Multicolinearity and Indicator Redundancy Problem in World University Rankings: An Example Using THEWUR 2013-2014 Data

Abstract

World university ranking systems used the weight-and-sum approach to combined indicator scores into Overall scores on which the universities are then ranked. This approach assumes that the indicators all independently contribute to the Overall score in the specified proportions. In reality, this assumption is doubtful as the indicators tend to correlate with one another and some highly so. This causes the multicolinearity problem rending some predictors redundant. At the same time, some indicators may contribute so little to the Overall score and thus making them non-contributing. When overlapping and non-contributing indicators are retained, the Overall score takes on a meaning very different from what it is originally intended to be.

Using data for the top 101 universities of the Times Higher Education World University Ranking 2013-2014, these problems are demonstrated and a solution is suggested. This resulted in a new Overall scores made up of only two of the six indicators, namely Research and Citation. The universities were then ranked on the basis of the new Overall score and compared with the original. It was also noticed that some of the universities have had their original Overall scores inflated by the non-contributing indicators and hence were over-ranked. Implications for using the new Overall and ranking are discussed.

Keywords: Academic excellence; Correlation; Multicolinrearity; Multiple regression

Generally, world university ranking (WUR) systems used the weight-and-sum (WAS) approach. They first selected indicators which are believed to index the various aspects of academic excellence and then assigned weights to each indicator, often *a priori* without explicit rationale for their different weightings, presumably to reflect their relative importance. Overall scores are then calculated for the participating universities which are thereby ranked. The rankings are then taken seriously (often, too seriously) by universities and by relevant people including sponsoring governmental and private agencies, potential students and their parents and, of course, the general public.

The WAS approach of arriving at the Overall scores and the resultant rankings seem to be objective and scientific, and is often touted to be so. It is this mistaken notion of being objective and scientific that has garnished WUR with a misplaced trust. There have been criticisms on conceptual and utilization grounds (e.g., Hazelkorn , 2007; Lincoln, 2012; Mumby, 2012; Raughvagers, 2011; Staff Writers; 2012). Statistical evidence of shortcomings of WUR has only been recently surfaced (Soh, 2013a, 2013b). As rightly pointed out by one of the anonymous reviewers, "*There have been some general criticisms on similar issues but a sound methodological issue of the problem of multicolinearity and indicator redundancy was missing so far.*"

An excellent example is the study by d'Hombres and Saisana (2013) of the EU-funded Centre for Research on Lifelong Learning. They re-analysed the 2007 data of the Shanghai Jiao Tong University Rankings and the Times Higher Education World University Rankings (two of the currently most popular WUR systems) and concluded, thus,

> The main findings of the report are the following. Both rankings are only robust in the identification of the top 15 performers on either side of the Atlantic, but unreliable on the exact ordering of all other institutes. And, even when combining all twelve indicators in a single framework, the space of the inference is too wide for about 50 universities of the 88 universities we studied and thus no meaningful rank can be estimated for those universities (last paragraph).

The use of statistical concepts and techniques to re-analyse ranking data is able to detect methodological flaws hidden in the data and hence oblivious to the untrained, naked eyes. Problems surfaced recently include, *inter alia*, spurious precision, inconsistency between assigned and attained weights, and assumed mutual compensation among indicators (Soh, 2013a). Multicolinearity is yet another problem of WUR which has not been explicitly studied. This problem occurs because the WAS approach tacitly assumes that the several indicators each is a measure of an aspect of academic excellence which is not measured by the other indicators in the same system. Statistically speaking, this assumption requires that the indicators are independent of one another but all correlate substantially with the criterion, the Overall. In Streiner's (2003) conception, a difference needs be made between a set of indicators forming a *scale* and another set of indicators forming an *index*. Such a subtle difference does not seem to be made by the WUR agents.

In view of the nature of the indicators which index different aspects of the university's academic achievement and administrative arrangements, the assumption of indicator independence is doubtful. When such independence is not obtained, there are considerable overlapping among indicators which render some of them unneeded and therefore redundant. More serious than this is that the Overall, based on which the universities are ranked, is not what the ranking agent has promised and specified but takes on a different meaning. For instance, if quality of teaching and quality of research are highly correlated, as indicators within a WUR system, they are doubly counted in the Overall and may not carry the weights stipulated by the system. This gives rise to multicolinearity and the solution is to identify and exclude the redundant indicators as they are not only non-contributing but also misinforming.

With the above as a backdrop, the research questions the present study tries to answer, by using THEWUR 2013-2014 data, are as follows:

1. How do the indicators and Overall correlate?
2. Is there a multicolinearity problem?
3. What are the indicators to retain?
4. What is the impact of the reduced model?

Method

Data THEWUR uses 13 data sources grouped into five broader indicators with different weightings (TLS Education Ltd., 2012b), thus:

- Teaching (30%): The dominant data is a by-invitation-only survey of academic reputation. This index also includes objective data such as staff-to-student ratio as a proxy measure for teaching quality and ratio of doctoral to bachelor's degree awarded. The survey is said to include a total of about 16,600 responses. With such a large sample, the ranking agent claimed that the results can be trusted.
- Research (30%): This is based on an academic reputation survey involving 10,000-plus responses. This index also includes objective data such as university research income, research productivity, and number of papers published in academic journals. Like Teaching, the simple size is used to claim trustworthiness of the results,

- Citation (30%): This is labelled by the ranking agent as the flagship indicator in view of the role the university plays in generating new knowledge that is valued also by other researchers who cited the published and indexed papers. The data pertains to the past five years and a total of 12,000 academic journals were examined.
- Industry income (2.5%): This is measured in terms of a university's ability to help industry with innovations, inventions, and consultancy. It seeks to capture knowledge transfer from the university to the business.
- International outlook (7.5%): This indicator indexes the diversity on campus in terms of students and staff as well as international academic collaboration, thus reflecting the international outlook of the university in the modern-day concept of higher education.

The first three indicators are relevant to the university directly and these are given a total weight of 90% thus reflecting the academic emphasis of THEWUR. Moreover, research is traditionally the main although not the only function of the university and it leads to publication which in its turn leads to indexing and citation. However, there is no explanation for how the different weightings were decided.

The data used for this study was gleaned from the relevant THEWUR 2013-2014 ranking website (TLS Education Ltd., 2012a). Only the Overall and the six indictor scores for the top 101 universities were used, with the last two universities having a tied ranking of 100. This number of universities was considered as more than adequate for the analysis to be performed in the present study.

Analysis

Statistical techniques of various types were used to suit the needed information relevant to the research questions. To begin with, correlation analysis was employed to ascertain the relationships among indictors and between these and the Overall. The results of the correlation analysis served as a preliminary investigation of the mutual influences of the indicators and Overall. Then, to verify the existence of multicolinearity, multiple regression analyses were run and the resultant Tolerances were used to judge the severity of the problem of indicator redundancy. Once the indicators were selected for retention in the model, stepwise multiple regression analyses were run for the purpose of finding out the predictive power of the identified indicators. Finally, using the indicators selected on the basis of multiple regression analysis, a new Overall was calculated and the universities were then ranked. Comparison were made between the original and new ranking for a better understanding.

Results

How do the indicators and Overall correlate?

Since the ranking involves using indictor scores to calculate the Overall score used for ranking, the first question comes readily to mind is the correlations among the indicators and these with the Overall.

As shown in Table 1, there are very high correlations among Overall, Teaching, and Research, indicating that universities rated high on teaching also were rated high on research and such universities also obtained high rating for Overall. Surprisingly, Citation has low and moderate correlation with Teaching and Overall, respectively, but a non-significant correlation with Research. The lack of a higher correlation between Citation and Research indicates that universities active in research may and may not have produced research papers that are valued by other researchers, and this suggests that quantity of research and quality of research are two different outputs which therefore cannot be equated.

Table 1 about here

It is of note that Internationalism has no significant correlations with any of the other four indicators and the Overall. This suggests that where Overall assessment of academic excellence is concerned, diversity among staff and students on campus is irrelevant.

Interestingly, Industry Income has a low positive correlation with Research but a low *negative* correlation with Citation. The positive correlation suggests that research has a mutual synergic benefit to the university and the business. However, the negative correlation suggests a conflict of interest between attaining academic excellence and serving the business; in this situation, staff who aim to publish citable papers (research output) may have to give up the opportunity of working with the business, and vice versa.

However, the high correlations among some of the indicators suggest that there might be multicolinearity which renders some of them redundant. This is verified next.

Is there a multicolinearity problem?

To verify the existence and severity of multicolinearity among the indicators, a multiple regression was run with the results shown in Table 2. As can be seen therein, the five indicators explained practically all variance of Overall with an adjusted coefficient of multiple determination (adjusted R^2) of 99.8%. As would be expected, the unstandardized b-weights are largely consistent with the assigned weights of the indicators (e.g., 30% each for Teaching, Citation, and Research and 10% for Internationalism and Industry Income combined), since the weights were used to calculate the Overall.

<u>Table 2 about here</u>

However, the standardized beta-weights show that the six indictors do not contribute to the Overall in the proportions of the weights as assigned by THEWUR. In terms of variance components, Teaching explains 31%, Citation 22%, Research 34%, Internationalism 9%, and

Industry Income 4%. In short, the Overall score is not made up of the indicator scores as intended and, therefore, to think of an Overall score as comprising indicator scores in the ranker-stipulated proportions is misleading. For example, Citation is originally given a weight of 30% but it turns out that it is under-weighted as 22% in the data. A discussion on this problem and issue is beyond the scope of this paper but the implications and a solution are discussed and demonstrated elsewhere (Soh, 2013a, 2013b).

In the context of the present study, it is worthy of note that although all six indicators contributed to the Overall with statistical significance (see the t- and p-values) but the results also caution that there is a high degree of multicolinearity. The problem of multicolinearity is that it signals that there are considerable overlap among the indicators such that some of them are redundant. More seriously, it make the estimate (in this case, Overall) unstable and hence cannot be trusted.

Evidence of multicolinearity are shown in the Tolerance column in Table 2. Tolerance is defined as $(1 - R^2)$ where R is the multiple regression coefficient of a particular predictor predicted by all other predictors in the regression analysis. If R is large, much of the variance of the particular predictor is predicted by the other predictors, and this renders the particular predictor redundant because what it can explain is already explained by the other predictors. Thus, Tolerance shows what a predictor can explain what is already explained by other predictors in the model, hence it is redundant. The convention is that a Tolerance smaller than 0.2 signals that the predictor is redundant and therefore not needed.

By applying the concept of Tolerance, it is obvious from Table 2 that four of the five indicators are redundant since they have values less than 0.2, except Industry Income. Thus, there is a rather severe problem of multicolinearity and most of the indicators are therefore not needed for the calculation of the Overall score. This gives rise to the question of which indictor to retain and which can be dispensed with. To this we next turn.

Which are the indicators to retain?

As Table 2 shows, the multicolinearity problem is found mainly among the first three indicators (i.e., Teaching, Research, and Citation). And, as shown in Table 1, Teaching and Research correlate highly and they also have high correlations with Overall. Moreover, Teaching has a higher correlation with Citation than has Research. These suggest that Research can be retained. At the same time, Citation has a moderate correlation with Overall but a low one with Teaching. These suggest that Citation can be retained, especially when it has a non-significant correlation with Research, because it explains part of the variance not already explained by another predictor, that is, Research.

The two remaining predictors (i.e., Internationalism and Industry Income) both have non-significant and low positive correlations with Overall, indicating that these did not contribute to Overall. Moreover, as shown in Table 2, Internationalism also has a multicolinearity problem. This leaves Industry Income as a possible indictor to be retained.

In short, Research, Citation, and Industry Income seem to be good predictors of Overall. They also covers the main functions of the present-day university as a generator of knowledge and partner of the business. At the same time, Citation can be seen as a direct indication of academic excellence. A multiple regression was then run with Research, Citation, and Industry Income as predictors. The results in Table 3a show that 95.3% of Overall is predicted. However, the Tolerance for Research is low, indicating that it is still a multicolinearity problem.

<u>Table 3a about here</u>

A further analysis was run with Research excluded to get rid of multicolinearity. The results in Table 3b show that this is successful since the Tolerances are both greater than 0.2. However, only 31.6% of Overall is predicted and this is rather low. This gives rise to the issue of tolerating multicolinearity or having much of the criterion unexplained. This obviously is a choice between two evils.

<u>Table 3b about here</u>

It may be argued that the two-indicator model is rather inefficient in predictive power and it is more desirable to retain Research so that more criterion variance can be explained, with due caution that there is still multicolinearity. With this in view, a stepwise multiple regression was run and the results in Table 4 show that Research alone explains 84.6% of Overall variance. When Citation is entered as the second predictor, it explains an additional 10.5% and Industry Income explains a further 0.1% only when entered as the third predictor. Thus, it can be argued that only Research and Citation need be retained to predict Overall.

<u>Table 4 about here</u>

At this point, with the exclusion of Industry Income which does not add much to the prediction, new Overall scores were calculated by summing the Research and Citation scores. When a stepwise multiple regression was run, Research explains 75.7% of the new criterion variance. When Citation was entered, it explains an additional 24.3% of the criterion variance. Thus, the two indictors were able to explain all criterion variance. Since both Tolerances are far greater than 0.2, this two-indicator model is free from multicolinearity

Table 5 about here

What is the impact of the reduced model?

With new Overall calculated for the two-indicator model, an important question is how well the correlation between the original rankings and the new rankings is. The Spearman's rank difference correlation turns out to be ρ=.914, which is very high for practical purpose. The new Overall, however, yielded new rankings for the universities, with some gaining positions and other losing. Table 6 shows the extent of changes in rankings.

Table 6 about here

Another relevant and interesting question is which kind of universities gain and which lose. This was studied by finding the correlations between gain in ranking (i.e., difference between the original and new rankings) and the original indicators scores. As the correlations in Table 7 suggest, universities with higher original Overall and five of the six indicator scores tend to lose and thereby getting lower new rankings, and vice versa. The exception is with Citation where those with higher Citation tend to gain in position. This indicates that Citation has become more critical in the new Overall (and new ranking) which have enhanced academic quality in view of the nature of Citation. Moreover, the negative correlations for Internationalism and Industry Income (which have been excluded in the new Overall) suggest that, in the original ranking, universities with high scores for these two indicators have had their original Overall grossly inflated and were thereby assigned ranking better than they deserved for non-academic reasons.

Table 7 about here

Universities with their respective new rankings are listed in the Appendix. It is interesting to note that the top 25 universities have small positive or negative changes in positions and those beyond have much wider range of gain or loss.

Discussion and Conclusion

Weight-and-sum (WAS) is an approach commonly used for social and education ranking systems. A ranking agent first choses a set of indicators (each of which may subsumes data at an even lower level for various sources), decides on the weight for each indicator as the coefficient of its score, with the weight supposedly reflecting its relative importance. After weighting, the scores for the several indicators are summed to form the Overall score. Based on the Overall, the participating universities or countries are ranked. The rankings thus form a league table released for public consumption. This looks rational enough but has several conceptual and statistical/measurement problems (Soh, 2013a). More often than not, the rationale for choosing the indicators and their weights and not others is not articulated. This is so for THEWUR and many other ranking systems.

At this point of time, the three most popular WUR are the Academic Ranking of World Universities (ARWU; ShanghaiRanking Consultancy, 2013), the QS World University Ranking (QSWUR; Quacquarelli Symonds Limited, 1994-2013), and the Times Higher Education World University Rankings (THEWUR; TLS Education Ltd. 2012a). All three systems use the WAS approach but operationalized academic excellence differently.

For indicators, the ARWU has Alumni and Staff winning Nobel Prizes and Field Medals (together 30%), Highly Cited Researchers (20%), Articles Published in *Nature* and *Science* (20%), Science Citation Index and Social Science Citation Index (20%), and Per Capita Performance on the Above (10%). QSWUR has as its indicators Academic Peer Review 40%), Employer Review (10%), Citation per Faculty Member (20%), Student/Faculty Ratio (20), International Students (5%), and International Faculty (5%), while THEWUR has Teaching (30%), Research (30%), Citation 32.5%), Industry Income (2.5%), and International Mix (5%). It is obvious that ARWU's indicators all pertain to academic achievements in various forms and are outcome measures indexing academic excellence in its strict definition. In contrast, QSWUR and THEWUR both have a mixture of academic and administrative measures rendering their Overall less 'pure' as indication of the strictly defined academic excellence.

Where THEWUR is concerned, Teaching, Research, and Citation may be taken as true reflection of the level of performance outcome of the university's main functions of instruction and generation of knowledge. However, whether Industry Income and International Mix are to be considered causal or effect factors is not clear. Moreover, while citation logically depends on publication which in turn logically depends on research, and therefore a high correlation between the two indicators can be expected. However, the non-significant correlation (r=.183, in Table 2) suggests that this logical flow of events may not actualized; this indicates that research does not necessary result in articles published in learned journals but in some other formats or even not published at all but used for other practical purposes (e.g., feeding the relevant industries). Thus, strictly speaking, only Teaching, Research, and Citation are *bona fide* indicators of academic excellent in the case of THEWUR.

Even if the choice of indicators and their weights are accepted, for the WAS approach to work, it needs to satisfy several conditions in it implementation. First, the data of the indicators need to meet certain measurement standards such as reliability and validity. Reliability has to do with the precision of the data such that it is stable, reproducible, or internally consistent, in short, it must contains the possible minimal random measurement errors caused by sampling of, say, test items and respondents. Validity has to do with accuracy in that what is measured is really what needs be measured. Unfortunately, Many social and educational ranking take these for granted and do not report the extent of reliability and validity. Such information is needed by the rank-users for deciding how much the ranking results can be trusted.

Secondary, the data of the indicators should be minimally vulnerable to subjectivity or conscious faking (as a form of acquiescence bias) if it is to be reliable and valid. However, many of the data used in social and educational rankings are social measures which are notoriously fallible, and triangulation, for example, has been commonly used as a check against such fallibility. Unfortunately, this safeguard procedure does not seem to have been consciously applied in social and educational ranking.

Thirdly, since social and educational rankings deal with human activities, the data tend to be correlated, with some indicators repeating the information already available in other indicators in the same system. Some of the overlapping indicators are therefore redundant and are not needed. For the multiple regression perspective, the indicators function best when they each correlate highly with the criterion (e.g., the Overall) but among them the correlations are at the minimum or very low. Inconsistency between assigned and intended indicator weights is one problem of WUR uncovered through multiple regression (Soh, 2013b). Multicolinearity, the main concern of this article, is another problem.

With the above possible pitfalls, rank-users are well-advised to be cautious when reading a set of WUR results and not to be oblivious to possible interpretation problems which may lead to unwarranted decisions and misguided actions. The analysis of the THEWUR 2012-2013 ranking results done here illustrate these issues.

Recall that Teaching and Research are two important indicators in the THEWUR as they each is given a weight of 30% summing up to 60%. The data for these indicators was obtained dominantly from by-invitation-only surveys, although other more objective information was also used together. For instance, in addition to survey data, Teaching also has staff/student ratio as a proxy measure for teaching quality and doctoral/bachelor ratio degrees awarded. Likewise, for Research, in addition to survey data, includes objective data such as university research income, research productivity, and number of papers published in academic journals. The ranking agent reported that a total of about 16,600 responses were collected for Teaching and 10,000-plus responses for Research. This is worrisome because subjective survey data plays a major part in these two heavily weighted indicator and the ranking agent implies that very large sample sizes ensure trustworthiness, when the issue is in fact not *reliability*, which large sample sizes may and may not guarantee, but the more critical *validity*.

Moreover, it is not made clear how the by-invitation surveys were conducted. If groups of respondents from different regions were asked to respond to different lists of universities, the groups in actuality completed different surveys. Even if they were given the full list of several hundred participating universities to rate, it is highly doubtful that any respondent is so knowledgeable to validly evaluate all the universities there are. In both cases, pooling the respondents as one huge sample does not constitutes a very large scale survey, which the ranking agent seem to be wanting to convince the rank-users. Besides, it is a well-known fact that social measures are highly fallible and have to be interpreted with due caution; and, the two reputation surveys fall into this category.

Citation is another important indicator carrying a weight of 30%. Citation depends on publication which is in turn dependent on quality research. In this sense, Citation is critical to the academic excellence of university and is a *bona fide* measure of the academic quality of an institution. Moreover, the information takes the form of actual counts of indexed articles in respectable citation indices. Citation is therefore less vulnerable to extraneous subjective influences like Teaching and Research which are mainly social surveys. However, this does not rule out influences working before THEWUR uses the information and therefore has not control over the data quality. Thus, even Citation needs be used with due caution without assuming perfect validity, for instance, Panaretos and Malesios (2012) commented thus,

> Measuring the research performance of a scientist by using only his or her bibliometric data is already more or less restrictive by default, let alone by measuring the citation performance with only a single one of the metrics available (p. 165).

It is readily appreciated that language has an impact on publication and hence citation, since most citation indices are in English and are more likely to include journals published in that language. For instance, the new edition of the Science Citation Index includes over 100 Chinese journals when it covers a total of 6,595 journals, with China's more than 500 science and technology journals excluded, most of which are in Chinese (Jia and Sun, 2009). The situation is so miserable that the editor-in-chief of China Medication Association journals suggested that SCI be read as 'Stupid Chinese Index'. The language problem applies to other East Asian countries like Korea and Japan as it was reported that the top 10 countries there had a Citation mean of 54.6 while 20 Western universities, matched for THEWUR 2011-2012 Overall rankings, had 83.2 (Soh, 2012, Table 2).

Having many indicators in a WUR system gives the impression that the Overall, which is a weighted sum of indicator scores, is comprehensive to cover important aspects of academic excellence. While this may be a conceptually welcome feature, its validity (and hence usefulness and usability) depends on whether the Overall in actuality is what has been promised by the ranking agent, in other words, whether the Overall really is made up of the indictor scores in the stipulated proportions without being confounded by inter-indicator correlations.

As demonstrated above, the correlations show that Teaching and Research overlap considerably but Internationalism and Industry Income made very little contribution to Overall. From the measurement perceptive, although Teaching and Research are supposed to be two different aspects of the work of university, but they tell very much the same story about the efficacy of the university in terms of academic excellence when the two indictors correlate so highly. The overlap suggest that one of the two would suffices and the other is redundant as it does not give much additional information about the university's efficacy. As for the two non-contributing indictors (i.e., Internationalism and Industry Income), their presence in the Overall just cloud the clarity of what Overall measures and renders its meaning muddled.

In short, with the THEWUR's six indicators, the Overall's validity is murky and it lacks conceptual clarity, because there are highly overlapping measures as well as non-contributing measures simply weighted and summed. Such confounding make the meaning of the Overall (and hence its interpretation and usefulness) vague and uncertain.

Here, the ranking agent and the rank-user are faced with a dilemma: whether to follow clearly defined *new* Overall (as represented by Research and Citation) or to retain a confounded and confused *original* Overall (made up of six indicators, two of which overlap considerably and two no-contributing). Bear in mind also that the original Overall is suffering severely from multicolineartiy while the new Overall does not.

Although the choice clearly is slanting toward the new Overall, such a rational choice may not be preferred especially by universities which will lost position for it and thereby have poorer ranking. Universities which have gains may be happy to see better images of themselves, and those with neither gain nor loss may be indifferent; naturally, those universities which have lost position may see the new Overall with mistrust and even objection. Nonetheless, the critical issue is which of the two Overalls represents academic excellence more truthfully and which misinforms and misguides. The new Overall is a purer, uncompounded measure of academic excellence without multicolinearity. The larger number of indicators and the seeming comprehensiveness of the original Overall are irrelevant, since its meaning is masked by overlapping and non-contributing indicators.

Multiconlinearity and non-contributing indicators have been found for THEWUR. The present writer is confident that what have been found for THEWUR will also be obtained for other WUR systems had their data been re-analysed likewise, and even other social and educational rankings as long as they use the WAS approach to select and weight indictors. WAS is a common-sense approach which has its attraction in its seeming simplicity, but this can turn out to be simplistic. With the example shown by the THEWUR data and the discussion on the pros and cons of using fewer indicators for clarity, the way ahead is clear, unless the ostrich's approach to hide the problem by not seeing it is preferred.

References

d'Hombres, B. and Saisana, M. (19-20 June 2013). CRELL: critiquing global university rankings and their methodologies. *GlobalHigherEd*. http://globalhighered.wordpress.com/2009/01/27/crell-critiquing-global-university-rankings-and-their-methodologies/

Hazelkorn, E. November 11, 2007. ‘‘OECD: Consumer Concept Becomes a Policy Instrument.’’ University World News, Issue No. 5. http://www.universityworldnews.com/article.php?story=20071108144018483&query=university+ranking

Jia, H. and Sun, T. (2009). Science index a target for criticism in China. *Chemistry World*. Royal Society of Chemistry.

http://www.rsc.org/chemistryworld/Issues/2009/August/ScienceIndexATargetForCriticismInChina.asp

Lincoln, D. (2012). Rankings: an idea whose time has come, and gone. *InsideHigher Ed*, February 28. http://www.insidehighered.com/blogs/world-view/rankings-idea-whose-timehas-come-and-gone.

Mumby, D. G. (June 25, 2012). Choosing a University for an Undergraduate Degree? Ignore the Ranking Lists. *My Graduate School*. http://mygraduateschool.wordpress.com/2012/06/25/choosing-a-university-for-an-undergraduate-degree-ignore-the-rankings-lists/

Panaretos , J. and Malesios, C. C. (2012). Commentary on Ruscio et al.: Measuring ScholarlyImpact Using Modern Citation-based Indices. *Measurement, 10*, 164-166.

Quacquarelli Symonds Limited (1994-2013). *QS Top Universities: World University Rankings, Guides, and Events.*

http://www.topuniversities.com/university-rankings

Rauhvargers, A. (2011). *Global University Ranking and Their Impact. EUA Report on Rankings 2011*. European University Association.

ShanghaiRanking Consultancy (2013). *Academic Ranking of World Universities*.

http://www.shanghairanking.com/ARWU2012.html

Soh, K. (2013a). *Social and Educational Rankings: Problems and Prospect*. New York: Untested Ideas Research Center.

Soh, K. (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. doi.org/10.1080/1360080X.2013.775959

Soh, K. (2012). World university rankings: What is in for top ten East Asian universities. New Horizons in Education, 60(2), 36-50.

Staff Writers (5 June 2012). 'Twelve questionable factors behind traditional university ranking.' Online Colleges: http://www.onlinecolleges.net/2012/06/05/12-questionable-factors-behind- traditional-university-rankings/

Streiner, D. L. (2003). Being inconsistent about consistency: when coefficient alpha does not and doesn't matter. *Journal of Personality Assessment*, *80(3)*, 217-222.

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

TSL Education Ltd. (2012b). *World University Ranking 2013-2014 Methodology*. http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/world-ranking/methodology

Table 1, Correlations

	Overall	Teaching	Research	Citation	Internat-ionalism	Industry Income
Overall	1.000	**.927**	**.920**	**.485**	.156	.153
Teaching		1.000	**.876**	**.289**	-.011	.143
Research			1.000	.183	.108	**.239**
Citation				1.000	-.081	**-.297**
Internationalism					1.000	.139
Industry Income						1.000

Note: Coefficients in bold are highly correlated (p<.01, df 99, two-tailed)

Table 2. Multiple Regression

	b-weight	Beta-weight	t	p	Tolerance
Intercept	-.269	.000	-	-	-
Teaching	.307	.426	40.245	.001	.011
Citation	.297	.299	44.153	.001	.023
Research	.295	.463	56.749	.001	.009
Internationalism	.079	.127	26.085	.001	.114
Ind. Income	.027	.049	9.564	.001	.425

Adjusted R^2 =.998

Table 3a. Multiple Regression

	b-weight	Beta-weight	t	p	Tolerance
Intercept	3.623	.000	-	-	-
Research	.538	.843	36.547	.001	.060
Citation	.344	.347	14.818	.001	.261
Industry Income	.028	.051	2.156	.034	.801

Adjusted R^2 = 0.953

Table 3b. Multiple Regression

	b-weight	Beta-weight	t	p	Tolerance
Intercept	9.071	.000	-	-	-
Citation	.171	.580	6.725	.001	.630
Industry Income	.574	.307	3.561	.001	.815

Adjusted R^2 = 0.316

Table 4. Stepwise Multiple Regression

	Step 1	Step 2	Sep 3
Adjusted R^2	.846	.952	.953
R^2 increase	.846	.105	.001
Beta-weight			
Research	.921	.858	.843
Citation	-	.330	.347
Industry Income	-	-	.051

Table 5. Stepwise Multiple Regression for New Overall

	Step 1	Step 2	Tolerance
Adjusted R^2	.757	1.000	-
R^2 increase	.757	.243	-
Beta-weight			
Research	.872	.777	.964
Citation	-	.499	.964

Table 6. Changes in Rankings

Change in position	Frequency
20 or more	5
11 to 20	12
1 to 10	32
No change	4
-1 to -10	32
-11 to -20	11
-20 or more	5

Table 7. Correlations

	Correlation with gain
Overall	-.118
Teaching	-.317
Research	-.186
Citation	.550
Internationalism	-.314
Industry Income	-.425

Appendix: Universities with new Ranking

Institution	Rank	Gain	Institution	Rank	Gain
California Institute of Technology	1	0	Utrecht	52	22
Harvard	2	0	Australian National	53	-5
Princeton	3	3	Rice	54	11
California. Berkeley	4	4	Pohong U of Science & Tech	55	5
Stamford	5	-1	Erasmus University Rotterdam	56	17
Oxford	6	-4	KU Leuvem	57	4
Cambridge	7	0	Ohio State	58	1
Massachusetts Institute of Tech	8	-3	Hong Kong U of Science & Tech	59	-2
California, Los Angeles	9	3	Hong Kong	60	-17
Chicago	10	-1	Queensland Australia	61	2
Yale	11	0	Colorado Boulder	62	35
Johm Hopkins	12	3	California, Irvine	63	30
Imperial College London	13	-3	Pittburgh	64	14
Michigan	14	4	Purdue	65.5	-3.5
Pennsylvinia	15	1	Manchester	65.5	-7.5
Columbia	16	-3	Amsterdam	67	16
Cornell	17	2	Kyoto	68	-16
Swiss Federal Institute of Technology	18	-4	Korea Ad Institute of Sc & Tech	69	-13
Northwestern	19	3	Notre Dame	71	19
Carnegie Mellon	20	4	Durham	71	9
Duke	21	-4	Seoul National	71	-27
Toronto	22	-2	Bristol	73	6
Washington	23	2	Tufts	74	6
University College London	24	-3	Tsinghua	75	-25
Texas at Austin	25	2	Vanderbilt	76	12
California, San Diego	26	14	Southern California	77	-7
Boston	27	23	Helsinki	78	22
Illinois at Urbana Champaign	28	1	Ecole Normle Suprrieure	79	-14
Tokyo	29	-6	Ghent	80	5
California, Santa Barbara	30	3	Freie Universitat Berlin	81.5	4.5
Georgia Institute of Technology	31	-3	Ecole Polytechnique	81.5	-11.5
Wisconsin-Madison	32	-2	York	83	17
British Columbia	33	-2	Emory	84.5	-4.5
Washington University in St Louis	34	8	Universitat Basel	84.5	-10.5
New York	35	5	McMaster	86	6
North Carolinaat Chapel Hill	36	11	Sydney	87	-15
Pennsylvania State	37	12	Nayang Technological	88	-12
Edinburgh	38	1	Case Western Reserve	89	-1
King's College London	39.5	-1.5	Peking	90	-45
Melbourne	39.5	-5.5	Universitat Heidelberg	91	-23
Ecole Poly. Federale de Lausanne	41.5	-4.5	Michigan State	92	-9
National University of Singapore	41.5	-15.5	Groningen	93	5
Minnesota	43	3	Rochester	94	1
Karolinska Institute	44	-8	Wagenigen U. and Research Center	95	-18
London Sch. of Econ. & Political Sc.	45	-13	Humboldt-Universitat zu Berlin	96	-2
Brown	46	6	Technische Universitat Munchen	97	-10
McGill	47	-12	Delft University of Technology	98	-29
Georg-ugust Universitat Gottingen	48	15	Monash	99	-8
California, Davis	49	3	Maastricht	100	-2
Leiden	50.5	16.5	Universite Pierre et Marie Curie	101	-5
Ludwig-Maximilians U. Munchen	50.5	4.5	-	-	-