), the nominal weights assigned to its six indicators. When their original indicator scores were entered into a multiple regression to predict the overall scores, the unstandardised b-weights show consistency with the nominal weights. This has to be the case since the b-weights are those used to calculate the total or overall scores. For illustration, to evaluate the relative contribution, ratio between the first and the other indicators are calculated and shown in the parentheses. The ratios among all indicator scores can be likewise calculated. As shown in Table 1, the unstandardised b-weights and the ratios are consistent with those for the assigned weights.

As the unstandardised b-weights do not show the relative predictive power of the indicators, it is necessary to look at the standardised beta-weights. As shown in Table 1, the beta-weights for the original raw scores have ratios somewhat different from those for the nominal weights. For instance, in comparison with Alumni, while award is overweighted by 0.4, the other indicators are all underweighted, specifically HiCi by 0.2, N&S by 0.3, PUB by 0.7 and PCP by 0.3. However, when the raw scores were *T*-scaled before weighting and summing, the beta-weights are now consistent with the nominal weights, actualising the intention of the ranking system. The discrepancies for the other indicators are much less.

Table 1. Multiple regression of ARWU 2011.

	Nominal weight	Original score		*T*-score
		b-weight	beta-weight	beta-weight
Intercept	—	0.002	0.000	0.000
Alumni	10 (1)	0.102 (1)	0.140 (1.0)	0.124 (1)
Award	20 (2)	0.205 (2)	0.334 (2.4)	0.250 (2)
HiCi	20 (2)	0.202 (2)	0.245 (1.8)	0.245 (2)
N&S	20 (2)	0.213 (2)	0.236 (1.7)	0.249 (2)
PUB	20 (2)	0.206 (2)	0.182 (1.3)	0.248 (2)
PCP	10 (1)	0.102 (1)	0.093 (0.7)	0.124 (1)

Notes: (1) Alumni = Alumni; Award = Staff winning Nobel Prizes and Field Medals; HiCI = highly cited researchers; N&S = articles published in *Nature* and *Science*; PUB = science citation index and social science citation index; PCP = per capita performance on the above. (2) Figures in parentheses are the ratios between the first and the other indicators, for example, award/alumni, and such.

The universities were ranked according to the new *T*-score total, as shown in Tables 2a and 2b. A 'zero' gain indicates no change between the original and the new rankings, for example, Harvard and Stanford. A positive gain indicates an upward movement in position, for example, California (Berkeley) moved from the original fourth to the new third. And, a negative gain indicates the opposite movement, for example, MIT dropped from the original third to the new fourth. It is of note that such changes (gains) are smaller for the high-ranking universities and increase down the line to the low-ranking

Table 2a. New ranking (1–50) of top 100 in ARWU.

Rank	Gain	University	Rank	Gain	University
1	0	Harvard	26	−1	Illinois at Urbana-Champaign
2	0	Stanford	27	1	Minnesota, Twin Cities
3	1	California, Berkeley	28	−1	Kyoto
4	−1	MIT	29	1	Northwestern
5	0	Cambridge	30	−1	New York
6	0	CalTECH	31	0	Washington in St. Louis
7	1	Columbia	32	3.5	Duke
8	−1	Princeton	33	−1	Colorado at Boulder
9	1	Oxford	34	3	British Columbia
10	1	Yale	35	−1.5	California, Santa Barbara
11	−2	Chicago	36	−0.5	Texas at Austin
12	0	California, Los Angeles	37	5	North Carolina at Chapel Hill
13	1	Pennsylvania	38	0.5	Maryland, College Park
14	−1	Cornell	39	2	Pierre and Marie Curie
15	1	Washington	40	−1.5	Manchester
16	−1	California, San Diego	41	2	Copenhagen
17	0	California, San Francisco	42	7	California, Davis
18	4	Michigan – Ann Arbor	43	2	Pennsylvania State
19	2	Tokyo	44	−10.5	Rockefeller
20	−2	Johns Hopkins	45	−1	Karolinska Institute
21	−1	University College London	46	−6	Paris Sud
22	−3	Wisconsin – Madison	47	−1	Southern California
23	1	Imperial College	48	9.5	Pittsburgh
24	2	Toronto	49	0	Utrecht
25	−2	Institute of Technology Zurich	50	3	Edinburgh

Table 2b. New ranking (51–100) of top 100 in ARWU.

Rank	Gain	University	Rank	Gain	University
51	−4	Technical Munich	76	6.5	Osaka
52	−3	California, Irvine	77	2	Utah
53	7	Melbourne	78	−3	Oslo
54	0	Munich	79	−6	Geneva
55	8	The Ohio State – Columbus	80	7	Queensland
56	8	McGill	81	−3	Arizona State – Tempe
57	−5	Vanderbilt	82	14	Sydney
58	−2	Zurich	83	7	Ghent
59	13	Florida	84	8	Michigan State
60	−9	Texas SW Medical Center	85	13	Tohoku
61	4.5	Leiden	86	−2	Rochester
62	−3	Rutgers, New Brunswick	87	−10	Moscow State
63	−2	Purdue, West Lafayette	88	12.5	Texas A&M – College Station
64	−2	Heidelberg	89	−2	Aarhus
65	0.5	Brown	90	−5	Nottingham
66	−8.5	Hebrew Jerusalem	91	−8.5	Indiana Bloomington
67	3.5	Australian National	92	−2	McMaster
68	2.5	Bristol	93	−3	Basel
69	−14	Carnegie Mellon	94	−7	Göttingen
70	−2	King's College London	95	−14	Stockholm
71	3	Helsinki	96	−1.5	Nagoya
72	−5	Uppsala	97	1	Sheffield
73	3	Boston	98	−3.5	Bonn
74	6	Arizona	99	1.5	Frankfurt
75	−6	Ecole Normale Superieure	100	−2	Case Western Reserve

Table 3. Changes in ranking in ARWU 2011.

Gain	Frequency
More	4
6 to 10	10
1 to 5	27
0	12
−1 to –5	36
−6 to –10	9
Less	2

ones. This also suggests that the ranking is more stable for the top end of the league table and it becomes increasingly less stable as the ranking goes lower. Thus, the ranking at the higher end is more reliable and trustworthy than at the lower end.

Table 3 summarises the gains (changes in positions) of the top 100 universities. As shown therein, there are 12 universities which retained their original rankings when the *T*-score was used. At the same time, 27 universities have positive gains between 1 and 5 positions, and 10 gain between 6 and 10 positions and four gained by more than 10 positions. Over the other end of the scale, 36 universities have negative gains between one and five positions, nine lose between 6 and 10 places and two more than that.

Quacquarelli Symonds World University Ranking (QSWUR) (2012)

When the same analysis was applied to the data from QSWUR 2012, a similar discrepancy was observed (Table 4), in fact, more severe than that found for ARWU 2011 above.

Table 4. Multiple regression of QSWUR (2012).

		Original score		*T*-score
	Assigned weight	b-weight	beta-weight	beta-weight
Intercept	–	0.021	0.000	0.000
Academic	40 per cent (1)	0.404 (1)	0.489 (1)	0.685
Employer	10 per cent (.25)	0.101 (.25)	0.205 (.42)	0.173 (.253)
Faculty/Students	20 per cent (.5)	0.202 (.5)	0.498 (1.02)	0.344 (.502)
Int. Faculty	5 per cent (.125)	0.051 (.126)	0.156 (.32)	0.085 (.124)
Int. Students	5 per cent (.125)	0.050 (.124)	0.136 (.28)	0.086 (.126)
Citation	20 per cent (.5)	0.202 (.5)	0.416 (.85)	0.344 (.502)

Note: Figures in parentheses are the ratios of the first to the other indicators.

Comparisons between the unstandardised b-weights and the standardised beta-weights for the indictors show up the discrepancies. For instance, taking Academic as a reference, four indicators (i.e., employer, faculty/students, international faculty and international students) are overweighted while citation is underweighted, when raw indicators were added to derive the overall. However, when the *T*-scores for the indicators were used to calculate the overall, the attained weights show that the original weighting scheme was maintained, disregarding the deviations in the third decimal values.

The new ranking and the corresponding gain for the top 100 universities are shown in Tables 5a and 5b. It is to be noted that the top nine universities retained their original

Table 5a. New ranking (1–50) of top 100 in QSWUR (2012).

Rank	Gain	University	Rank	Gain	University
1	0	MIT	26	−3	Hong Kong
2	0	Cambridge	27	3	The Tokyo
3	0	Harvard	28	7	Kyoto
4	0	University College London)	29	−3	King's College London
5	0	Oxford	30	6	Melbourne
6	0	Imperial College London	31	−4	Northwestern
7	0	Yale	32	0	Manchester
8	0	Chicago	33	−5	Bristol
9	0	Princeton	34	3	Seoul National
10	1	Columbia	35	4	Sydney
11	1	Pennsylvania	36	2	Wisconsin–Madison
12	−2	Caltech	37	−3	Ecole normale supérieure
13	2	Stanford	38	−5	Hong Kong S&T
14	−1	ETH Zurich	39	6	British Columbia
15	−1	Cornell	40	4	Peking
16	1	Michigan	41	2	New York
17	2	Toronto	42	6	Tsinghua
18	0	McGill	43	−14	Ecole Polytechnique Lausanne
19	−3	Johns Hopkins	44	−4	Chinese Hong Kong
20	2	California, Berkeley (UCB)	45	1	Queensland
21	−1	Duke	46	10	Illinois at Urbana-Champaign
22	−1	Edinburgh	47	5	New South Wales
23	8	California, Los Angeles	48	2	Osaka
24	0	Australian National	49	−7	Brown
25	0	National Singapore	50	−1	Carnegie Mellon

Table 5b. New ranking (51–100) of top 100 in QSWUR 2012.

Rank	Gain		Rank	Gain	
51	8	Washington	76	−9	Trinity College Dublin
52	8	Ludwig-Maximilians- München	77	1	Helsinki
53	−6	Nanyang Technological	78	−12	Sheffield
54	1	Heidelberg	79	6	Utrecht
55	15	California, San Diego	80	−16	Boston
56	12	Texas at Austin	81	−4	Birmingham
57	−4	Technische, München	82	−6.5	Tohoku
58	22	National Taiwan	83	−11	Nottingham
59	2	Monash	84	−5	Western Australia
60	−6	Glasgow	85	−12	Southampton
61	−10	Copenhagen	86	8	Leeds
62	−5	North Carolina, Chapel Hill	87	3.5	Zurich
63	−1	Amsterdam	88	0	Georgia Institute Technology
64	−6	Warwick	89	6.5	Purdue
65	−24	Ecole Polytechnique	90	−4	Nagoya
66	16	Katholieke, Leuven	91	−17	Geneva
67	4	Lund	92	8	California, Davis
68	22.5	Fudan	93	−1	Durham
69	14	Auckland	94	−5	Aarhus
70	−7	Korea Adv. Institute S&T	95	0.5	City Hong Kong
71	16	Freie, Berlin	96	−12	Washington in St. Louis
72	3.5	Leiden	97	−4	St Andrews
73	−4	LSE	98	−1	Pohang S&T
74	7	Uppsala	99	−1	Pittsburgh
75	−10	Tokyo Institute Technology	100	−1	Erasmus Rotterdam

rankings with neither gain nor loss in their position. It is also to be noted that positional changes occurred more drastically at the lower end of the league table, some as much as positive 22.5 (Fudan) and negative –24 (Ecole Polytechnique). Recall that this trend is also true of ARWU as mentioned earlier.

As Table 6 shows, 15 universities do not change when the *T*-score was used for ranking, 21 moved up by one to five positions and 20 by six or more positions. Over the other end, 27 dropped by one to five positions, nine by six to 10 positions and eight by more than that.

Times Higher Education World University Rankings (THEWUR 2012–2013) (TSL Education Ltd. 2012)

As for THEWUR 2012–2013, the discrepancies seem to be less severe (Table 7), but still visible nonetheless.

Table 6. Changes in ranking in QSWUR (2012).

Gain	Frequency
More	7
6–10	13
1–5	21
0	15
−1 to –5	27
−6 to –10	9
Less	8

Table 7. Multiple regression of THEWUR (2012).

		Original score		*T*-score
	Assigned weight	b-weight	beta-weight	beta-weight
Intercept	–	0.130	0.000	0.000
Teaching	30 per cent (1)	0.302 (1)	0.408 (1)	0.415 (1)
Research	30 per cent (1)	0.301 (1)	0.453 (1.11)	0.416 (1)
Citation	30 per cent (1)	0.299 (1)	0.362 (.887)	0.417 (1)
International Mix	7.5 per cent (.25)	0.073 (.24)	0.129 (.316)	0.104 (.25)
Industry Income	2.5 per cent (.08)	0.024 (.08)	0.050 (.123)	0.036 (.08)

Note: Figures in parentheses are the ratios of the first to the other indicators.

The first three indictors are supposed to be equally weighted at 30 per cent each, but when teaching is referenced to, Research is lightly overweighted and citation slightly underweighted. At the same time, both International Mix and Industry Income are obviously overweighted. Again, when *T*-scores for the indicators were used to calculate the overall, the relative importance of the indicators is maintained.

The less severe discrepancy suggests that there will be fewer changes in positions and this is confirmed by a comparison between the original and the new rankings as shown in Tables 8a and 8b. A cursory inspection of the tables shows that more universities retain their original rankings, and even when there are changes, they are smaller than those found in the other two systems. The greatest positive gain is 10 which brings California (Irvine)

Table 8a. New ranking (1–50) of top 100 in THEWUR (TSL Education Ltd. 2012).

Rank	Gain	University	Rank	Gain	University
1	0.0	Caltech	26	1.0	Tokyo
2	0.5	Stanford	27	4.0	Wisconsin–Madison
3	1.0	Harvard	28	0.0	Melbourne
4	−1.5	Oxford	29	1.0	British Columbia
5	0.0	MIT	30	2.0	Edinburgh
6	0.0	Princeton	31	7.0	California, San Diego
7	0.0	Cambridge	32	3.5	California, Santa Barbara
8	1.0	California, Berkeley	33	−4.0	National Singapore
9	1.0	Chicago	34	−1.0	Illinois at Urbana Champaign
10	−2.0	Imperial College London	35	−1.0	McGill
11	0.0	Yale	36	−10.5	Texas at Austin
12	1.0	California, Los Angeles	37	0.0	Australian National
13	2.0	Pennsylvania	38	−2.5	Hong Kong
14	0.0	Columbia	39	2.0	New York
15	−3.0	ETH Zürich	40	2.5	North Carolina at Chapel Hill
16	0.0	Johns Hopkins	41	−1.0	École Polytechnique, Lausanne
17	0.0	University College London	42	2.5	Washington, St Louis
18	1.0	Northwestern	43	1.5	California, Davis
19	−1.0	Cornell	44	−1.5	Karolinska Institute
20	0.0	Michigan	45	−6.0	London School E&P
21	0.0	Toronto	46	1.0	Minnesota
22	0.0	Carnegie Mellon	47	1.0	Ludwig-Maximilians, München
23	0.0	Duke	48	3.0	Brown
24	0.0	Washington	49	1.0	Pohang S&T
25	0.5	Georgia Institute Technology	50	−4.0	Peking

Table 8b. New ranking (51–100) of top 100 in THEWUR (TSL Education Ltd. 2012).

Rank	Gain	University	Rank	Gain	University
51	−2.0	Manchester	76	0.0	Pittsburgh
52	2.5	Boston	77	−6.5	Wageningen Research Center
53	0.0	Ohio State	78	0.0	Heidelberg
54	2.0	South California	79	1.0	Durham
55	4.5	École Normale Supérieure	80	1.0	Pierre et Marie Curie
56	−1.5	Kyoto	81	6.0	Tufts
57	4.0	Pennsylvania State	82	9.0	Colorado Boulder
58	−1.0	King's College London	83	−1.0	Lund
59	−7.0	Tsinghua	84	−1.0	Amsterdam
60	2.5	École Polytechnique	85	9.5	Notre Dame
61	3.0	Leiden	86	10.0	California, Irvine
62	−4.0	KU Leuven	87	−10.0	Delft
63	−3.5	Seoul National	88	4.0	Paris
64	−1.5	Sydney	89	0.5	Zürich
65	5.5	Georg-August, Göttingen	90	−2.0	McMaster
66	1.0	Utrecht	91	6.0	Maryland, College Park
67	−1.5	Queensland	92	2.5	Michigan State
68	−2.5	Hong Kong S&T	93	−9.0	Montreal
69	6.0	Rice	94	−4.5	Groningen
70	2.5	Massachusetts	95	−10.0	New South Wales
71	−3.0	Korea Adv. Institute S&T	96	2.0	Arizona
72	−3.0	Purdue	97	−4.0	Ghent
73	1.0	Bristol	98	−12.0	Nanyang Technological
74	5.0	Emory	99	0.5	Humboldt-Universität zu Berlin
75	−2.5	Erasmus Rotterdam	100	−0.5	Monash

Table 9. Changes in ranking in THEWUR (TSL Education Ltd. 2012).

Gain	Frequency
More	7
6–10	35
1–5	23
0	27
−1 to –5	7
−6 to –10	1
Less	0

from its original 96 to the new 86. The greatest negative gain of –12 goes to Nanyang Technological changing its position from the original 76 to the new 98.

As Table 9 shows, 27 universities retained their original rankings. On the whole, there are more positive gains than negative ones. Specifically, 23 universities moved up between one and five positions, 35 between six and 10 positions and seven more than 10 positions. On the other hand, seven dropped by 1 to five positions and one by six to 10 positions.

Discussion and conclusion

Discrepancies between the nominal indicator weights (intended by the ranker), and the attained indictor weights (what the overall score really is made up of) misinform the rank

consumers and leads to misjudgement and unrewarding, fruitless effort. Misjudgement may take the form of unjustified self-praise and unnecessary disappointment in the competitive context of today. Fruitless effort is simply directing resources to improve ranking but on the wrongly identified errant indicators. If reputation is important and resources need to be channelled optimally, the university cannot afford to ignore discrepancies such as those shown in the analyses.

Stephen Hawking acknowledged his error in believing that everything swallowed by a black hole must be lost forever, and Albert Einstein recognised that the universe is expanding contrary to his earlier belief (Connor, 2013). We (less significant earthlings) need to acknowledge our blunder in a matter as important as world university ranking, even if there is no reason other than that such exercises take up so much energy, time and resources which could be used more beneficially otherwise.

On a smaller scale but still vital are medical blunders made by doctors (Gastaldo, 2012) as cited below:

> When it comes to medicine, 'more is not always better,' writes neurosurgeon and CNN chief medical correspondent Sanjay Gupta in the *New York Times*. Doctors make thousands of mistakes each year—in 1999, as many as 98,000 Americans were dying annually due to medical errors, and that number could have more than doubled by now—and, in Gupta's opinion, this is partly because doctors are increasingly practicing 'defensive medicine.' Far more tests and procedures are performed in the US than anywhere else in the world, and these numbers—as well as the number of drugs prescribed—are rising. The problem is that many of these tests, procedures, and prescriptions are entirely unnecessary . . . Unnecessary tests can lead to false positives, which can lead to unnecessary procedures. Unnecessary procedures can go horribly wrong. Unnecessary medications can cause serious side effects . . . It's also important to talk about these mistakes: 'The only thing we can do is learn each time one happens, and reduce future errors in the process.'

Do these examples of admitting mistakes in cosmological science and medicine make sense to higher education? Yes, they certainly do. The university is an intellectual powerhouse where high-level fine thinking is supposed to take place, and this goes together with the wisdom to recognise honest errors and the courage to admit them when found. An error in university ranking and its consequential judgement, when compared with science and medicine, is definitely less damaging, but an error nevertheless.

Ranking world universities has been around for the past decade, and it is not known the extent to which it has helped or harmed higher education. This definitely is an area of research worthy of pursuit. Judgements and decisions would have been made and actions would have been taken (together with the resources thus utilised) on the face value of the ranking results released hitherto. In the past, before methodological error as that found in the present study had yet to come to light, blind faith was the rule of the game. Now that an error has been uncovered in the discrepancy between the nominal and attained weights, those rankings of the past surveys had meanings not the same as what they were thought or made out to appear. In those years, judgements, decisions and actions would have been based on doubtful information.

It is true that what has been done cannot be undone. And, this analysis is based on historical data (although fresh enough to be relevant) and the findings can be easily dismissed as only of historical interest. But, then, ranking will not stop here and will stay and continue (UNESCO, 2011), and the same error has a very high chance of being perpetuated if nothing is done to it. There is obviously a need for re-orientation among the rank consumers to become more discerning and critical – to become better-informed consumers

who always check the information they received and cannot be misguided, demanding the highest integrity of the data and information. This means that the consumers of ranking need to refrain from rushing to meet the stop-press with immediate press releases when a set of ranking results are published by the rankers; not to be overexcited by the ranking, but to take a careful look at the soundness of methodology as the details are more critical than the superficial ranking and its basis (the overall score).

It is not known how the universities will accept and react to the new ranking based on *T*-scores as presented in the relevant league tables here. Predictably, those having no change in ranking will be indifferent, those gain in position will be happy to see themselves moving up the scale, and those losing positions may be, naturally, sceptical. Whatever the response is, it is most important that all realised that the current methodology yields misleading results, the overall scores are not what they are promised to be, and these need to be rectified for common good and *for the future*. It is unfortunate that information such as those presented here comes late, but that is that nature of the matter as research needs data and takes time, and publication of research findings takes even more time.

As for the rankers, going from the raw scores to the *T*-scores (or any other suitable standardisation) which more accurately reflect the relative contributions of individual indicators is only a small technical refinement which costs very little time and effort – all it takes is a few more lines of instructions in the computer programmes for processing the data. It is the rankers' moral obligation to keep their promise by ensuring that the overall score truly reflects the indicator weights pronounced so that the rank consumers are properly informed, given that the indicator weighting schemes are valid.

The observed discrepancy may look as a small technical flaw but its impact is much greater than it has been hitherto realised. If ranking is to stay, it should not be in the present form but should ensure what is promised is what is delivered. Otherwise, it appears as if the blind is leading the blind.

Caveats

This study is basically concerned with the relative importance of indicators within each of the three main ranking systems: whether the indicator weights are actualised as intended by the rankers in the derivation of the overall scores for ranking the universities. Other methodological issues which are as important if not more, such as spurious precisions and mutual compensation of indicators, are not dealt with here as they have been covered elsewhere (e.g., Soh, 2012a, 2012b).

On methodology, this study compares the nominal weights (as indicated by the unstandardised regression weights or *b*-weights) and the attained weights (as indicated by the standardised regression weights or Beta-weights). Admittedly, this is but one of the several methods for comparing relative importance of indicators suggested by Nathans, Oswald, and Nimon (2012). However, using the beta-weights is the most straightforward method which, hopefully, will get the message across to all interesting in world university ranking, particularly the rankers and rank-users.

References

Connor, S. (2013, April 11). Stephen Hawking admits the biggest blunder of his scientific career – early belief that everything swallowed up by a black hole must be lost forever. *The Independent.* Retrieved from http://www.independent.co.uk/news/science/stephen-hawking-admits-the-biggest-blunder-of-his-scientific-career–early-belief-that-everything-swallowed-up-by-a-black-hole-must-be-lost-forever-8568418.html

Gastaldo, E. (2012, August 1). Doctors: To stop erors, stop over-treating: Sanjay Gupta reflects on the increasing number of medical mistakes. *Newser*. Retrieved from http://www.newser.com/story/151260/doctors-to-stop-errors-stop-over-treating.html

Nathans, L.L., Oswald, F.L., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research, & Evaluation, 17,* 1–19.

Quacquarelli Symonds World University Ranking. (2012). *QS World University Rankings – 2012*. Retrieved from http://www.topuniversities.com/university-rankings/world-university-rankings/2012

Shanghai Ranking Consultancy. (2012). *Academic Ranking of World Universities 2011*. Retrieved from http://www.shanghairanking.com/ARWU2011.html

Soh, K. (2013a). Misleading university rankings: Cause and cure for discrepancies between nominal and attained weights. *Journal of Higher Education Policy and Management, 35*, 206–214.

Soh, K. (2013b). Nominal versus attained weights in Universitas 21 Ranking. *Studies in Higher Education, 1–8,* iFirst article.

Soh, K.C. (2012a). Profiling universities, not only ranking them: Maximizing the information of predictors. *Higher Education Review, 44*, 27–42.

Soh, K.C. (2012b). Simpson's paradox and confounding factors in university rankings: A demonstration using QS 2011–12 data. *European Journal of Higher Education, 2*, 389–402.

Times Higher Education. (2012). *THE 100 under 50. Methodology*. Retrieved from http://www.timeshighereducation.co.uk/stiry.asp?stirycode=420090

TSL Education Ltd. (2012). *World University Rankings 2012–2013*. Retrieved from http://www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking

UNESCO. (2011). *University rankings are here to stay, reports University World News*. Retrieved from http://www.unesco.org/new/en/education/themes/strengthening-education-systems/higher-education/single-view/news/university_rankings_are_here_to_stay_reports_university_world_news/

Williams, R., Rassenfosse, G., Jensen, P.H., & Marginson, S. (2012). *U21 Ranking of National Higher Educations 2012*. Birmingham: Universitas 21.