Problems of indicator weights and multicolinearity in world university rankings: comparisons of three systems

Kaycheng Soh

World university rankings (WUR) use the weight-and-sum approach to arrive at an overall measure which is then used to rank the participating universities of the world. Although the weight-and-sum procedure seems straightforward and accords with common sense, it has hidden methodological or statistical problems which render the meaning of the overall measure (and therefore the ranking results) suspect. The study investigates indicator weight inconsistency and multicolinearity. The 2013 ranking results of the ARWU, QSWUR, and THEWUR for the top 100 universities were re-analysed. Correlations, weight discrepancies, and tolerances were estimated and then compared across the three ranking systems. Weight discrepancy and multicolinearity were found in all three systems to different degrees. These can be attributed to the nature of the indicators and their inter-correlations. Each system has strengths and weaknesses. Rank users need to be aware of the interpretive problems, to avoid equating the rankings from the systems, and to make specific reference to the system concerned when citing rankings. Re-conceptualisation and possible modifications to the systems are discussed.

Keywords: Indicators, multicolinearity, multiple regression, world university ranking.

World university ranking (WUR) has become an annual affair ever since its first appearance in 2003 with the release of the Academic Rankings of World Universities (ARWU) (Shanghai Ranking Consultancy, 2013) conducted at the Shanghai Jiaotong University (Liu, Cheng, and Liu, 2005). This was soon followed by the QS World University Rankings (QSWUR; Quacquarelli Symonds Limited, 1994-2014) and the Times

Higher Education World University Rankings (THEWUR; TSL Education Ltd, 2012). Universities the world over take to the ranking results like a fish to water, almost always without any apparent reservations. Even when the systems have been criticised on conceptual grounds (in the early years, eg Lincoln, 2012) and for statistical problems (in recent years, eg Soh, 2013a), the rank producers and consumers have continued to conduct their respective businesses in the same way, failing to take into account the issues and problems. WUR has perhaps become so engrained in higher education that it is difficult to dismantle the systems or even to modify them for greater trustworthiness. As Hazelkorn (2009) pointed out after a comprehensive review of WUR, rankings are here to stay.

Even though universities might need the information provided by the various systems for whatever reasons, and the rank producers need to continue their business for commercial purposes, consumers need to be discerning and cautious when using the information to avoid being misinformed and thereby misled, and producers should consider

TABLE 1

Indicators and assigned weights

Indicator	Weight per cent	Code
ARWU		
Alumni winning Nobel Prizes and Fields medals	20	Alumni
Staff winning Nobel Prizes and Fields medals	20	Award
Highly cited researchers	10	HiCi
Articles published in *Nature* and *Science*	20	N&S
Science Citation Index and Social Sciences Citation Index	20	PUB
Per capita academic performance on the above	10	PCP
QSWUR		
Academic peer review	40	Academic
Employer/recruiter review	20	Employer
Citation per faculty member	20	FacStd
Faculty-student ratio	10	IntFac
International students	5	IntStd
International faculty	5	Citation
THEWUR		
Teaching	30	Teaching
Research	30	Research
Citations	30	Citations
International mix	7.5	Internat
Industry income	2.5	Industry

modifications so as to serve their clients better with more trustworthy information.

Most WUR systems use the common approach of weight-and-sum. They choose a set of indicators which are believed to be necessary to index the university's academic excellence. The indicators are assigned different weights to reflect the system's perception of their relative importance. Weighted indicator scores are then summed to form a total overall score for each university. Universities are then ranked based on their overall scores. Finally, the rankings form a league table used by universities to evaluate their relative positions on a scale of academic excellence. Table 1 shows the indicators and their assigned weights for the three most popular WUR systems of the day.

Problems

World University Rankings seem like common sense, enough for them to be widely accepted and emulated by various social and educational ranking systems. Nonetheless, when looked at more carefully, ranking systems are not as simple as they appear (Soh, 2013a) and there are problems such as spurious precision, mutual indicator consumption, and scaling. Of special interest to the present study are the two related problems of indicator weights and multicolinearity.

Indicator weights

WUR systems assign different weights to their chosen indicators. In doing so they tacitly assume that the total or overall score is contributed by the indicator scores in the assigned weights, that is, correctly in the stipulated proportions. Recent studies demonstrate that this is not always true and it may in fact be always untrue (Soh, 2013b, 2013c). When the overall score is *not* made up of the indicator scores in the proportions assigned by the rank producer and, at the same time, the rank users are not aware of the discrepancies between the assigned and obtained indicator weights, the users are not given what they have been promised by the producer. Taking the unrealised indicator weights as if they have been actualised in the overall score misleads the users. An indicator is over-weighted when its actualised or obtained weight is greater than the assigned or nominal weight, and thereby becomes more powerful or determining than it should be, without this exaggeration being known to the unsuspecting user. A decision based on it is therefore misguided because of misinformation and this may be costly in terms of resources misallocated. To date there are no data available to show the extent and costs of such misinformation: this obviously is a problem worthy of further research.

Multicolinearity

A problem closely related to the weight-discrepancy problem is multicolinearity. Since the indicator scores are based on the data collected for a university as an entity, the data may be correlated highly, moderately, lowly, or not at all, depending on their nature. If two indicators are very similar in nature, even if data are collected for them from different sources, their indicator scores are likely to be correlated. If the correlation is high, the information provided by one will be much the same as that provided by the other indicator. This means one of the two indicators is redundant: it does not provide new information in addition to what is already known from the other. The totality of such correlations among indicators of a system is shown up as multicolinearity. Ideally the predictors should correlate highly with the criterion (eg the overall score) and, at the same time, have low or no correlations among them. The problem of multicolinearity in WUR has earlier been found for the THEWUR results (Soh, 2014).

In multiple regression, multicolinearity occurs when predictors are highly correlated, rendering some of them redundant, contributing information which is already covered by other indicators. This makes the understanding of the criterion murky and gives the redundant predictors undue importance. The existence of multicolinearity is evaluated by *tolerance* which statistically is operationalised as $(1 – R^2)$. In this formulation, R^2 is the multiple coefficient of determination, which is the variance of a particular predictor predicted by all remaining predictors in a multiple regression model. Understandably, when the variance of that particular predictor is accounted for by the other predictors, $(1 – R^2)$ will be rather small, indicating that whatever that predictor is able to predict is already covered by the others, thus, that predictor is redundant. There are various recommendations for acceptable levels of tolerance. Tabachnick and Fidell (2001) recommend a value of .10 as the minimum level of tolerance, but this is not a hard and fast rule. Menard (1995) recommends a value of .20 and Huber and Stephens (1993) a value of .25. It appears that researchers use different criteria that suit their research purposes. Perhaps, the relative magnitudes of tolerance would be more meaningful for evaluating predictor redundancy than any fixed figure. This is the approach adopted for the present study.

In sum, the present study is an attempt to evaluate the severity of the problems of indicator weight discrepancy and multicolinearity. However, a preliminary analysis of the inter-indicator correlations is useful for a preview of the probable causes of the problem. Therefore, the study will look for answers to the following three research questions:

How do the three systems compare in correlations among indicators, between indicators and overall?

How do the three systems compare in actualising the indicator weights?

How do the three systems compare in multicolinearity?

Method

Data

The twin problems of indicator weights and multicolinearity have been surfaced only in recent years (Soh, 2013b, 2013c). It is logical to continue this line of research in replications using fresh data and to go beyond the individual systems to make comparisons between them, to evaluate their respective efficacies with reference to the two problems. For this purpose, data from the ARWU, QSWUR, and THEWUR for the year of 2013 were used.

The data were gleaned from the three systems' websites. Only the top 100 (or 101 where there were ties) universities were selected as this number was considered to be large enough to provide data needed for the kind of analyses to be performed.

Analysis

With reference to the research questions, Pearson's product moment correlation coefficients were calculated to find out how indicators correlate with one another and also with the Overall score. This is relevant to the first research question.

Then, stepwise forward multiple regression analysis was run for each system using the Overall as the criterion and the indicators as the predictors. For this, the resultant standardised regression coefficients (beta-weights) were used to calculate the indicators' variance components as percentages of criterion variance predicted by each predictor. Comparisons were then made between the indicator's assigned weights and variance component to detect its magnitude of weight discrepancy. This is relevant to the second research question.

The multiple regression analysis also yielded tolerances for each indicators within a system. These provide information for evaluating multicolinearity and are relevant to the third research question.

Results

How do the three systems compare in correlations among indicators, between indicators and overall?

Before an attempt is made to answer the primary research questions, it is useful to look at correlations among the indicators and how indicators correlate with the Overall. For this, Pearson's correlation coefficients were calculated and shown in Table 2.

As Table 2 shows, the ARWU indicators correlate highly with Overall. The coefficients vary from r = 0.39 (PUB) to r = 0.93 (N&S), with a median of r = 0.82. Among the six indicators, the coefficients vary from r = -0.15 (N&S and PCP; the only negative coefficient) to r = 0.86 (N&S and HiCi) with a median of r = 0.62.

TABLE 2

Correlations among indicators and with overall

ARWU							
	Overall	Alumni	Award	HiCi	N&S	PUB	PCP
Overall	1.00	.80	.84	.88	.93	.62	.73
Alumni		1.00	.78	.60	.64	.39	.67
Award			1.00	.67	.69	.32	.70
HiCi				1.00	.86	.64	.52
N&S					1.00	.61	.63
PUB						1.00	(.15)
PCP							1.00

QSWUR							
	Overall	Academic	Employer	Fac Std	Int Fac	Int Std	Citation
Overall	1.00	.65	.52	.52	.24	.28	.38
Academic		1.00	.44	.23	(.05)	(.12)	(.16)
Employer			1.00	(.15)	.36	.30	(-.07)
FacStd				1.00	(-.05)	(-.03)	(.03)
IntFac					1.00	.61	(-.07)
IntStd						1.00	(-.07)
Citation							1.00

THEWUR						
	Overall	Teaching	Research	Citations	Internat	Industry
Overall	1.00	.93	.92	.49	(.16)	(.14)
Teaching		1.00	.88	.31	(.00)	(.14)
Research			1.00	(.19)	(.12)	.23
Citations				1.00	(-.09)	-.28
Internat					1.00	(.14)
Industry						1.00

Note: all coefficients are statistically significant (p<.05, two-tailed) except those in parentheses.

For QSWUR, the indicators correlate with Overall with a median of r = 0.45, and the coefficients vary from r = 0.24 (IntFac) to r = 0.65 (Academic). Among the six indicators, the coefficients are generally low, with nine being non-significant and even negative. The median for these is r = -0.04.

For THEWUR, the indicator-Overall correlations vary from r = 0.14 (Industry) to r = 0.93 (Teaching), with a median of r = 0.49. Among the five indicators, the coefficients are generally low and some are negative, except the r = 0.88 between Teaching and Research. The inter-indicator coefficients vary from r = -0.28 (Citations and Industry) to r = 0.88 (Teaching and Research), with a median of r = 0.12.

With the results above, considering the medians of correlations for the Overall and among indicators, it is concluded that ARWU has a set of more homogeneous indicators when compared with the other two systems. The medians for the Overall indicate that QSWUR and THEWUR have similar degrees of indicator homogeneity, but

TABLE 3

Comparisons on indicator weights, variance components, and tolerance

	Assigned weight per cent	Beta-weight	Variance component per cent	Weight discrepancy	Tolerance
ARWU (R = 0.99, R^2 = 0.977)					
N&S	20	.367	31.2	11.2	.195
Award	20	.220	18.7	-1.3	.287
Alumni	10	.158	13.4	3.4	.336
HiCi	20	.175	14.8	-5.2	.222
PUB	20	.129	11.0	-9.0	.467
PCP	10	.127	10.8	0.8	.381
QSWUR (R = 0.88, R^2 = 0.765)					
Academic	40	.361	23.1	-16.9	.747
Fac-Std	20	.397	25.4	5.4	.938
Citation	20	.346	22.1	2.1	.947
Employer	10	.268	17.1	7.1	.731
IntStd	5	.192	12.3	7.3	.902
IntFac	5	-	-	-	-
THEWUR (R= 0.99, R^2= 0.998)					
Teaching	30	.426	31.2	1.2	.196
Citation	30	.299	21.9	-8.1	.788
Research	30	.463	33.9	3.9	.200
Internat	7.5	.127	9.3	1.8	.932
Industry	2.5	.049	3.6	1.1	.834

Note: Variance components per cent are based on the proportions of the predicted criterion variance (adjusted coefficient of multiple determination).

THEWUR has rather greater indicator homogeneity than QSWUR. Such patterns of correlations have implication for multicolinearity to be dealt with later.

Before looking into multicolinearity, a study was made to find out the extent to which the systems' indicator weights were actualised in their data. Table 3 shows the comparisons on this aspect.

How do the three systems compare in actualising the indicator weights? Stepwise multiple regression was run with the Overall as the criterion and the indicators as the predictors. The beta-weights are the standardised regression coefficients showing the relative contribution of each indicator to the Overall, after the indicator distributions were transformed in to z-score distributions. The order of the indicators indicates the precedence for entering into the multiple regression, decided by the computer usually selecting the one that predicted the criterion variance most first.

Then, variance components were calculated using the beta-weights and the adjusted R^2, with each figure showing how much a particular indicator predicts the *explained variance* of the Overall. With the two sets of weights, the weight discrepancy was calculated to indicate the extent to which an indicator's contribution to the Overall differs from the expected contribution (as indicated by its assigned weight). The order of indicator entry may not coincide with the order of listing by the ranking system. For instance, ARWU list Alumni, Award, and N&S in that order with the tacit importance or preference of the indicators, but the results show that, in terms of proportions of variance explained, the order for the three indicators are N&S, Award, and then Alumni.

For ARWU, the multiple R is 0.99, with an Adjusted R^2 of 0.977 indicating the six indicators together predicted 97.7 per cent of the criterion variance of the Overall. Of the indicators, N&S was given a weight of 20 per cent by the systems, but it turned out to contribute 31.2 per cent to the Overall, with an over-weight of 11.2 per cent. Similarly, Alumni over-weighted by 3.4 per cent and PCP by 0.8 per cent. On the other hand, PUB was under-weighted by -9.0 per cent, HiCi by -5.2 per cent, and Award by -1.3 per cent. Such discrepancies signal that care should be exercised when using the Overall as it is not actually reflecting the stipulated relative importance of the indicators.

For QSWUR, the multiple R is 0.88 with an Adjusted R^2 indicating that five indicators predicted 76.5 per cent of the criterion variance, with the exception of IntFac. Academic was under-weighted by –16.9 per cent, a discrepancy so large that it should be taken seriously. At the same time, FacStd is over-weighted by 5.4 per cent. The other indicators over-weighted from a small 2.1 per cent (Citation) to a sizable 7.3 per cent (IntStd). Note that IntFac was rejected by the computer to enter the regression process as it did not contribute to the Overall.

For THEWUR, the multiple R is 0.99 with an Adjusted R^2 of 0.998 indicating that 99.8 per cent of the criterion variance was predicted by the five indicators. Citation is under-weighted with a discrepancy of -8.1 per cent, while the other four indicators all over-weighted, with Research over-weighted by 3.9 per cent, followed by 1.8per cent for International Mix, and then around 1.0 per cent by Teaching and Industry Income.

With the findings above, one conclusion is that the predictors of ARWU and THEWUR were able to predict almost the total variance of the Overall within each system. This compared more favourably for these two systems than for QSWUR for which only about three-quarter of the criterion variance was predicted. More specifically, the total of indicator weight discrepancy is - 0.1 for ARWU and THEWUR, but 5.0 for QSWUR. Another conclusion is that all three systems show obvious discrepancies between the assigned and attained indicator weights, some more serious than the others.

How do the three systems compare in multicolinearity?

As noted earlier, the lowest recommended maximum tolerance is 0.10 and the most lenient is 0.25. The stringent criterion of 0.10 means that when 90per cent of a predictor's variance is predicted by all other indicators in the regression model, it has not more than 10 per cent variance left to explain and is therefore considered redundant. For the most lenient tolerance of 0.25, all other predictors together predict 75 per cent of the variance, with 25 per cent left to be explained. However, it was argued above that since there is no fixed cut-off point, the relative magnitudes of predictors may be more useful for evaluating multicolinearity, and this approach is adopted here.

As Table 3 shows, for ARWU, the tolerances vary from 0.195 (N&S) to 0.467 (PUB), with a median of 0.312. This could be due to the sizable correlations among the six indicators as shown in Table 1. Here, N&S and HiCi have the two lowest tolerances and both are below 0.25. This suggests that the two indicators are more likely redundant as compared with the other four indicators.

For QSWUR, the tolerances vary from 0.731 (Employer) to 0.947 (Citation), with a median of 0.902. This could be expected from the rather low inter-predictor correlations as reported in Table 2. The two indicators with the lowest tolerances are Employer (0.731) and Academic (0.747) in comparison with the other indicators (not counting IntFac which was not entered to the model). However, as these two tolerances are sizable, it can be argued that they can be retained in the model.

As for THEWUR, the tolerances vary from as low as 0.196 (Teaching) to as high as 0.932 (Internat), with a median of 0.788. This is also attributable to the relatively low inter-indicator correlations (Table 2). The two lowest tolerances go to Teaching (0.196) and Research (0.200).

In sum, when multicolinearity is seen as a modelling problem, it is most severe in ARWU with its average of 0.312, but QSWUR and THEWUR seem to be fine with their respective medians of 0.853 and 0.590.

Discussion and Conclusions

The findings of the analyses are summarized in Table 4, with the three systems ranked on each aspects studied. In terms of homogeneity of indicators as reflected in the inter-indicator correlations and the correlations with the Overall, ARWU has the best choice of indicators, followed by THEWUR and then QSWUR. In terms of indicator weight discrepancy, ARWU and THEWUR are on a par with one another and they have a lower discrepancy than QSWUR. In terms of multicolinearity, QSWUR has the least problem, followed by THEWUR, and then ARWU. Aggregating the ranks for these three aspects it is clear that ARWU and THEWUR rank higher than QSWUR; this indicates that QSWUR has more methodological (statistical) issues to handle than the other two systems do.

TABLE 4

Summary of findings: rank order of systems

	Inter-indicator correlation	Weight discrepancy	Tolerance	Sum of ranks
ARWU	1 (.62)	1= (-.01)	3 (.312)	5.5
QSWUR	3 (-.04)	3 (5.0)	1 (.902)	7
THEWUR	2 (.12)	1= (-.01)	2 (.788)	5.5

Note: Figures in parentheses are averages.

However, as is often true, summary statistics such as the average and rank may mask important variations, since different systems have their respective strengths and weaknesses relative to others. This implies that the strengths and weaknesses of a particular system should not be neglected when interpreting and using their ranking results, lest misinterpretation occurs, which may have undesirable consequences in real terms.

The quality and trustworthiness of a WUR system depend, *inter alia*, on its choice of indicators, their nature and inter-correlations, as these factors determine, first, the discrepancy between assigned weights and the actual weights obtained from the data, and whether some indicators are redundant because they overlap with others.

ARWU has all its indicators based on academic achievement. This, by its nature, results in high inter-indicator correlations, which may lead to higher overall reliability or internal consistency and hence invite greater faith in the ranking results. However, as a consequence, this leads to greater overlaps among the indicators and thus causes higher multicolinearity (ie lower tolerance) which is a disadvantage in constructing the Overall used for ranking. In other words, the findings suggest that ARWU could be made more parsimonious by dropping some of the overlapping indicators and still could be trusted to rank the universities on academic excellence, if this is strictly defined as dealing only with academic achievement. For example, the extremely high correlation r = 0.93 for N&S with the Overall suggests that this indicator by itself might be sufficient as the rankings of the universities on N&S and on the Overall will not be much different. Of course, there may be objection that this strong definition of academic excellence is too narrow and the reduction of indicators is too drastic. In fact, ARWU has been criticised for a bias towards research in natural sciences which naturally result in publication (and citations) of papers in the two most prestigious platforms for scientists, the journals *Nature* and *Science*.

QSWUR aims to broaden the definition of excellence by including administrative as well as academic measures (ie employer survey, faculty-student ratio, proportion of international faculty members and students). Since its inter-indicator correlations are low, the tolerance is naturally high (showing low multicolinearity). This should turn out to be an advantage to the system, but the system suffers from having a high weight discrepancy. That shows that there is gross inconsistency between the assigned and obtained weights, with gross under-weighting of Academic Reputation (Academic) and over-weighting in four other indicators, with one indicator International Faculty (IntFac) clearly redundant. There is obviously a need to re-conceptualise the weighting scheme if all the indicators are to be retained. Note that 60 per cent of the weight is assigned to two indicators of an academic nature (ie Academic Reputation and Citation) while the remaining 40 per cent is for the nonacademic indicators. In view of this, it seems that it is more appropriate to use the Overall as an indicator of *institutional excellence* (to reflect the importance placed on administration in higher education institutions) rather than academic excellence (which can be seen as a narrower and more intellectual dimension in the higher education context).

THEWUR has a pattern of low inter-indicator correlations which is advantageous where multicolinearity is concerned. Its average tolerance suggests that each of its five indicators predicted the Overall to some extent without too much overlap. The very high correlations of Teaching and Research with the Overall may look like an advantage, but since these two indicators are very highly correlated, one of them could well be excluded. International Mix and Industry Income have high tolerances which suggest that they leave much variance to be predicted by the other indicators. However, their lack of sizeable correlations with the Overall render their functions in the regression model suspect. Obviously, there is also a need to reconceptualise the model and to re-assign indicator weights for the systems to function with greater efficacy.

Since different WUR systems have different indicators differently weighted, they are bound to yield different rankings even for the same university. The extent to which the three systems rank the universities consistently does not seem to have been systematically studied. An attempt is made here as a follow-up.

When the three lists of top 100 universities were compared, surprisingly only 40 appeared in all three lists. The correlations for the Overall of these universities were then calculated, resulting in $r = 0.68$ between ARWU and QSWUR, $r = 0.86$ between QSWUR and THEWUR, and $r = 0.83$ between THEWUR and ARWU. Although all the coefficients are statistically significant ($p<.05$, two-tailed), they respectively indicated only 46 per cent, 74 per cent, and 69 per cent shared variances.

Obviously, consistency among the three sets of ranking is moderate at best. The inconsistency of ranking happens even for the top universities, as clearly illustrated in Table 5. Specifically, Harvard ranks first in ARWU, second in QSWUR and shares the second position with Oxford in THEWUR. Stanford ranks second in ARWU, does not appear in the top five in QSWUR, but reappears as the fourth in THEWUR. Cambridge ranks fifth in ARWU, third in QSWUR but not within the top five in THEWUR. Four of the top five are American universities in ARWU, three in QSWUR, and four in THEWUR. Of course, if all universities have the same respective ranks in the three systems, there needs be only one such system. Having three systems as currently practised leads to inconsistency in ranking, but the inconsistency problem can be solved by a statistical procedure (ie T-scaling) which is easy to implement (Soh, 2013c).

TABLE 5

Top five universities in the three rankings

<table>
<tr><th></th><th>ARWU</th><th>QSWUR</th><th>THEWUR</th></tr>
<tr><td>1</td><td>Harvard</td><td>MIT</td><td>California Institute of Technology</td></tr>
<tr><td>2</td><td>Stanford</td><td>Harvard</td><td rowspan="2">Harvard
Oxford</td></tr>
<tr><td>3</td><td>California, Berkeley</td><td>Cambridge</td></tr>
<tr><td>4</td><td>MIT</td><td>University College London</td><td>Stanford</td></tr>
<tr><td>5</td><td>Cambridge</td><td>Oxford</td><td>MIT</td></tr>
</table>

In view of the inconsistency, rankings assigned by different systems cannot be taken as equivalent, that is, the ranks of a university cannot be assumed to be the same across systems. At present each WUR system functions like a currency: a rank is not a universal quantity, just as a US dollar does not have the same purchasing power as a Singapore dollar or any other currency. This is obvious but practically neglected most of the time when a university quotes a rank of, say, 20 without specifying the ranking system. This implies that it needs to become a common practice that when a university quotes its ranking, it also quotes at the same time the specific system and the ranking year.

Moreover, since a rank does not have the same meaning across systems due to the differences in indicators and their weights, ranking should not be the focus of attention although it can serve as an index to locate a university on the league table. A university wishing to compare itself with others should look into its profile provided by the indicator scores which give more and specific information for strategic planning or some other academic and administrative purposes. And, it is worth repeating that a difference may need to be drawn explicitly between *academic excellence* (a narrower and specific quality) and *institutional excellence* (a broader and general quality).

Since different WURs thus provide different information, rank users (not only the universities) are well-advised to take the necessary precaution against the limitations of rankings. It has been suggested that universities need to look at their profiles of indicator scores rather than the rankings (Soh, 2012). Ranking is not ‘one currency’ but many currencies; rankings should be treated as such, to avoid misinformation, misinterpretation, and being misguided.

Address for correspondence
Kaycheng Soh, Independent Education Consultant, Singapore (Retired from National Institute of Education, Nanyang Technological University, Singapore). e-mail: sohkaycheng@hotmail.com. Mailing address: 50 Lorong 40 Geylang #07-29, The Sunny Spring, Singapore 398074

References

Hazelkorn, E (2009) *The impact of global rankings on higher education research and the production of knowledge*, Dublin: Centre for Social and Educational Research, Dublin Institute of Technology

Huber, E and JD Stephens (1993) 'Political parties and public pensions: a quantitative analysis', *Acta Sociologica* 36 309-325

Lincoln, D (2012) 'Rankings: an idea whose time has come, and gone', *InsideHigher Ed* http://www.insidehighered.com/blogs/worlf-virw/rankings-idea-whose-time-has-come-and-gone

Liu, NC, Cheng, Y and L Liu (2005) 'Academic ranking of world universities using scientometrics – a comment to the "Fatal Attraction"', *Scientometrics* 64(1) 101-109

Menard, S (1995). *Applied logistic regression analysis*, Sage University Series on Quantitative Applications in the Social Sciences, Thousand Oaks, CA: Sage

Quacquarelli Symonds Limited (1994-2014) *QS World University Rankings 2013*, http://www.topuniversities.com/university-rankings/world-university rankings/2013#sorting =rank+region=+country=+faculty=+stars=false+search=

Shanghai Ranking Consultancy (2013) *Academic Ranking of World Universities 2013* http://www.shanghairanking.com/ARWU2013.html

Soh, Kaycheng (2014) 'Multicolinearity and indicator redundancy problems in world university rankings: an example using THEWUR 2013-2014 data', *Higher Education Quarterly*, in press

Soh, Kaycheng (2013a) *Social and educational rankings: problems and prospects*, New York: Untested Ideas Research Centre

Soh, Kaycheng (May 2013b) 'Misleading university rankings: cause and cure for discrepancies between nominal and attained weights', *Journal of Higher Education Policy and Management,* 35(2) 206-214

Soh, Kaycheng (October 2013c) 'Rectifying an honest error in world university rankings: a solution to the problem of indicator weight discrepancies', *Journal of Higher Education Policy and Management* http://dx.doi.org/10.1080/1360080X.2013.844670

Soh, Kaycheng (2012) 'Profiling universities, not only ranking them: maximizing the information of predictors', *Higher Education Review,* 44(2) 27-42

Tabachnick, BG and LS Fidell (2001) *Using Multivariate Statistics* 4th edn, Boston, MA: Allyn and Bacon

TSL Education Ltd (2012) *THE World University Rankings 2013-2014*
http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/world-ranking/institution/boston-university