JRC Scientific and Technical Reports

Higher Education Rankings: Robustness Issues and Critical Assessment

How much confidence can we have in Higher Education Rankings?

Michaela Saisana and Beatrice D'Hombres

EUR 23487 EN 2008

JRC EUROPEAN COMMISSION

ipsc Institute for the Protection and Security of the Citizen

CRELL Centre for Research on Lifelong Learning

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen
Centre for Research on Lifelong Learning (CRELL)

Contact information
Address: Michaela Saisana, JRC, TP361, via E. Fermi 2749, 21027 (VA), Italy
E-mail: michaela.saisana@jrc.it
Tel.: +39-0332-786572
Fax: +39-0332-785733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

Composite Indicators website: http://composite-indicators.jrc.ec.europa.eu/

JRC 47028

EUR 23487 EN
ISBN 978-92-79-09704-1
ISSN 1018-5593
DOI 10.2788/92295

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Abstract

The Academic Ranking of World Universities carried out annually by the Shanghai's Jiao Tong University (mostly known as the 'Shanghai ranking') has become, beyond the intention of its developers, a reference for scholars and policy makers in the field of higher education. For example Aghion and co-workers at the Bruegel think tank use the index – together with other data collected by Bruegel researchers - for analysis of how to reform Europe's universities, while French President Sarkozy has stressed the need for French universities to consolidate in order to promote their ranking under Jiao Tong. Given the political importance of this field the preparation of a new university ranking system is being considered by the French ministry of education.

The questions addressed in the present analysis is whether the Jiao Tong ranking serves the purposes it is used for, and whether its immediate European alternative, the British THES, can do better.

Robustness analysis of the Jiao Tong and THES ranking carried out by JRC researchers, and of an ad hoc created Jiao Tong-THES hybrid, shows that both measures fail when it comes to assessing Europe's universities. Jiao Tong is only robust in the identification of the top performers, on either side of the Atlantic, but quite unreliable on the ordering of all other institutes. Furthermore Jiao Tong focuses only on the research performance of universities, and hence is based on the strong assumption that research is a universal proxy for education. THES is a step in the right direction in that it includes some measure of education quality, but is otherwise fragile in its ranking, undeniably biased towards British institutes and somehow inconsistent in the relation between subjective variables (from surveys) and objective data (e.g. citations).

JRC analysis is based on 88 universities for which both the THES and Jiao Tong rank were available. European universities covered by the present study thus constitute only about 0.5% of the population of Europe's universities. Yet the fact that we are unable to reliably rank even the best European universities (apart from the 5 at the top) is a strong call for a better system, whose need is made acute by today's policy focus on the reform of higher education. For most European students, teachers or

researchers not even the Shanghai ranking – taken at face value and leaving aside the reservations raised in the present study – would tell which university is best in their own country. This is a problem for Europe, committed to make its education more comparable, its students more mobile and its researchers part of a European Research Area.

Various attempts in EU countries to address the issue of assessing higher education performance are briefly reviewed in the present study, which offers elements of analysis of which measurement problem could be addressed at the EU scale.

Higher Education Rankings: Robustness issues and Critical Assessment

Michaela Saisana and Beatrice D'Hombres

Executive Summary

It has been often stated that European growth has been disappointing during the past three decades, remaining persistently lower than in the United States and this is to be attributed to the state of innovation and higher education in Europe. Such a conclusion has been in part stirred by the publication, since 2003, of the 'Shanghai Jiao Tong University Academic Ranking of World Universities' (henceforth SJTU), which measures university research performance across the world. The SJTU ranking tends to reinforce the evidence that the US is well ahead of Europe in terms of cutting-edge university research. Since 2004, the Times Higher Education Supplement (henceforth THES), also publishes every year an international university ranking which follows a somewhat different set of indicators some of which are based on expert opinion. These rankings receive at present worldwide attention. They can potentially be used by prospective students when choosing which university to attend. University Ranking might also have a much wide impact on institutions' reputations and potentially on the behavior of academics, businesses and potential benefactors. Governing bodies take an interest in them as a means of assessing institutional performance, sometimes seizing on them in default of other, more sensitive indicators of institutional performance.

There clearly is a demand for rankings in the field of higher education, but there are also questions about the quality, impact and eventual validity of the conclusions, which in turn depend heavily upon the choice of indicators, the suitability of the methodologies used, the transparency of the processes and the robustness of the rankings.

University rankings are very appealing, in that they provide a single number that allows, at a glance, to situate a given university in the worldwide context. However, this very simplicity of use can be highly misleading in that most rankings are based on a simple formula that aggregate subjectively chosen indicators.

This JRC report has set four main goals:

- To throw light on the methodological issues and eventual limitations of the SJTU and THES rankings;
- To assess the robustness of the two higher education ranking systems with a view to identify for which universities these ranking can be reliably used to draw conclusions;
- To propose, if possible and despite the known limitations of the currently available indicators in the THES and SJTU, an approach that combines these pieces of information in the least biased way;
- To identify whether the average European university lags indeed behind the average US university based on the set of twelve indicators of the THES and SJTU frameworks.

To achieve these goals, besides the classical tools of multivariate analysis, we carry out a thorough uncertainty and sensitivity analysis of the 2007 SJTU and THES rankings for the universities that are ranked Top100 in the SJTU and Top200 in the THES (88 universities in total). We take into account a plurality of scenarios in which we activate simultaneously different sources of uncertainty. In these scenarios we deviate from the classic approach – also taken in the two university ranking systems - to build a composite indicator by a simple weighted summation of indicators. The sources of uncertainty we acknowledge cover a wide spectrum of methodological assumptions (all with their advantages and implications). Subsequently, a frequency matrix of the university ranks is calculated across the different simulations. Thus simulations may differ from one another in the inclusion/exclusion of an indicator, the choice of weights, the aggregation rule and so on. Such a multi-modeling approach and the presentation of the frequency matrix, rather than the single ranks, allows one to deal with the criticism, often made to league tables and rankings systems, that ranks are presented as if they were calculated under conditions of certainty while this is rarely the case.

Multi-modelling approach results on the assessment of university ranks

THES & SJTU (130 scenarios)

	Rank 1-5	Rank 6-10	Rank 11-15	Rank 16-20	Rank 21-25	Rank 26-30	Rank 31-35	Rank 36-40	Rank 41-45	Rank 46-50	Rank 51-55	Rank 56-60	Rank 61-65	Rank 66-70	Rank 71-75	Rank 76-80	Rank 81-88	(1) -THES &SJTU	(2) - THES	(3) - SJTU	Median rank
Harvard Univ	100																				1
Univ Cambridge	98																				2
California Inst Tech	59	41																			4
Univ California - Berkeley	72	28																	X		4
Massachusetts Inst Tech (MIT)	81	19																			5
Stanford Univ	45	43	12																		6
Columbia Univ		90	8																		8
Univ Oxford	22	76																			8
Princeton Univ	7	78	15																		8
Univ Chicago		72	28																		9
Yale Univ	18	42	41																		10
Cornell Univ		6	94																		12
Univ Pennsylvania			100																		13
Imperial Coll London			79	20																	14
Univ Coll London			75	24																	15
Johns Hopkins Univ			22	75																	17
Univ California - Los Angeles			15	70	7	8													X		18
Swiss Fed Inst Tech - Zurich			8	49	21	18	5												X		19
Tokyo Univ				66	28																19
Univ Wisconsin - Madison				50	35	8	6												X		20
Univ Toronto				28	49	20													X		22
Kyoto Univ				22	47	22	9														23
Univ Michigan - Ann Arbor				14	42	5			8	11	7		8					X			24
Univ California - San Diego				13	42	17	12	8	7									X	X		25
Duke Univ				22	35	7				6	7	8	8	7				X			25
Univ Washington - Seattle				17	42	29	6	6											X		25
Northwestern Univ					44	54															26
Univ Illinois - Urbana Champaign					22	35	23	10	5	5								X	X		29
McGill Univ				11	16	30	6			7		6	16	5				X	X	X	29
Carnegie Mellon Univ				9	15	32	35	8												X	30
Univ Manchester				5	20	28	37	10												X	30
Australian Natl Univ					14	32	42	8											X	X	31
Univ British Columbia						40	48	10													31
Univ Edinburgh					11	37	42	10												X	31
New York Univ					5	27	40	20	7												33
Univ Bristol						6	27	42	18											X	37
Brown Univ							10	53	28	8										X	40
Ecole Normale Super Paris						5	18	34	22	12	5							X	X	X	40
Tech Univ Munich						12	17	22	18	13	5							X			40
Univ Texas - Austin							12	36	36	12											40
King's Coll London						5	14	27	21	12	12							X	X	X	41
Univ Maryland - Coll Park							29	17	28	19	5								X		41
Vanderbilt Univ							8	26	32	22	7							X			42
Univ Melbourne							9	28	31	19	8							X	X	X	43
Univ Munich							10	22	28	26	11										44
Univ Sheffield								26	22	20	14	7		5				X		X	46
Univ Heidelberg								12	25	27	20	12						X			48
Washington Univ - St. Louis							8	12	16	8	5		6	8	16	8	7	X	X	X	49
Univ Nottingham								8	21	22	13	12	8		7	5		X		X	50
Univ Colorado - Boulder								9	22	16	22	13	8	5				X	X	X	51
Univ Pittsburgh - Pittsburgh								5	11	25	12	6	6	22	10			X			53
Univ Leiden									8	26	24	28	11					X			55
Purdue Univ - West Lafayette									5	15	32	23	14	5				X			55
Univ Paris 06								7	10	12	19	16	22	7				X	X	X	55
Univ Birmingham									12	18	18	19	12		8	9		X		X	56
Case Western Reserve Univ										6	36	25	12	5	8	5		X		X	57
Univ Basel										16	22	14	11	11	6	8	8	X		X	58
Univ Helsinki										5	22	35	25	8	5			X			59
Univ Rochester										5	12	40	23	11	6			X			59
Univ Zurich										14	10	34	18	6	8			X	X		59
Uppsala Univ										12	17	16	17	16	12	7		X	X		61
Boston Univ									7	18	15	8	8	15	20	8		X	X	X	62
Univ California - Santa Barbara								10	8	8	12	8	9	18	12	10	5	X		X	62
Osaka Univ								5	12	5	9	10	12	12	10	14	10	X	X		64
Rice Univ										9	21	7	18	22	14	5		X		X	64
Univ Southern California											5	17	27	32	11			X		X	65
Univ Minnesota - Twin Cities								5	9	11	8	9	8	20	22	5		X	X	X	65
Hebrew Univ Jerusalem												13	28	20	18	9	6	X			66
Pennsylvania State Univ - Univ Park										8	14	8	9	19	21	12	5	X		X	68
Rutgers State Univ - New Brunswick											10	13	12	14	10	20	19	X	X	X	70
Ohio State Univ - Columbus											5	7	9	32	24	20		X		X	70
Univ California - Davis									5	10	8	8	6	11	8	6	37	X		X	71
Univ California - Irvine											8	8	10	23	15	15	20	X		X	71
Tokyo Inst Tech													9	32	25	22	12		X	X	72
Univ Florida												5	8	22	27	30	7	X		X	73
Tohoku Univ												5	8	28	22	20	15	X			73
McMaster Univ													11	25	12	10	39	X			75
Univ Strasbourg 1												7	13	9	14	15	38	X			77
Lund Univ														15	19	31	31		X		78
Univ Oslo											9	12	13		7	16	39	X		X	78
Michigan State Univ														13	18	23	42	X			79
Indiana Univ - Bloomington												7	13	18		12	48	X			79
Univ North Carolina - Chapel Hill													7	9	15	22	48	X		X	80
Univ Freiburg												6	13	15	5	12	48	X			80
Univ Arizona													5	6	15	21	52	X		X	81
Nagoya Univ														5	5	22	65	X	X		83
Texas A&M Univ - Coll Station																18	77				84
Univ Goettingen																13	78	X			85

Legend:
Frequency lower 15%
Frequency between 15 and 30%
Frequency between 30 and 50%
Frequency greater than 50%
Note: Frequencies lower than 4% are not shown

(1) the hybrid THES&SJTU rank is highly sensitive to the methodological assumptions, (2) the THES rank is biased towards the set of the indicators, the set of weights and the aggregation rule, and (3) the SJTU rank is biased towards the set of the indicators, the set of weights and the aggregation rule.

This report puts forward three findings and three recommendations.

- **Finding 1:** While indicators and league tables are enough to start a discussion on higher education issues in Europe and benchmark it worldwide, they are not sufficient to conclude it.

As already widely discussed in the literature, the choice of the indicators reflects more the league tables compilers' opinion and the availability of internationally comparable data than the result of a consensus from the academic community. Both the THES and the SJTU rankings rely highly on bibliometric indicators and thus they tend to be biased towards English-speaking and hard sciences intensive institutions, leaving aside social and human sciences. In addition, regarding the THES, numerous authors are concerned with the use of expert-based indicators (50% of the total weight) and with the lack of transparency surrounding the process of selection of experts involved in the review process.

- **Finding 2:** The robustness analysis reveals that the rank of more than half of the institutions is highly sensitive to the methodological assumptions and the choice of indicators. Thus no conclusive inference regarding the relative performance for the majority of the universities can be drawn from either ranking.

For the majority of the universities we analysed, the THES or SJTU ranks have proven impossible to capture with adequate statistical robustness. The THES rank was found to be biased in the case of 26 universities, whilst the SJTU was found to be biased in the case of 33 universities. 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. This outcome calls in turn for a revision of the current set of indicators, or a further collection of indicators.

This finding further implies that the THES and SJTU rankings should not be used to discuss about the determinants of university performance (Aghion *et al.*, 2008) or to deliver policy messages on educational issues, as the assigned university rank largely depends on the methodological assumptions made in compiling the two rankings. For instance, we cannot conclude that Paris VI University performs

significantly better than McGill University though the difference in positions suggests a disparity in quality or performance.

- **Finding 3:** The average US university is not necessarily superior to the average European university. Moreover, the European universities studied here display more homogeneity than US universities.

An analysis of the 27 European universities and 48 USA universities that are ranked Top100 in the SJTU and Top200 in the THES shows that – contrary to current wisdom - the average US university is not necessarily superior to the average European university. The average US university has a better performance than the average European university in the *number of articles in Science and Social Citation Index,* in the *number of highly cited researchers* (SJTU indicators) and in the *citations per faculty* (THES indicator). Yet, the average European university has a better performance than the average American university in the *proportion of international staff* and the *proportion of international students.* For the remaining seven indicators analysed (in particular the two indicators related to the number of *Alumni* or *Staff winning Nobel prizes and field medal*), the performance of the average European university is comparable to the average US university. Regarding homogeneity issues, the European universities analysed have a more homogenous performance than their American counterparts in the majority of the indicators.

We recommend that the university ranking systems can and should be improved as follows:

- First, the indicators should be revisited along the lines of the recommendations and suggestions already provided by the Berlin Principles (see Box 1 in Appendix).

The Berlin principles place emphasis on league tables for universities that recognize the diversity of institutions, provide clear information about the indicators and target groups. The principles also provide recommendations on the way data should be gathered, processed in a transparent way and how final rankings should be presented. The THES ranking fails to comply with the Berlin

principles as, for instance, there is clearly a lack of information surrounding the construction of the two expert driven indicators.

- Second, the compilation of university rankings should always be accompanied by a robustness analysis based on a multi modeling approach. We believe that this could constitute an additional recommendation to be added to the already 16 existing Berlin principles.

The multi-modeling approach adopted in this report allowed us to show that the rank of most of the 88 institutions is highly dependant on the methodology chosen for the compilation of both rankings. We selected numerous scenarios that represent distinct, diverse and at times contradicting approaches in order to aggregate information on university performance. The multi modeling approach employed, has already proven to be useful in the development and validation of several composite indicators (e.g., Environmental Performance Index, Composite Learning Index, Alcohol Policy Index, Knowledge Economy Index) and was also included in the JRC/OECD Handbook on Composite Indicators.

Thereafter, while university rankings can not inform us about the real position of most of universities, given the statistical uncertainty associated with the ranks, a multi-modeling approach, like the one implemented in this report, allows to rank institutions in a range bracket. The upshot is that this way of doing is probably better than assigning a specific rank which is not representative of the real performance of the university.

- Third, the assessment of the universities performance based on the hybrid set of the twelve indicators used in the THES and SJTU rankings provides a more reliable average rank of the institutions.

The two sensitivity measures we used showed that the impact of the methodological assumptions is much lower when using the set of twelve indicators as opposed to either the THES or the SJTU indicators alone. Given the diversity of the indicators, as confirmed by correlation analysis, and the fact that the number of statistical dimensions in the combined THES&SJTU framework is twice the number of statistical dimensions for either the THES or the SJTU (result of factor analysis), more diverse aspects of universities are captured if all twelve

indicators are considered. The linkages between the THES indicators on one side and the SJTU indicators on the other are positive and significant, yet fair ($r \leq 0.58$). This result evidences the relatively low degree of overlap of information between the two sets and suggests that an eventual merging of the twelve indicators may provide a more holistic picture of the universities performance.

Even if all three previous recommendations are taken into consideration, one further issue remains: the high volatility of more than half of the universities we analysed. We recall the reader that these universities are considered the "elite" of the thousands of universities world-wide. If the ranks of those universities are full of uncertainty, let alone the ranks of the universities further down the classification ladder. This high volatility calls for a revision of the set of indicators, either by enriching it with other dimensions that are crucial to assessing university performance or by revising some of the existing indicators in order to remove some of the bias present (e.g., eliminate bias in favour of old and/or big universities and/or hard sciences).

A legitimate question is raised: when will the revision of the dataset of indicators reach a satisfactory level? Uncertainty and sensitivity analysis should be employed as a guide to achieve this goal. We would argue that the stopping criterion for the revision is reached when, upon acknowledging the methodological uncertainties that are intrinsic to the development of a ranking system, the space of inference of the ranks for the majority of the universities is narrow enough to justify a meaningful classification.

We hope that the debate of this study will

- lead to improvements to league tables methodologies;
- enable users to better understand the complexities of the league tables, and avoid misunderstanding them; and
- help higher education institutions develop approaches that help them satisfy the legitimate information needs of their stakeholders.

Our analysis, findings and recommendations are detailed in the remainder of this report.

Table of Contents

1. Introduction....14

2. Two world university rankings: criteria and weights....18
2.1 SJTU university ranking – main features....18
2.2 THES university ranking....20
2.3 Good or bad criteria: literature overview....22
2.4 Consistency of the two worldwide rankings....30

3. Main drivers of the two international university rankings....32
3.1 Correlation analysis....32
3.2 Statistical dimensionality of the datasets....37
3.3 Relative contribution of indicators to the overall scores....39

4. Uncertainty and Sensitivity Analysis....41
4.1 Considerations on the robustness assessment....42
4.2 Robustness assessment of the THES ranking system....49
4.3 Robustness assessment of the SJTU ranking system....58
4.4 Hybrid framework, multi-modelling approach....65

5. Can university rankings be used as a guide for higher education policies?....75

6. Conclusions....83

7. References....89

ANNEX: Methodological boxes and additional information....92

List of Tables

Table 1. Shanghai Jiao Tong University Rankings (SJTU), 2007 19
Table 2. Times Higher Education Supplement Rankings (THES), 2007 21
Table 3. 2007 SJTU indicators: what has been said so far 26
Table 4. 2007 THES indicators: what has been said so far 27
Table 5. THES Ranking: Methodological changes over the four editions 28
Table 6. SJTU Ranking: Methodological changes over the five editions 29
Table 7. Pearson correlation coefficients between the twelve indicators in the THES& SJTU frameworks 36
Table 8. Eigenvalues of the THES, the SJTU or the hybrid framework 38
Table 9. Squared factor loadings of the twelve indicators in the four main factors 38
Table 10. Analysis of shares for the indicators in THES and SJTU 40
Table 11. Checklist for building a composite indicator 43
Table 12. Examples of composite indicators that were assessed using a multi-modelling approach 45
Table 13. Scenarios for the assessment of the THES and SJTU 47
Table 14. Frequency matrix of the Universities ranks according to the THES framework 51
Table 15. THES Index: 59 universities whose simulated rank is highly sensitive to the methodological assumptions 54
Table 16. THES: 10 universities whose THES rank is non-representative of the simulated scenarios 55
Table 17. Impact of the exclusion of a single indicator from THES 57
Table 18. Frequency matrix of the Universities ranks according to the SJTU 59
Table 19. SJTU Index: 52 universities whose simulated rank is highly sensitive to the methodological assumptions 61
Table 20. SJTU Index: 6 universities whose SJTU rank is not representative of the simulated scenarios 62
Table 21. Impact of the exclusion of a single indicator from SJTU 65
Table 22. Frequency matrix of university ranks - hybrid THES & SJTU 67
Table 23. Hybrid THES&SJTU Index: 50 universities whose simulated rank is highly sensitive to the conceptual and methodological assumptions 69
Table 24. Hybrid Index: universities whose THES or SJTU rank is non-representative of the simulated scenarios 71
Table 25. Conclusions from the robustness assessment of the university ranks 76
Table 26. Benchmarking performance of European universities along the THES and SJTU indicators with respect to the average European University 78
Table 27. Benchmarking universities scores in the six THES and six SJTU indicators with respect to the average US University 80
Table 28. Benchmarking average European university scores in the six THES and six SJTU indicators with respect to the average US University 82

List of Figures

Figure 1. SJTU ranking versus THES ranking (common universities) 31
Figure 2. Sensitivity analysis: impact of the assumptions on the THES ranks 57
Figure 3. Sensitivity analysis: impact of the assumptions on the SJTU ranks 64
Figure 4. Sensitivity analysis: impact of the assumptions on the THES ranks 73
Figure 5. Recommended set of weights to be used when combining the twelve indicators of the THES and the SJTU 73

1. Introduction

It is has been often stated that European growth has been disappointing during the past three decades, remaining persistently lower than in the United States. This situation triggered the launch of the EU Lisbon agenda in 2000. That higher education has become increasingly important for growth in Europe can be understood in light of recent work by Acemoglu, Aghion and Zilibotti (2006), Sapir *et al.* (2004) and Aghion *et al*. (2008).

The recognition of this fact has been stirred greatly by the publication, since 2003, of the 'Shanghai Jiao Tong University Academic Ranking of World Universities' (henceforth SJTU), which measures university research performance across the world. The SJTU ranking tends to reinforce the evidence that the US is well ahead of Europe in terms of cutting-edge university research. Its only rival is the ranking computed annually, since 2004, by the Times Higher Education Supplement (henceforth THES), which follows a somewhat different set of indicators some of which are based on expert opinion. Both these rankings are now receiving worldwide attention.

Besides these two international university rankings, there are more than 30 national university rankings that are flourishing around the world (see Usher and Savino, 2006; European Commission, 2008). Various indicators, ranging from input indicators to outcome and process indicators, are combined into a single number in an attempt to represent overall university "excellence" (Salmi and Saroyan, 2007; Usher and Savino, 2006).

University is a multidimensional phenomenon, which makes it difficult to condense the diversified work going on within universities into a single number or ranking. Comparing universities which differ in focus and mission is also questionable. However, for governance purposes, for example economic rationale behind the allocation of public funds, accountability and transparency, there is an increasing interest in measuring and benchmarking "excellence" across universities. The easier access to higher education and the growing mobility of students have also rendered the comparison of university quality across countries even more appealing.

For these reasons, despite being a controversial topic, university rankings are receiving widespread attention. The yearly publications of the two international university rankings, SJTU and THES, are widely covered by the media and constitute an occasion for national governments to comment on the relative performances of their national universities. In France, the publication of the SJTU is always associated with a surge of articles in newspapers which either bemoan about the poor performance of French universities or denounce the inadequacy of the SJTU ranking to properly assess the attractiveness of the fragmented French higher education institutions landscape (see Le Monde, 27 February 2008; Liberation, 26 February 2008). University rankings are also used by national policy makers to stimulate debates about national university systems and ultimately can lead to specific education policies orientations.[1]

Window to the market oriented academic world, university rankings have become more and more popular and are now an integral part of the higher education landscape. At the same time, however, these rankings are subject to a plethora of criticism. Several authors (e.g., Marginson, 2007; Taylor and Braddock, 2007; Zitt and Filliatreau, 2007; Williams, 2007; Van Raan, 2007) have questioned the relevance with respect to university excellence of the underlying indicators included in either the THES or the SJTU. They outline that the chosen indicators are mainly based on research performance with no attempt to take into account the more versatile mission of universities (in particular teaching), and are biased towards large, English-speaking and hard sciences intensive institutions. The choice of indicators reflects more the league tables compilers' opinion and the availability of internationally comparable data than the result of a consensus from the academic community.

In response to this growing concern about the quality of ranking, in 2006, the International Ranking Expert Group (IREG) proposed a set of rules, named "Berlin Principles", of quality and good practice in the compilation of higher education rankings. The principles range from recognizing the diversity of institutions, providing clear information about the indicators and target groups to the way data should be gathered, processed and final rankings presented in passing, by more transparency on the methodology used for the calculation of scores (see Box 1 in the appendix for a summary of the Berlin principles).

[1] See for instance the French university funding reform to make public universities more competitive. The French Research and Higher Education Minister Pécresse often justifies the need of university reforms on the base of the absence of French universities in the top 20.

Whilst the discussion of the limitations of the twelve indicators underlying altogether the THES and the SJTU Indices have received proper attention and discussion in the relevant literature, to the best of our knowledge, there has been no attempt so far to examine in depth the impact on the universities ranks of the methodological assumptions made in compiling these two university ranking systems. In fact, although the heated debate about rankings and league tables also centres on their robustness and consistency, and methodologies used in compiling them, a thorough analysis of the statistical robustness and the "consequences" of the methodological assumptions made in THES and SJTU are still lacking. The purpose of this paper is to fill in this gap by quantifying how much the university ranks depend on the methodology (set of selected indicators, weighting scheme, normalization and aggregation method) used to build either the THES or the SJTU.[2]

We carry out a sensitivity analysis of the 2007 SJTU and THES rankings under a plurality of scenarios in which we activate simultaneously different sources of uncertainty that cover a wide and versatile spectrum of methodological assumptions (all with their advantages and implications). In this multi-modelling approach we estimate the frequency of the university ranks obtained in the different simulations (triggering the exclusion of an indicator, the weighting and the aggregation rule). Such a multi-modelling approach allows one to deal with the criticism, often made to league tables and rankings systems, that ranks are presented as if they were calculated under conditions of certainty while this is rarely the case. Thus we deviate from the classic approach – also taken in the two university ranking systems - to build a composite indicator by a simple weighted summation of indicators. Leaving aside the conceptual framework and selection of indicators used in either ranking (see Taylor and Braddock 2007 for a comprehensive discussion of the limitations of the two rankings from the conceptual point of view), this report is a contribution to the discussions and concerns of the methodological issues inherent in the two rankings.

This report sets three main objectives. The first is to throw a considerable amount of light on the approaches and limitations of the SJTU and THES rankings. The second is to assess the robustness of the two ranking systems with a view to identify for which universities the THES and SJTU ranking systems can be reliably used to draw conclusions. The final objective is to identify factors behind the

[2] Recently, the Centre for Higher Education Research and Information (CHERI) has published a report in which the two world rankings (and three English national rankings) are examined. The object of the study is to investigate how higher education institutions are influenced by league tables. The authors also discuss the methodologies underlying the rankings. However, the authors do not implement a sensitivity analysis of the two rankings.

differences in university performance across Europe compared to the United States given the set of twelve indicators included altogether in THES and SJTU.

Section 2 of the report describes the main features of the 2007 SJTU and THES rankings, including the underlying indicators, the associated weights and the aggregation rule. An overview of the methodological changes that have occurred over years in the rankings is offered together with a summary of the critics specific to each of the indicators used in the compilation of both rankings. We also briefly examine whether the SJTU and THES rankings give consistent results. Section 3 provides insights into the degree of association between the indicators of the two international rankings using correlation analysis, principal components and path analysis. In section 4, we carry out a robustness assessment of the two international rankings. We aim at examining to which extent the THES and SJTU rankings depend on the statistical methodology chosen. The analysis involves the simultaneous activation of various and versatile sources of uncertainty (e.g., triggering the exclusion of an indicator, the weighting and the aggregation rule). Section 5 discusses the use of these university rankings as a guide for higher education policies and offers a summary of the main conclusions on the reliability of the ranks discussed in detail in the previous section. The strong and weak performance of a university across the twelve indicators are discussed here, together with some considerations on the comparison of the European universities with the US universities. Section 6 wraps up the aims, the main finings and the recommendations of the study.

2. Two world university rankings: criteria and weights

In the following section we describe the features of the SJTU and THES rankings. As we will see below, while the SJTU ranking focuses on the research dimension of universities, the THES ranking attributes a large importance to the expert-based indicators. Table 1 and Table 2 display the different indicators and associated weights for the two 2007 SJTU and THES rankings. Table 3 and Table 4 overview the changes that have occurred over years in the methodology of both rankings while Table 5 and Table 6 review the critics specific to each of the indicators used in the compilation of both rankings. The critical analysis of the two world rankings is largely drawn from the very comprehensive study of Taylor and Braddock (2007).

2.1 SJTU university ranking – main features

The SJTU Academic Ranking of World Universities is published by the Shanghai Jiao Tong University since 2003.[3] In 2007, the SJTU was also published by broad subject in five categories including Natural Sciences and Mathematics; Engineering, Technology and Computer Sciences; Life and Agriculture Science; Social Sciences; Clinical Medicine and Pharmacy. The original purpose of the SJTU was to assess the gap between the top Chinese universities and 'world-class' universities, particularly in terms of academic or research performance. It has been produced independently without financial support from any sources outside the Institute of Higher Education (IHE).

Sample selection of universities

Each university with Nobel and Field laureates and staff, highly cited researchers or articles published in Nature and Science has been included in the

[3] See http://www.arwu.org/ranking.htm for additional information.

sample. Universities with significant amount of articles indexed in Sciences Citation Index-Expanded and Social Sciences Citation Index are also included. In total, more than 2000 institutions have been selected, 1000 have been ranked and the rank of the top 500 is published. The first 100 top universities are assigned a single rank. The remaining 400 universities are assigned a range of ranks, such as 102-150, 151-202, 203-304, 305-402, 403-500.[4]

Table 1. Shanghai Jiao Tong University Rankings (SJTU), 2007

Criteria	Indicator	Weight
Quality of Education	Alumni of an institution winning Nobel Prizes and Fields Medals	10%
Quality of Faculty	Staff of an institution winning Nobel Prizes and Fields Medals	20%
	Highly cited researchers in 21 broad subject categories	20%
Research Output	Articles published in Nature and Science	20%
	Articles in Science Citation Index-expanded, Social Science Citation Index	20%
Academic performance	Academic performance with respect to the size of an institution	10%

Criteria

The 2007 SJTU ranking is based on 4 criteria: (1) quality of education, (2) quality of faculty, (3) research output and (4) academic performance. Six indicators were selected (Table 1). The criterion on quality of education is described by a single indicator, the number of alumni of an institution having won Nobel Prizes and Fields Medals with different weights set according to when the alumni obtained degrees.[5] This indicator is assigned a 10% weight. The quality of university is captured by two indicators, (a) the number of staff of an institution that are Nobel or Field laureate and (b) the number of highly cited researchers in 21 broad subject categories.[6,7] Each

[4] A similar approach was followed in "Higher aspirations: An agenda for reforming European universities" by Philippe Aghion, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell and André Sapir Bruegel Blueprint Series N. 5, 2008.

[5] Alumni are individuals who obtained a bachelor, Master's or doctoral degrees from the institution. The weight is 100% for alumni obtaining degrees in 1991-2000, 90% for alumni obtaining degrees in 1981-1990, 80% for alumni obtaining degrees in 1971-1980, and so on, and finally 10% for alumni obtaining degrees in 1901-1910. For Nobel prizes, if a prize is shared by more than one person, weights are set for winners according to their proportion of the prize.

[6] Staff is defined as those who work at an institution at the time of winning the prize. The weight is 100% for winners in 2001-2006, 90% for winners in 1991-2000, 80% for winners in 1981-1990, 70% for winners in 1971-1980, and so on, and finally 10% for winners in 1911-1920.

[7] The 21 broad subject categories are Agricultural sciences, Biology & Biochemistry, Chemistry, Clinical medicine, Computer Science, Ecology/Environment, Economics/Business, Engineering, Geosciences, Immunology, Materials Science, Mathematics, Microbiology, Molecular Biology &

indicator is assigned a 20% weight. The research output is quantified by three indicators: the number of articles (a) published in Nature and Science, (b) Science Citation Index-expanded and (c) Social Science Citation Index.[8] Research output gets 40% of the total weight. Finally, the fourth criterion academic performance is expressed in terms of the size of the institution and is equal to the weighted scores of the above 5 indicators divided by the number of full-time equivalent academic staff. This indicator is assigned a weight of 10%.

For each indicator, the best performing institution is assigned a score of 100, and other institutions are calculated as a percentage of the top score. The final score is a weighted average of each indicator. The highest scoring institution is assigned a score of 100, and other institutions are calculated as a percentage of the top score.

2.2 THES university ranking

The THES World University Ranking is published by the Times Higher Education since 2004. It was produced for the interest of THES readers but, according to the publishers, the THES rankings have taken on a life of their own. The THES league tables were not intended to guide undergraduates in choosing which universities to apply to (this was more the role of national rankings). The THES ranking aims to look at the standing 'in the round' of universities of a particular type, i.e. those that bill themselves as international. Inevitably, according to the THES, this is largely about research because that is what these universities value and compare themselves on. But it also includes numbers of international students and staff, and teaching in so far as this can be indicated by the staffing ratios. The World University Rankings focus on reputation because that is, according to the publishers, the only way they could achieve an up-to-date picture compared with merely statistics. The published guide book states it has the '…single intention, to provide a measure of the world's top universities on as even a basis as possible'. The THES ranking also publishes a faculty ranking in the following areas: science; technology; biomedicine; arts and humanities; and social sciences.

Genetics, Neuroscience, Pharmacology, Physics, Plant and Animal Science, Psychology/Psychiatry, Social Science, Space Sciences. See http://www.isihighlycited.com for additional information.

[8] See http://www.isiknowledge.com. 100% of the weight is attributed to the corresponding author affiliation, 50% for first author affiliation, 25% for the next author affiliation, and 10% for other author affiliations. When calculating the total number of articles of an institution, a special weight of two was introduced for articles indexed in Social Science Citation Index.

Sample selection of universities

The top 500 universities in terms of research impact were first selected. Then single-faculty institutions or exclusively postgraduate institutions were removed from the sample.[9] The THES finally published the ranking of the first top 200 institutions.

Table 2. Times Higher Education Supplement Rankings (THES), 2007

Criteria	Indicator	Weight
Research Quality	Academic Opinion: Peer review, 5,101 academics	40%
	Citations per Faculty: Total citation/ Full Time Equivalent faculty	20%
Graduate Employability	Recruiter Review: Employers' opinion, 1,471 recruiters	10%
International Outlook	International Faculty: Percentage of international staff	5%
	International Students: Percentage of international students	5%
Teaching Quality	Student Faculty: Full Time Equivalent faculty/student ratio	20%

Criteria

In the THES World University Ranking, the opinion of scientists and international employers plays a crucial role implying that reputation is considered as a good proxy for university performance.[10] In 2007, the THES ranking is based on 4 criteria (that group 6 indicators): (1) research quality, (2) graduate employability, (3) international orientation and (4) teaching quality. The research quality is measured by (a) the opinion of a sample of academics that are asked to indicate the best universities and (b) the number of citations divided by Full Time Equivalent Faculty.[11] For the indicator based on academic opinion, 5101 respondents are asked to identify both their subject area of expertise and their regional knowledge.[12] They have then to select up to 30 institutions from their region(s) that they consider to be the best in the respective field of expertise. In 2007 the sample was composed of 41% of experts from Europe, Middle and East Africa, 29% from Asia-Pacific and 30% from America. The peer review indicator counts for 40% of the total weight. In

[9] It also implies that non-university higher education institutions are not taken into account.
[10] See http://www.thes.co.uk/worldrankings/ for additional information.
[11] The sample size of active academics increases every year. In 2007, this includes respondents from 2007, 2006 and 2005. Only the most recent response is taken into account if the respondent has filled in the questionnaire more than once.
[12] QS Quacquarelli Symonds, an independent consultancy agency is in charge of assembling a sample of field-specific experts. See http://www.topuniversities.com/worlduniversityrankings/university_rankings_news/article/2006_peer_review_response_analysis/ for additional information

addition, a 20% weight is assigned to the number of papers published and citations received by the research staff over the period 2002-2006.[13] The criterion on graduate employability is based on a single indicator that is also derived from a survey. This indicator is based on a sample of 738 of employers (from the manufacturing, services, finance, transport and public sector) who work internationally or on a large national scale and are asked from which universities they would prefer to recruit graduates.[14] In the 2007 sample, 43% of recruiters were from America, 25% from Asia-Pacific and 32% from Europe. This indicator is assigned a 10% weight. The international orientation of the institution is captured by two indicators. The first indicator is the percentage of overseas staff in the university and the second indicator measures the percentage of overseas students.[15] Each indicator receives a 5% weight. Finally, teaching quality is described by a single indicator, the ratio between the full time equivalent faculty staff and the number of students enrolled at the university. Teaching quality receives 20% of the weight.

In 2007, a z-score was calculated for each indicator by subtracting the mean score from the raw score and then dividing by the standard deviation. Then, once standardized, the scores for each indicator have been normalised against a score of 100 for the best performing institution The overall score is the weighted average of the six indicators.

2.3 Good or bad criteria: literature overview

Both the SJTU and THES rankings have been subject to several and severe critics that are summarized in Table 3 and Table 4. Some of those critics apply in the same way to both rankings while others are specific to a given ranking. Most studies agree that it is preferable to use objective indicators of research outputs rather than subjective measures such as peers' opinion.

The SJTU ranking

The majority of the critics against the SJTU ranking are twofold. First, and most important, only the research dimension of universities is taken into account. The SJTU team justifies the use of research-oriented indicators by the impossibility to compare teaching quality between universities at an international level. However, a

[13] Scopus is the citation data supplier in 2007.

[14] This year, around 375 recruiters responded to the questionnaire. Adding those 375 graduate employers to those interviewed last year leads to a sample of 738 responses. QS Quacquarelli Symonds is also in charge of the assembling a sample of recruiters

[15] The information comes from the institutions directly or from national central statistics agencies.

general interpretation is often given to the SJTU ranking, although the relationship between research performance and teaching quality is far from being well established. In addition, given that a great number of students who take up university studies may not necessarily follow an academic career or undertake a research-based job, the ranking is of little use to them. The SJTU ranking thus overlooks at the numerous other social and private benefits associated with university education. Moreover, even if one agrees on the use of research-based indicators, it is questionable whether it is appropriate to consider as adequate measures of current research some rare and potentially lagged achievements such as Nobel prizes, which are, in addition, only awarded in a limited number of fields. Finally, although the SJTU indicators express objective measures of research quality, they strongly downplay social sciences and humanities.

Second, five out of six indicators (which represent 90% of the total weight) are size-dependant indicators. This strongly favours - *ceteris paribus* - large institutions and does not give information on the real productivity of the staff of the institution. In 2007 for example, the University of Basel ranks 82nd in the overall rating and 27th with the academic performance indicator. On the contrary, the Johns Hopkins University is in the top 20 of the list but drops by more than 60 positions when the academic performance indicator is used instead. This simple example shows that the choice of size-dependant *versus* size adjusted indicators makes, for some institutions, an enormous difference in the final result. In response to this critic, league table compilers are thinking to increase the weight of the last indicator in the next editions of the ranking.

The THES ranking

Numerous authors are concerned with the use of expert-based indicators (50% of the total weight) due to the possibly misleading conclusions on the final university rating stemming from the continuous changes in methodology over the four editions of the THES ranking.

First, the use of expert-driven indicators on university performance and the lack of transparency surrounding the process of selection of experts involved in the review process cast serious doubts about the reliability of the overall THES ranking. Not being transparent enough regarding the methodology is subject to criticisms. However, it might also be a conscious or unconscious way of avoiding detailed comments that would render less credible the final ranking. Besides, peer review

indicators measure the reputation of a university rather than the "contemporaneous" research performance. As we show in the statistical part of this report in Section 3, either of the two expert-based indicators and the indicator on citations exhibit low degrees of correlation that are in most cases not statistically different from zero.

Second, as shown in Table 6, the yearly changes have been very substantial: (a) the sample sizes of the academic and recruiter polls have increased across the years, (b) the indicator on citations was based on the previous 10 years of citations in the first 2 editions, and to the previous 5 years of citations in the last two editions, (c) Scopus has replaced Thomson Scientific as data supplier for citations in the fourth edition. Finally while in the first three editions, each institution's score was calculated as a percentage of the best performing one, the indicators in 2007 were first standardized before being converted into a score between 0 and 100. Even if the THES team argues that those changes were necessary in order to improve the quality of the ranking, it is very difficult to disentangle time variations in the performance of universities from changes that are the result of a statistical artefact. For example, the Adelaide University rises in the overall ranking by 40 positions between 2006 and 2007. Is this shift the consequence of an improvement in the university's performance or is it the result of the statistical changes implemented in the 2007 THES ranking?

General comments

Both rankings rely highly on bibliometric indicators. As commented by Zitt and Filliatreau (2006) and Van Raan (2007), bibliometric-based indicators tend to be biased towards English-speaking and hard sciences intensive institutions. Indeed, non English journals are often not included in the Citation Indices and articles published in non English journals are less cited on average than those published in English. In addition, journal coverage by SCOPUS or Thomson-ISI is still not satisfactory for social and human sciences.[16] Publications in refereed journals are also far from being the only publication practice in soft sciences. Finally, citation habits of different scientific disciplines vary a lot between disciplines with a bias in favour of hard sciences.

The methods used to compile league tables are not always justifiable. Important aspects of institutional performance have proved impossible to capture with adequate statistical robustness (Yorke and Longden, 2005). Data require interpretation

[16] In average, SCOPUS covers a larger number of papers and journals than the Thomson-ISI Database. In addition, more sources in languages other than English and in humanities and social sciences are included in the SCOPUS database than in the first one.

and some conceptual framework, but league tables often combine performance indicators in an ad hoc way that may not even reflect the compilers' own concept of quality or excellent performance as stated in their publicity materials. The indicators selected and weightings applied are often not supported by an explicit rationale (Clarke, 2002).

The methodologies used to compile league tables might lead to misleading conclusions. Indeed, the difference in scores between institutions placed several positions apart may not be statistically significant, even though the difference in positions suggests a disparity in quality or performance. Alterations in methodology from year to year – in the data sources, indicators, procedures for calculating scores, weightings, ranking methods, etc – produce fluctuations in institutional positions that have nothing to do with changes in quality or performance but maybe the result of the fact that '[…] tables are not immune from cultural bias' (Yorke and Longden, 2005: 19; Brown, 2006).

Table 3. 2007 SJTU indicators: what has been said so far

INDICATORS	WEIGHT	PROBLEMS	QUALITIES
		Rough measures of teaching and research quality (I-A, B)	**Proxy for research and teaching quality (IA, B, II)**
I - Nobel Prizes (P) & Field (M) Medals			
A - won by alumni	*10%*	1 – Many U have no N & F laureates: no distinction for those U	1 - Reward research quality and not only research quantity
B - won by faculty members	*20%*	2 - Attributing N & F laureates to teaching quality is not straightforward because of a self-selection bias 3 - N and F prizes long time ago: not representative of the current performance 4 - Affiliation at the time of the prize is problematic if prize winning work was done before in another U (I, B)	2 - Proxy for university ability to attract outstanding researchers 3 - Quality research output: exclude researchers with "soft" academic publications (in particular **II**)
		Hard Science bias (I A, B, II, III) N & F prizes in a limited number of fields	
II - Number of highly cited researchers	*20%*	Only 2 out of the 21 disciplines belong to social sciences	
III - Number of papers published by staff			
A - in Science and Nature	*20%*	It covers only hard sciences	
B - in wide academic journals	*20%*		4 - Focus more on research quantity (**III B**) 5 - Rewards more articles indexed in the social sciences/arts and humanities indexes to compensate for the hard science bias
IV - Academic performance/size U	*10%*	**Scant weight (IV)**	**Adjustment of size bias (IV)**

Note: this overview is largely drawn from Taylor and Braddock (2007).

Table 4. 2007 THES indicators: what has been said so far

INDICATORS	WEIGHT	PROBLEMS	QUALITIES
I - Peer review	40%	**Survey: Regional bias and lack of transparency (I and II)** 1 - Assessors were asked to assess the relative performance of institutions in their own geographical area 2 - Not clear what questions were asked and who was surveyed	
II - Employer review	*10%*	**Reputation indicator (I)** 1 - depends on past performance	
		Selection bias (II) Outstanding universities are initially well connected and recruit good students: nothing to do with U excellence produced within the U	
III - Citation per capita	*20%*	**Quantity without taking into account quality (III)** 1 - Measure research quantity without awarding in a special way high research quality 2 - Only 20% of total score while research on the main component of U excellence 3 - Bibliometric indicators: biased toward English publication journals and downplay the weight of social science and humanities	**Per capita: no size bias (III)**
IV - Student/teaching ratio	*20%*	**Crude measure of teaching quality (IV)**	**Proxy for teaching quality (IV)** 1 - Give 20% of the total score to one important aspect of U activities 2 - Objectively measurable 3 - Very difficult to find other ways to measure teaching quality
V - International orientation		**Not real criterion of U excellence (V)**	**Proxy for university quality (V)** 1 - Capacity to attract foreign staff and students 2 - International education
A - % of overseas students	*5%*	1 - Correlated with the characteristics' of the university's city population (multicultural city) 2 - Tell more about the quality of recruitment's methods (good advertisement) than about the university excellence	
B - % of overseas staff	*5%*		

Note: this overview is largely drawn from Taylor and Braddock (2007).

Table 5. THES Ranking: Methodological changes over the four editions

2004	2005	2006	2007
Peer review score (50%) Sample made of 1300 academics	**Peer review score (40%)** Sample has changed: 2375 academics (include data from the 2004 survey) Weight: this indicator counts now for 40% of the final score	**Peer review score (40%)** Sample has changed: 3,703 academics (data from 2004 and 2005 added to the new responses in 2006).	**Peer review score (40%)** Sample has changed 5,101 academics (this includes respondents from 2005 and 2006). Experts cannot choose their own university.
	New indicator **Recruiter review (10%)** Sample size : 333 recruiters	**Recruiter review (10%)** Sample has changed: 738 recruiter (include data from 2005).	**Recruiter review (10%)** Sample has changed: 1,471 recruiters (this includes respondents from 2005 and 2006).
International Faculty Score (5%) Percentage of overseas staff	**International faculty score (5%)**	**International faculty score (5%)**	**International faculty score (5%)**
International Students Score (5%) Percentage of overseas students	**International students score 5%)**	**International students score (5%)**	**International students score (5%)**
Faculty/student score (20%) Number of full time equivalent faculty on the number of student enrolled at the university	**Faculty/student score (20%)**	**Faculty/student score (20%)**	**Faculty/student score (20%)**
Citations/faculty score (20%) Citations over the period 1994- 2003. Data supplier: data from Thomson's Essential Science Indicators database.	**Citations/faculty score (20%)** Citations over the period 1995- 2004	**Citations/faculty score (20%)** The analysis uses data covering 5 years (2001-05) rather than 10 years as in the first two editions.	**Citations/faculty score (20%)** Scopus has replaced Thomson Scientific as supplier of citations data. Citations over the period 2002-2006

Note: Changes from year to year are marked in grey

Table 6. SJTU Ranking: Methodological changes over the five editions

2003	2004	2005	2006	2007
Quality of Education:	**Quality of Education**: New indicator: number of the alumni of an institution winning Nobel prizes and Fields Medals (weight: 10%)	**Quality of Education**: alumni of an institution winning Nobel Prizes and Fields Medals (10%)	**Quality of Education**: alumni of an institution winning Nobel Prizes and Fields Medals (10%)	**Quality of Education**: alumni of an institution winning Nobel Prizes and Fields Medals (10%)
Quality of faculty: - Nobel laureates in physics, chemistry, medicine and economics (20%)	**Quality of faculty:** New definition: - Staff of an institution winning Nobel Prizes and Fields Medals (20%);	**Quality of faculty:** - Staff of an institution winning Nobel Prizes and Fields Medals (20%);	**Quality of faculty:** - Staff of an institution winning Nobel Prizes and Fields Medals (20%);	**Quality of faculty:** - Staff of an institution winning Nobel Prizes and Fields Medals (20%);
- Highly cited researchers in 21 broad subject categories over the period 1981-1999 (20%)	- Highly cited researchers in 21 broad subject categories over the period 1981-1999 (20%).	- Highly cited researchers in 21 broad subject categories over the period 1981-1999 (20%).	- Highly cited researchers in 21 broad subject categories over the period 1981-1999 (20%).	- Highly cited researchers in 21 broad subject categories over the period 1981-1999 (20%).
Research output: - Number of articles published in Nature and Science. 2000-2002 (20%); - Number of articles indexed in Science Citation Index-expanded and Social Science Citation Index in 2002 (20%).	**Research output**: - Number of articles published in Nature and Science. 1999-2003 (4 years periods) (20%); - Number of articles indexed in Science Citation Index-expanded and Social Science Citation Index in 2003 (20%).	**Research output**: - Number of articles published in Nature and Science, 2000-2004 (20%); - Number of articles indexed in Science Citation Index-expanded, Social Science Citation Index and Arts & Humanities Citation Index Arts & Humanities Citation Index in 2004 (20%).	**Research Output:** - Number of articles published in Nature and Science, 2001-2005 (20%); - Number of articles indexed in Science Citation Index-expanded and Social Science Citation Index in 2005 (20%) Arts & Humanities Citation Index Arts & Humanities Citation Index removed	**Research Output:** - Number of articles published in Nature and Science, 2002-2006 (20%); - Number of articles indexed in Science Citation Index-expanded and Social Science Citation Index in 2006 (20%)
Size of institution: Academic performance per faculty (20%).	**Size of institution**: Academic performance with respect to the size of an institution (new weight: 10%).	**Size of institution**: Academic performance with respect to the size of an institution (10%).	**Size of institution:** Academic performance with respect to the size of an institution (10%).	**Size of institution:** Academic performance with respect to the size of an institution (10%).

Note: Changes from year to year are marked in grey

2.4 Consistency of the two worldwide rankings

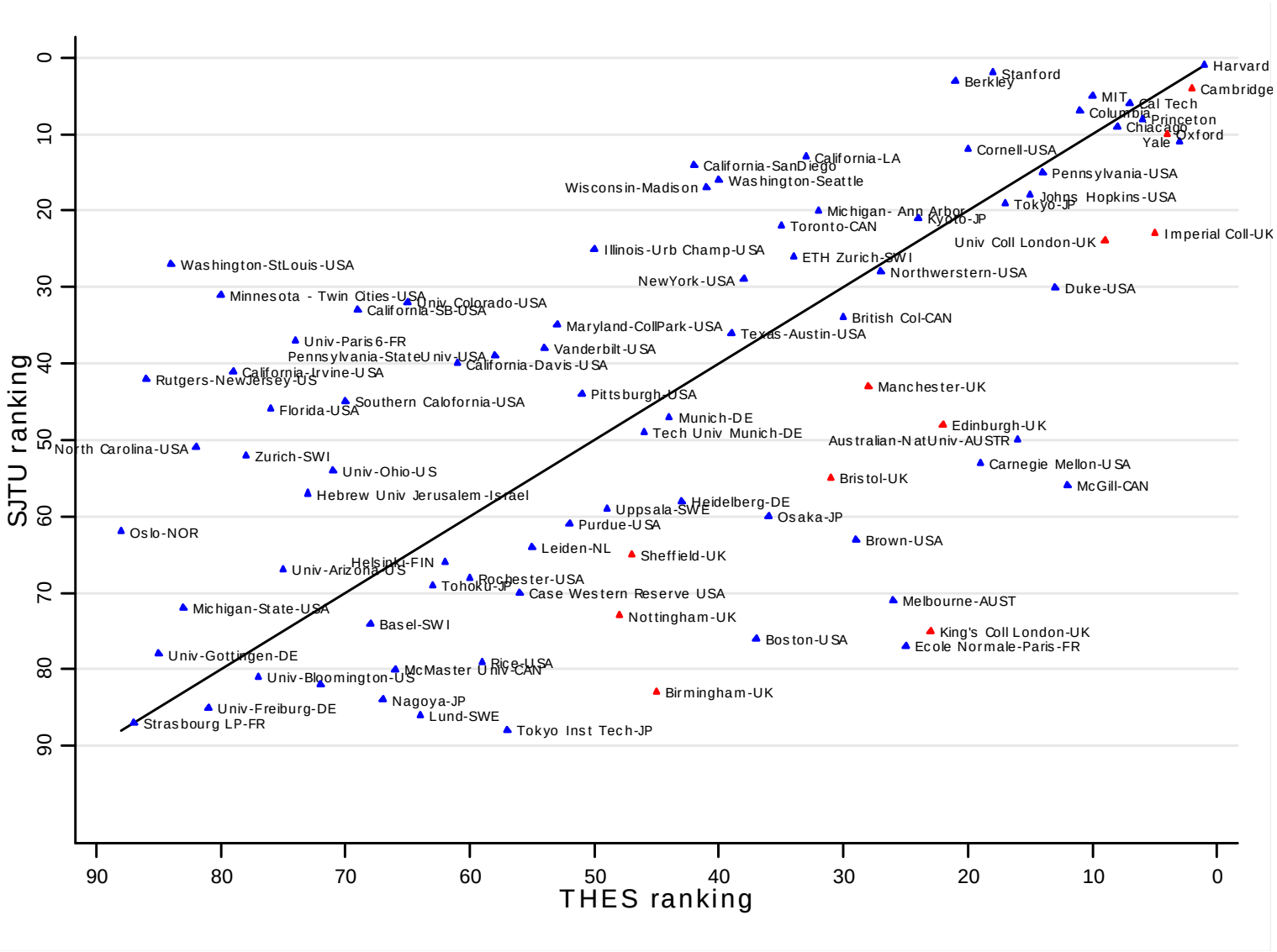
A common perception is that some universities will almost always appear at the top of any league table. Some commentators have suggested that 'different ranking systems provide consistent results for some institutions and inconsistent ones for others' (Usher and Savino, 2006: 32). So, for example, there may be consistency at the top and bottom of particular league tables but volatility for those institutions in the middle.

Figure 1 plots the THES ranking versus the SJTU ranking for the 88 universities which are common to the two rankings[17]. The THES and SJTU rankings are consistent for six of the Top 10 universities: Harvard, Cambridge, Princeton, California Inst Tech, MIT and Columbia are Top 10 in both rankings. The two rankings identify similar world-class universities, despite the diversity of indicators used. However, we observe much greater variation in the middle to lower end. There is indeed a fairly large number of universities, such as Mc-Gill and Melbourne that are in the Top 30 in the THES ranking but are situated after the 50th position in the SJTU ranking. Similarly, California-Santa Barbara or the University of Paris 6 are in the Top 40 in the SJTU ranking and respectively 69th and 74th in the THES ranking. What is also clearly apparent from Figure 1 is that UK universities tend, on average, to perform much better with the THES ranking than with the SJTU ranking. For instance, King's college of London and Bristol university respectively rank 75th and 55th in the SJTU ranking and 23rd and 31st with the THES ranking.

It is useful to discuss the drivers of such differences: Mc-Gill and Melbourne perform particularly well with the two expert-based indicators while all bibliometric indicators suggest that their research performance is relatively low. Given that peer reviewers are asked to rank universities in their region and not across the world, this result may point to a strong regional bias. For instance, Melbourne only competes with universities in Asia Pacific. We also observe that the two universities are classified as Top in terms of the proportion of international staff and students. This might be due to the fact that those two universities are very good at attracting foreign students or that Montreal and Melbourne are two very international cities. The honourable position of University of Paris 6 is likely due to the fact that this university is specialized in Sciences and Medicine and benefits from a comparative advantage with the bibliometric indicators relatively to other institutions more oriented towards social sciences.

[17] The analysis is based on the "common THES rank" (or simply THES rank) and the "common SJTU rank" (or simply SJTU rank). These ranks range from 1 to 88 for the universities we are studying. For example, the University of Strasbourg in France is originally ranked 99th in the SJTU, but it is ranked 87th in our classification of the 88 universities. Similarly, the University of Strasbourg is ranked 184th in the original THES but the common THES rank is 87th.

Figure 1. SJTU ranking versus THES ranking (common universities)



3. Main drivers of the two international university rankings

In this section, we perform correlation analysis and factor analysis to examine the relationship between the indicators in each framework and across the two frameworks. We study the scores and ranks of the 88 universities that are ranked both in the SJTU and in THES. The degree of association between the SJTU indicators is shown to be much higher than among the THES indicators. Additionally, the expert-based indicators on university quality are not systematically related to the more objective indicators on bibliometric measures. The analysis of the relative contribution of each indicator to the overall university score (which is equal to the proportion of the product of the indicator score with the associated weight compared to the total score) also shows that, on average, the relative contribution of an indicator to the overall score is not necessarily captured by the weight attached to the indicator. Furthermore, the THES scores are indeed dominated by three indicators: the academic review, the citations per faculty and the teacher to student ratio. The SJTU scores are also dominated by three indicators (articles in Science and Social Citation Indices, number of highly cited researchers, articles published in Nature and Science), contrarily to the more balanced structure suggested by the weights.

3.1 Correlation analysis

Correlation analysis is performed to examine the relationship between the indicators in each framework and across the two frameworks. Correlation analysis is a basic but widely used tool for "confirming" the mathematical design of indices. Booysen (2002) recommends that a weak correlation between an underlying indicator and an index should result in the exclusion of the respective indicator. A major drawback of correlation analysis though is the fact that strong correlation does not necessarily imply strong influence or representation of the indicator in the overall index. In other words, any random variable could potentially show strong correlation with the index without actually being part of the index. Yet, the higher the number of cases analysed the lower the probability that spurious correlations occur.

Associations between the components of the THES ranking

A simple correlation between the THES scores and either of the six indicators scores reveals relatively strong associations in particular with two of the six indicators (Table 7). The THES score has a relatively high correlation with the *academics review* score ($r = 0.81$) and the *recruiter review* score ($r = 0.71$) and a fair relationship with the *percentage of international staff* ($r = 0.60$) and the *teacher to student ratio* ($r = 0.55$). Low is the association of the THES scores to the *citations per faculty* and the *percentage of international students* ($r < 0.40$). Thus, the THES scores are more associated to the two peer review indicators, in part due to the 50% weighting attached to the two indicators altogether. The built in bias, which is acknowledged by the THES publishers, is that old, big universities are more recognisable. As a consequence, these universities receive a higher *academic review* score. Similarly, recruiters' responses, that come largely from human resources departments, are very predictable because they generally want to hire graduates from a narrow selection of universities.

Relationships among the six indicators vary but are generally low, which implies that there is limited overlap in what is being measured. The most pronounced association is between the two expert-derived indicators, namely the *academic review* score and the *recruiter review* score ($r = 0.57$). This positive and fair correlation can be interpreted in different ways. First, research quality may be strongly associated to student performance in the labour market. On one side, this could be the result of a sample selection bias, namely students from prestigious universities are also endowed with high level of individual ability and/or have benefited from a good pre-university education. On the other hand, teaching and research quality might be intrinsically related. Second, as it has been claimed several times, the two review-based indicators are both measuring the reputation of the university. However, reputation and current university excellence are not necessarily related to each other. Indeed, the correlation between the *citations per faculty* and the *recruiter review* score is not significantly different from zero. Also, the association of the *citations per faculty* and the *academic review* score is relatively low ($r = 0.31$). Many of the associations between the THES indicators are not significant. This result further confirms that the six THES indicators account for different aspects of universities features. At this point, we emphasise that the correlations we study are carried out across the 88 universities included in the Top 100 THES list and are also ranked by the SJTU (Top 200 list). However, the correlation analysis carried out by CHERI/HEFCE (2008) on the entire set of universities confirms these conclusions.

Associations between the components of the SJTU ranking

The associations between the SJTU scores and the six indicators scores are stronger than in the THES dataset (Table 7). The SJTU scores have an almost perfect association to the *number of articles published in Nature and Science* ($r = 0.94$). Also very high is the association between the SJTU scores and the *Alumni* or *Staff winning Nobel prizes and field medals,* the *number of highly cited researchers,* and the overall *academic performance with respect to the size of an institution* ($r \geq 0.79$). The correlation between the SJTU scores and the *articles in Science Citation Index expanded and the Social Science Citation Index* is slightly weaker but still significant ($r = 0.66$).

Relationships among the six SJTU indicators vary and are generally higher than in the THES case, implying that the amount of overlap is higher in this case. The most pronounced association is between the *number of highly cited researchers* and the *number of articles published in Nature and Science* ($r = 0.88$). Also strong are the associations including the *Alumni* or *Staff winning Nobel prizes and field medals*, the *number of articles published in Nature and Science* and the *academic performance with respect to the size of an institution.* This is not surprising as these indicators are all capturing research quality, although they are biased towards hard sciences. All remaining associations are positive and significant. This result further confirms that there is significant overlap in the amount of information provided by the six SJTU indicators. Again, the associations are studied only for the 88 common universities and may not necessarily reflect the situation in the entire SJTU dataset, but the conclusions of the study carried out by CHERI/HEFCE (2008) that were based on the entire set of universities point in this direction.

Associations between the components of the THES and the SJTU indicators

The linkages between the indicators in the THES on the one side and the SJTU indicators on the other side have several interesting points to reveal (Table 7). In most cases the associations are positive and significant, but without exceeding an $r = 0.58$ value. These modest associations are found between the *academic review indicator* of the THES and either the *number of highly cited researchers* ($r = 0.55$) or the *number of articles published in Nature and Science* ($r = 0.58$) of the SJTU. This result makes evident that there is a relatively low degree of information overlap between the two frameworks. Therefore, an eventual merging of the two dataset may provide a more holistic picture of the universities performance.

The *academic review* score in the THES has a modest association to the overall SJTU score ($r = 0.60$). Even lower is the association between the *recruiter review* score of the THES and any of the SJTU indicators or the overall SJTU score ($r \leq 0.39$). It is therefore likely that recruiters opinion on which graduates would rather recruit is not really associated with qualities related to winning Nobel Prizes and Fields Medals, or citations and research outcomes. The features that the recruiters had in mind when assigning their recruiter score were therefore different than what the twelve indicators in the THES and SJTU are capturing.

The low association between the *number of citations per faculty* in the THES and either the *number of highly cited researchers* or the *articles published in Nature and Science* or the a*rticles in the Science and Social Citation Indices* in the SJTU ($r \leq 0.39$) raises an issue that although all four indicators are related to publications, thus one would have expected a higher association between them, it is evident that there is a different type of information provided by those four bibliometric indicators.

A few associations appear to be random, in particular those involving some of the THES indicators on one side, such as the *teacher to student ratio*, or the *percentage of international students* or *percentage of international staff*, and some of the SJTU on the other side, such as *highly cited researchers*, the *number articles published in Nature and Science* and the *academic performance with respect to the size of an institution*. In other words, the two THES criteria related to teaching quality and the international outlook bear no clear association with the bibliometric indicators in the SJTU. This might point that the teacher to student ratio is a too crude measure of teaching quality with, in addition, great variations between disciplines or that research and teaching performances are not significantly associated altogether. However, recall that since we are studying the performance of roughly 0.5% of world's universities that were ranked Top100 in SJTU (or equivalently in the Top200 list of the THES), research and teaching quality may not necessarily be associated. However, if the analysis had included a much broader number of universities, this association would need to be stronger.

Table 7. Pearson correlation coefficients between the twelve indicators in the THES& SJTU frameworks

		THES framework							SJTU framework					
		Academic review	Recruiter review	Teacher to student ratio	Citations per faculty	International students	International staff	THES Score	Alumni winning Nobel prizes	Staff winning Nobel prizes	Highly cited researchers	Articles in Nature & Science	Articles in Science & Social CI	Academic performance - size
THES framework	Academic review	1.00												
	Recruiter review	0.57	1.00											
	Teacher to student ratio	0.09*	0.19*	1.00										
	Citations per faculty	0.31	0.16*	-0.16*	1.00									
	International students	0.26	0.25	0.09*	0.02*	1.00								
	International staff	0.33	0.42	0.44	-0.03*	0.44	1.00							
	THES Score	0.81	0.71	0.55	0.34	0.38	0.60	1.00						
SJTU framework	Alumni winning Nobel prizes	0.46	0.29	0.39	0.24	0.28	0.48	0.61	1.00					
	Staff winning Nobel prizes	0.48	0.25	0.28	0.26	0.29	0.44	0.56	0.78	1.00				
	Highly cited researchers	0.55	0.39	0.12*	0.39	0.02*	0.10*	0.52	0.55	0.61	1.00			
	Articles in Nature & Science	0.58	0.34	0.26	0.33	0.08*	0.21	0.59	0.66	0.71	0.88	1.00		
	Articles in Science & Social CI	0.46	0.39	0.12*	0.28	-0.08*	-0.06*	0.43	0.36	0.26	0.71	0.69	1.00	
	Academic performance - size	0.45	0.24	0.39	0.25	0.30	0.42	0.59	0.65	0.76	0.65	0.76	0.34	1.00
	SJTU Score	0.60	0.38	0.29	0.35	0.17*	0.31	0.65	0.79	0.85	0.89	0.94	0.66	0.81

All coefficients are significant ($p < 0.01$, $n = 88$); * coefficient non significant ($p >> 0.05$).

3.2 Statistical dimensionality of the datasets

The conclusion drawn before on the stronger associations between the SJTU indicators than between the THES indicators are further confronted to the statistical dimensionality of each dataset (Table 8). There are two principal factors in either framework that explain at least as much information as a single indicator alone ($eigenvalue \geq 1.0$). In the THES framework, where there is less information overlap than in the SJTU, the first two factors account for 60.8% of the variation. In the SJTU framework, the first two factors account for 86.1% of the variation.

The set of twelve indicators can be described adequately by four principal factors with eigenvalues greater than unity (78.6% of the total variation). Upon factor rotation, so as to enhance the interpretability of the results, the indicators are grouped quite intuitively in particular as far as the first two factors are concerned. The first Factor is represented by two THES indicators: the *international staff* and the *international students*; the second Factor is described by five SJTU indicators (except for the *articles in Science and Social Citation Index*); the third Factor is captured by the two *peer review* indicators (academic and recruiter) of the THES and the *articles in Science and Social Citation Index* of the SJTU; finally, the fourth Factor is described by the *teacher to student ratio* and *citations per faculty* of the THES (Table **9**). In brief we could argue that the first Factor clearly captures mobility issues, the second factor clearly describes research outcomes, the third Factor is a combination of subjective indicators and research (not directly linked to the research expressed by Factor 2), and finally the forth Factor contains the two remaining THES indicators (*teacher to student ratio* and *citations per faculty*) not because of the similarities of the two indicators (recall that the correlation between the two indicators was found to be non significant) but because of the dissimilarity they bear to the other three factors. Another remark is that three of the principal factors are described exclusively by either the THES or SJTU indicators, and only one factor is based on indicators from both frameworks. Given the diversity of the indicators and the fact that the number of statistical dimensions in the combined THES&SJTU framework is twice the number of statistical dimensions for either the THES or the SJTU, more diverse aspects of universities are captured if all twelve indicators are considered. This conclusion was already drawn previously based on the correlations, but it is further confirmed here.

Table 8. Eigenvalues of the THES, the SJTU or the hybrid framework

	Framework THES		Framework SJTU		Hybrid framework THES&SJTU	
	Eigenvalue	Cumulative variance explained (%)	Eigenvalue	Cumulative variance explained (%)	Eigenvalue	Cumulative variance explained (%)
1	2.30	38.3	4.16	69.4	5.30	44.2
2	1.35	60.8	1.00	86.1	1.90	60.0
3	0.87	75.3	0.39	92.6	1.19	69.9
4	0.68	86.6	0.22	96.2	1.04	78.6
5	0.42	93.6	0.14	98.5	0.66	84.1
6	0.38	100.0	0.09	100.0	0.51	88.4
7					0.38	91.5
8					0.36	94.5
9					0.28	96.9
10					0.16	98.2
11					0.14	99.3
12					0.08	100.0

Table 9. Squared factor loadings of the twelve indicators in the four main factors

	Indicator	Factor 1	Factor 2	Factor 3	Factor 4
THES	Academic review	0.07	0.14	0.48	0.03
	Recruiter review	0.09	0.00	0.75	0.01
	Teacher to student ratio	0.01	0.13	0.01	0.63
	Citations per faculty	0.00	0.13	0.03	0.45
	International students	0.66	0.01	0.01	0.01
	International staff	0.50	0.07	0.04	0.16
SJTU	Alumni winning Nobel prizes and field medals	0.08	0.64	0.03	0.02
	Staff winning Nobel prizes and field medals	0.09	0.75	0.01	0.00
	Highly cited researchers	0.04	0.55	0.25	0.03
	Articles published in Nature and Science	0.02	0.72	0.18	0.00
	Articles in Science and Social Citation Index	0.17	0.18	0.46	0.00
	Academic performance with respect to size	0.04	0.74	0.02	0.01
	Eigenvalue	**1.78**	**4.06**	**2.25**	**1.34**

3.3 Relative contribution of indicators to the overall scores

The analysis of the relative contribution of the indicators to each university score can provide useful information as to whether some indicators dominate the overall scores. The relative contribution is estimated as the proportion of an indicator score multiplied by the respective weight with respect to the overall university score. The minimum, average and maximum value of those shares (across the 88 universities) are presented in Table 10.

The results show that, on average, the shares of the THES indicators in the overall score are only in few cases consistent with the corresponding weights and that the THES scores appear to be dominated by three indicators: the *academic review*, the *citations per faculty* and the *teacher to student ratio*. The *academic review* was assigned a 40% weight, yet the relative average share of this indicator in the overall score is slightly higher (43.5%). For some universities the contribution of the *academic review* to the overall score can be as high as 54.7% (University of California-Santa Barbara) to as low as 28.3% (Michigan State University). Also higher than the respective weight is the contribution of the *citations per faculty* to the overall THES score (weight:20%, share:22.1%). For the remaining four indicators, the average contribution is lower than the corresponding weight.

The SJTU framework has similar results to reveal: the average contribution of the indicators to the overall scores are only in few cases consistent with the corresponding weights. Also, the overall scores are dominated by three indicators, contrarily to the more balanced structure suggested by the weights. In fact, the *articles in Science and Social Citation Indices,* the *number of highly cited researchers* and the *articles published in Nature and Science* each contributes by more than 18%. Interestingly, the *staff winning Nobel prizes and field medals* has an average contribution of 12.8%, which is much lower than the 20% weight attached to that indicator. Only in the case of *articles in the Science and Social Citation Indices* is the average contribution higher than the corresponding weight. For the remaining five indicators, the average contribution is close to or lower than the actual weight. Some extreme cases are observed for the *articles in Science and Social Citation Index* that was assigned a 20% weight: the contribution of this indicator to the overall score can be as high as 47.8% (Tohoku University in Japan) to as low as 15.6% (California Institute of Technology).

Table 10. Analysis of shares for the indicators in THES and SJTU

	Indicators	Weight	Share average	Share Min	Share Max
THES	Academic review	40%	43.5%	28.3%	54.7%
	Recruiter review	10%	9.2%	0.0%	14.4%
	Teacher to student ratio	20%	17.8%	4.3%	33.9%
	Citations per faculty	20%	22.1%	0.3%	31.1%
	International students	5%	3.5%	0.9%	8.1%
	International staff	5%	4.0%	1.1%	7.3%
SJTU	Alumni winning Nobel prizes and field medals	10%	7.5%	0.0%	19.5%
	Staff winning Nobel prizes and field medals	20%	12.8%	0.0%	28.2%
	Highly cited researchers	20%	20.2%	9.1%	29.5%
	Articles published in Nature and Science	20%	18.1%	8.2%	27.7%
	Articles in Science and Social Citation Index	20%	32.1%	15.6%	47.8%
	Academic performance with respect to size	10%	9.4%	5.5%	15.5%

These results show than an analysis of the shares is of added value, given than the relative contribution of an indicator to the overall score is not necessarily reflected in the weight attached to the indicator. In fact, the THES and SJTU frameworks have shown that this is not to be taken for granted.

4. Uncertainty and Sensitivity Analysis

In this section we carry out a thorough uncertainty and sensitivity analysis of the 2007 SJTU and THES rankings under a plurality of scenarios in which we activate simultaneously different sources of uncertainty. The sources cover a wide and versatile spectrum of methodological assumptions (all with their advantages and implications). Subsequently, a frequency matrix of the university ranks is calculated across the different simulations (triggering the exclusion of an indicator, the weighting and the aggregation rule). Such a multi-modeling approach allows one to deal with the criticism, often made to league tables and rankings systems, that ranks are presented as if they were calculated under conditions of certainty while this is rarely the case. Thus we deviate from the classic approach – also taken in the two university ranking systems - to build a composite indicator by a simple weighted summation of indicators.

For the majority of the universities we analysed, the THES or SJTU rank have proven impossible to capture with adequate statistical robustness. In fact, the difference in scores between institutions placed several positions apart are not statistically significant even though the difference in positions suggests a disparity in quality or performance. Even when combining all twelve indicators in a single framework, the space of the inference is too wide (no meaningful rank can be estimated) for about 50 universities (of the 88 studied). This outcome calls in turn for a revision of the current set of indicators, or a further collection of indicators. In addition, the THES rank was found to be biased in the case of 26 universities, whilst the SJTU was found to be based in the case of 33 universities.

4.1 Considerations on the robustness assessment

The selection of an appropriate methodology is central to any exercise attempting to capture and summarize the interactions among the individual indicators included in a composite indicator or ranking system (Saisana and Tarantola, 2002). The literature review offered in the JRC/OECD (2008) Handbook on composite indicators discusses the plurality of the approaches that have been used in building a composite indicator and shows that some of the methodologies are suited (more or less) to the purposes for which they are employed. In particular, the authors stress the need for an explicit conceptual framework for the index, and the usefulness of multivariate analysis prior to the aggregation of the individual indicators. They review tools for imputation of missing information, methodologies for weighting and aggregation, and finally methods for assessing the robustness of the index using uncertainty and sensitivity analysis. In Table 11 we present a stylised 'checklist' to be followed in the construction of a composite indicator, which we have rearranged from the information contained in the Handbook. In brief, the ten main steps to be followed are:

1. Conceptual or theoretical framework
2. Data selection
3. Imputation of missing data
4. Multivariate analysis
5. Normalisation
6. Weighting and aggregation
7. Uncertainty and sensitivity analysis
8. Back to the underlying indicators data
9. Links to other indicators
10. Visualisation of the results

Table 11. Checklist for building a composite indicator

Step	Why it is needed
1. Theoretical framework Provides the basis for the selection and combination of variables into a meaningful composite indicator under a fitness-for-purpose principle (involvement of experts and stakeholders is envisaged at this step).	• To get a clear understanding and definition of the multidimensional phenomenon to be measured. • To structure the various sub-groups of the phenomenon (if needed). • To compile a list of selection criteria for the underlying variables, e.g., input, output, process.
2. Data selection Should be based on the analytical soundness, measurability, country coverage, and relevance of the indicators to the phenomenon being measured and relationship to each other. The use of proxy variables should be considered when data are scarce (involvement of experts and stakeholders is envisaged at this step).	• To check the quality of the available indicators. • To discuss the strengths and weaknesses of each selected indicator. • To create a summary table on data characteristics, e.g., availability (across country, time), source, type (hard, soft or input, output, process).
3. Imputation of missing data is needed in order to provide a complete dataset (e.g. by means of single or multiple imputation).	• To estimate missing values. • To provide a measure of the reliability of each imputed value, so as to assess the impact of the imputation on the composite indicator results. • To discuss the presence of outliers in the dataset.
4. Multivariate analysis should be used to study the overall structure of the dataset, assess its suitability, and guide subsequent methodological choices (e.g., weighting, aggregation).	• To check the underlying structure of the data along the two main dimensions, namely individual indicators and countries (by means of suitable multivariate methods, e.g., principal components analysis, cluster analysis). • To identify groups of indicators or groups of countries that are statistically "similar" and provide an interpretation of the results. • To compare the statistically-determined structure of the data set to the theoretical framework and discuss possible differences.
5. Normalisation should be carried out to render the variables comparable.	• To select suitable normalisation procedure(s) that respect both the theoretical framework and the data properties. • To discuss the presence of outliers in the dataset as they may become unintended benchmarks. • To make scale adjustments, if necessary. • To transform highly skewed indicators, if necessary.
6. Weighting and aggregation should be done along the lines of the underlying theoretical framework.	• To select appropriate weighting and aggregation procedure(s) that respect both the theoretical framework and the data properties. • To discuss whether correlation issues among indicators should be accounted for. • To discuss whether compensability among indicators should be allowed.
7. Uncertainty and sensitivity analysis should be undertaken to assess the robustness of the composite indicator in terms of e.g., the mechanism for including or excluding an indicator, the normalisation scheme, the imputation of missing data, the choice of weights, the aggregation method.	• To consider a multi-modelling approach to build the composite indicator, and if available, alternative conceptual scenarios for the selection of the underlying indicators. • To identify all possible sources of uncertainty in the development of the composite indicator and accompany the composite scores and ranks with uncertainty bounds. • To conduct sensitivity analysis of the inference (assumptions) and determine what sources of uncertainty are more influential in the scores and/or ranks.
8. Back to the data is needed to reveal the main drivers for an overall good or bad performance. Transparency is primordial to good analysis and policymaking.	• To profile country performance at the indicator level so as to reveal what is driving the composite indicator results. • To check for correlation and causality (if possible). • to identify if the composite indicator results are overly dominated by few indicators and to explain the relative importance of the sub-components of the composite indicator.
9. Links to other indicators should be made to correlate the composite indicator (or its dimensions) with existing (simple or composite) indicators as well as to identify linkages through regressions.	• To correlate the composite indicator with other relevant measures, taking into consideration the results of sensitivity analysis. • To develop data-driven narratives based on the results.
10. **Visualisation of the results** should receive proper attention, given that the visualisation can influence (or help to enhance) interpretability.	• To identify a coherent set of presentational tools for the targeted audience. • To select the visualisation technique which communicates the most information. • To present the composite indicator results in a clear and accurate manner.

Note: Source JRC/OECD (2008) Handbook on composite indicators

Several practitioners have noted that the encoding process of building a composite indicator or a ranking system is fraught with uncertainties of different order (Saisana et al., 2005). As a result, an uncertainty analysis should naturally include a careful mapping of all these uncertainties (/assumptions) onto the space of the output (/inferences). When this is done, two things can happen:

- The space of the inference is still narrow enough as to be meaningful (the possible rank range is relatively narrow)
- The space of the inference is too wide (no meaningful rank can be estimated for the universities).

The latter outcome calls in turn for a revision of the ranking system, or a further collection of indicators.

To this end, we developed several scenarios that are described in detail below. The results of this scenario analysis were then used to build frequency matrices of the university ranks and account for the uncertainty underlying the construction of the ranking systems in the most comprehensive way possible. This multi-modeling approach we employed, has already proven to be useful in the development and validation of several composite indicators listed in Table 12 (e.g., Composite Learning Index, Environmental Performance Index, Knowledge Economy Index, Alcohol Policy Index) and was also included in the JRC/OECD (2008) Handbook on Composite Indicators.

Table 12. Examples of composite indicators that were assessed using a multi-modelling approach

Source	Composite Indicator	Brief description
Saisana (2008)	Composite Learning Index (developers: Canadian Council on Learning, Ottawa, Canada) http://www.ccl-cca.ca/CCL/Reports/CLI/	The Index aims at providing an aggregate measure that puts lifelong learning on a map by measuring how well Canadians are doing across the full spectrum of learning (school, home, workplace, community). The composite learning index is the first national learning index in the world.
Saisana & Munda (2008)	Knowledge Economy Index (developers: FP6 KEI project) http://kei.publicstatistics.net/	Development of a robust composite indicator that captures the multi-dimensional nature of knowledge economy in Europe.
Saisana & Saltelli (2008; 2006)	Environmental Performance Index (developers: Yale university and Columbia university) http://epi.yale.edu/Home	The Environmental Performance Index aimed at providing benchmarks for current national pollution control and natural resource management results in more than 130 countries by identifying specific targets and measuring how close each country comes to these established goals.
Brand et al. (2007)	Alcohol Policy Index (developers: New York Medical College, Valhalla)	The index was developed with a view to assist public health leaders and policymakers to gauge the strength of a country's alcohol control policies.
Hoskins et al. (2006)	Active Citizenship Index (developers: Centre for Research on Lifelong Learning, JRC) http://crell.jrc.ec.europa.eu/	The Active Citizenship Index attempts to put the spotlight on the Lisbon Strategy and measure values, representative democracy and civil society using as basis the 2002 European Social Survey data.
Cherchye et al. (2008) Saisana et al. (2005)	Technology Achievement Index (developers: United Nations)	The study aimed at discussing how the combined use of data envelopment analysis and uncertainty & sensitivity analysis can provide useful tools in the construction of composite indicators using the Technology Achievement Index as an illustration.
Munda & Saisana (2008)	Sustainable Development in Spanish Regions (funded by: Autonomous University of Barcelona- Economics and Economic History, Spain)	The goal of the study was to develop and validate a regional sustainability ranking system in Spain using multi-criteria analysis and sensitivity analysis (plurality of scenarios and assessment of their impact on the final outcome).
Saisana et al. (2005) Munda et al. (2008)	Environmental Sustainability Index (developers: Yale University and Columbia University)	The index aimed at benchmarking the ability of more than 130 nations to protect the environment over the next several decades by integrating data that track natural resource endowments, past and present pollution levels, environmental management efforts, and a society's capacity to improve its environmental performance.

The approach consisted of four main steps: (a) the consideration of the two conceptual frameworks, the THES and the SJTU, (b) the design and application of the multi-

modelling approach, (c) the calculation of the simulated university ranks, and (d) the identification of sensitive and/or non-representative ranks.

4.1.1 Two conceptual frameworks – the THES and the SJTU

The underlying indicators of the THES and the SJTU ranking systems were considered in the analysis. At this point we should add that we take for granted that the indicators selected, compiled and normalised by the THES and SJTU contain good quality data.[18] To allow for a comparative assessment of the two ranking systems, only the universities that are common in the THES and the SJTU were included. Our analysis thereafter focused on a total of 88 universities: 27 Universities in European countries, 48 Universities in the USA, two Universities in Australia, four Universities in Canada, one University in Israel and six Universities in Japan. In principle, the robustness assessment of each ranking system could have been done independently using the full set of Universities in either case. However, our aim is to comparatively assess the two ranking systems while studying their behaviour on the same sources of uncertainty. To this end, the same set of universities has to be considered. The indicators scores are only provided in scaled form in the relevant literature (see Table A 1 in the appendix) and thereafter our analysis will not include any uncertainty on the normalisation issue.

4.1.2 Multi-modelling approach

A multi-modelling approach was applied to assign weights and to further aggregate the indicators scores into the final composite indicator score (Table 13). The approach consisted of simulations (saturated sampling) based on combinations of three main assumptions on (a) the number of indicators included, (b) the weights attached to the indicators and (c) the aggregation rule. We carry out a total of 70 simulations ($=7\times4+7\times3\times2$) for either the THES or the SJTU framework[19].

[18] However, one should note that even if a dataset has been submitted to rigorous quality check, the following problems have still to be tackled to arrive at a composite measure: the consideration of measurement error in the data, the imputation of missing data, the treatment of outliers and extreme values, the transformation of skewed indicators, the standardization/normalization of the data (e.g., re-scaling, standardisation).

[19] Data envelopment analysis is only used with the additive aggregation rule.

Table 13. Scenarios for the assessment of the THES and SJTU

Assumption	Alternatives	Comments
number of indicators	▪ all six indicators included or ▪ one-at-time excluded (6 options)	A set of indicators is not the reality but only a descriptive model of it. It is important to check the relevance and the explicative capacity of the conceptual framework used. We deal with this issue in the case of the THES and SJTU by looking at the sensitivity of results to the exclusion of a single indicator from the dataset. Note however, that, as we said earlier, it is beyond the scope of this study to discuss the relevance of the indicators for measuring university excellence.
weighting method	▪ original set of weights, ▪ factor analysis, ▪ equal weighting, ▪ data envelopment analysis [20]	There is both a technical and a socio-political component in this consideration. In the THES or the SJTU ranking systems, there are no theoretical justifications for the selected weights. In our analysis, in order to examine whether different weights give a very different final ranking, we have considered four popular weighting methods.
aggregation rule	▪ additive, ▪ multiplicative, ▪ Borda multi-criterion	When a set of individual indicators is aggregated, a fundamental issue is compensability that refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some indicators by a sufficiently large advantage on another indicator, whereas smaller advantages would not do the same. Different functional forms to deal with the issue of compensability are the multiplicative (or else termed geometric aggregation) and the Borda-type multi-criteria approach (see Box 2 in the appendix).

4.1.3 Calculation of the simulated university ranks

The *frequency matrix* of a university's rank in either framework, the THES or the SJTU, is calculated across all the simulated scenarios. These frequency matrices can be very informative in the presentation of the results of a ranking system as they synthesize and make explicit the uncertainty contained in the university rankings. For each university, the numbers in the frequency matrices indicate the percentage of times a given university was ranked in a given position (groups of 5 positions are displayed, namely rank 1-5, rank 6-10, etc.) across all simulations carried out.

[20] Endogenous weights derived by data envelopment analysis. These weights allow to check how stable the bottom position of a university is – since the best set of weights for that university is used – and to derive policy priority. See Box 3 for more details on the data envelopment analysis.

4.1.4 Identification of sensitive or non-representative ranks

Our objective in undertaking this multi-modeling approach is two-fold:

(a) to identify universities for which the classification in the performance ladder of the world-class universities is highly sensitive to the methodological assumptions,

(b) to identify universities whose originally proposed rank (either by THES or the SJTU) is non-representative of the plurality of the methodological scenarios considered.

We will use non-sophisticated formulas in quantifying these objectives. With respect to the first objective and due to lack of any pre-established threshold or criterion, we will use the 90% confidence interval of a university's rank, and classify a university as having:

- ***Low sensitivity*** *to the methodological assumptions* if:
  $\text{simulated rank range} \leq 13$, (roughly 1/7 of the positions in the classification)
- ***Median sensitivity*** *to the methodological assumptions* if:
  $13 < \text{simulated rank range} < 22$,
- ***High sensitivity*** *to the methodological assumptions* if:
  $\text{simulated rank range} \geq 22$, (roughly 1/4 of the positions in the classification)

As far as the second objective is concerned, we confront the common THES rank or the common SJTU rank with the simulated median rank estimated over all methodological scenarios. The simulated median for each university could be seen as an unbiased "summary picture" of a university's performance given the selected set of indicators and across all the scenarios considered, provided that the scenarios are representative of the space of inference. We believe the scenarios we have employed represent distinct, diverse and at times contradicting approaches for aggregating information on university performance and it is therefore advantageous to include them in our analysis. As for the previous objective, the criterion to classify a university rank proposed by THES or SJTU as unreliable or non-representative is set at a 22-rank difference between the common THES or SJTU rank and the simulated median:

Unreliable/ non-representative: if $\left|\textit{Common rank} - \textit{simulated median rank}\right| > 13$,

(roughly 1/7 of the positions in the classification)

In the context of sensitivity analysis, we will study the impact of the methodological assumptions on the university ranks. To this end, we calculate the absolute difference between a simulated university rank in a given scenario and the corresponding THES (or SJTU) rank. We will then use as our two "sensitivity measures" for each scenario: (a) the median and (b) the 90th percentile of the absolute differences over the entire set of the 88 universities. To be more explicit, the first sensitivity measure shows how many positions of shift is observed for half of the universities in the dataset when the assumptions of the scenario are taken into account compared to the original assumptions made in either the THES or the SJTU. The second sensitivity measure shows how many positions of shift is observed for the 10% most affected universities in the dataset. In the case of the 88 universities studied, a value of 10.0 in the first sensitivity measure implies that 44 universities shift 10 positions or less with respect to the THES rank for this given scenario, whilst a value of 10.0 in the second sensitivity measure implies that the 9 (10% of the total number of universities) most affected universities shift 10 positions or more with respect to the THES rank for this given scenario.

First sensitivity measure: $median(|Common \quad rank - simulated\ median\ \text{rank}|)$

Second sensitivity measure: $90^{th}\ percentile(|Common \quad rank - simulated\ median\ \text{rank}|)$

4.2 Robustness assessment of the THES ranking system

4.2.1 Simulated ranks

The frequency matrix of the university ranks based on the THES framework is shown in Table 14. Harvard, Cambridge, Oxford, Yale and the Imperial College of London are the Top 5 universities according to the original THES. When the methodological assumptions in the development of the THES Index are acknowledged, it is evident that Harvard, Cambridge and Oxford remain undoubtedly in the Top5 positions (in more than 80% of the simulations). In fact, Harvard is a Top5 university (100% frequency), followed by Oxford (90% frequency) and then by Cambridge (81% frequency). The Yale University is more likely to perform between the 6th and the 10th position (64% frequency), than in the Top5 (36% frequency) as suggested by the THES ranking. As we move towards the middle ranked universities, the impact of the assumptions on the rank becomes even more pronounced. For

example, the Kyoto University, which is ranked in the 24th position in THES, has a very uncertain position when acknowledging the uncertainties: it could be ranked anywhere between the 21st and 65th position, with an almost equal frequency. The case of the University of Basel offers another pronounced example. THES ranks the University of Basel in the 68th position, but the uncertainties are acknowledged, the University of Basel could be ranked between the 21st and 88th position. Universities that are ranked in the lower end with a certain degree of confidence are the Hebrew University of Jerusalem, the Irvine University of California, the Chapel Hill University of North Carolina, the St. Louis University of Washington, the University of Goettingen and the Rutgers State University of New Brunswick: these universities are ranked in the top 81-88 universities more than 50% of the times.

Table 14. Frequency matrix of the Universities ranks according to the THES framework

THES (70 scenarios)

	Rank 1-5	Rank 6-10	Rank 11-15	Rank 16-20	Rank 21-25	Rank 26-30	Rank 31-35	Rank 36-40	Rank 41-45	Rank 46-50	Rank 51-55	Rank 56-60	Rank 61-65	Rank 66-70	Rank 71-75	Rank 76-80	Rank 81-88	Original THES score	Original THES rank	Common THES rank
Harvard Univ	100																	100	1	1
Univ Cambridge	81	17																97.6	2	2
Univ Oxford	90	9																97.6	2	2
Yale Univ	36	64																97.6	2	2
Imperial Coll London	66	27																97.5	5	5
Princeton Univ	27	56	17															97.2	6	6
California Inst Tech	23	43	23	10														96.5	7	7
Univ Chicago	9	86	6															96.5	7	7
Univ Coll London	36	41	16	6														95.3	9	9
Massachusetts Inst Tech (MIT)	26	14	24	14	19													94.6	10	10
Columbia Univ		19	34	20	21													94.5	11	11
McGill Univ		50	30	14														93.9	12	12
Duke Univ		6	21	19	16	11	9	7										93.4	13	13
Univ Pennsylvania		16	46	37														93.3	14	14
Johns Hopkins Univ			20	14	23	33		7										92.9	15	15
Australian Natl Univ		14	37	31	14													91.6	16	16
Tokyo Univ			7	16	14	10	14	23	10									91.1	17	17
Stanford Univ		9	20	16	24	11	10											90.6	19	18
Carnegie Mellon Univ			43	40	13													90	20	19
Cornell Univ			10	27	24	24	11											90	20	19
Univ California - Berkeley	7	7	30	40	14													89.7	22	21
Univ Edinburgh			11	50	29	7												88.8	23	22
King's Coll London		7	40	10	24	7	9											88.2	24	23
Kyoto Univ					13	16	13	6	13	11	10	6	9					87.2	25	24
Ecole Normale Super Paris				14	39	37	9											87.1	26	25
Univ Melbourne			11	27	34	21												85.9	27	26
Northwestern Univ					17	40	41											85	29	27
Univ Manchester			17	29	19	20	6	10										84.7	30	28
Brown Univ					14	40	39											84.5	32	29
Univ British Columbia						26	30	30	9									84.3	33	30
Univ Bristol			11	24	20	24	10	7										84.1	37	31
Univ Michigan - Ann Arbor						20	40	24	7									83.8	38	32
Univ California - Los Angeles					6	6	17	11	10	13	7	6	19					82.8	41	33
Swiss Fed Inst Tech - Zurich			13	14	23	13	19	11	6									82.5	42	34
Univ Toronto			6	6	13	14	7	19	10	11	6							80.6	45	35
Osaka Univ							7	14	16	7	7	10	14	14	6			80	46	36
Boston Univ						11	34	40	6	6								79.7	47	37
New York Univ								19	33	20	7	9	7					77.8	49	38
Univ Texas - Austin					7	6	13	19	20	6	16	7	6					77.1	51	39
Univ Washington - Seattle								19	9	19	20	19	10					76.7	55	40
Univ Wisconsin - Madison							14	17	17	23	11	10						76.7	55	40
Univ California - San Diego							7		11	6	7	6	6	16	21	17		76.3	58	42
Univ Heidelberg							13	36	29	10	10							75.5	60	43
Univ Birmingham					13	19	16	19	17	6	6							74.1	65	44
Univ Munich							16	24	19	17	10	9	6					74.1	65	44
Tech Univ Munich						16	24	14	17	9	6			7				73.9	67	46
Univ Sheffield					13	13	21	17	10	9	7							73.7	68	47
Univ Nottingham					26	14	14	14		9								73.2	70	48
Uppsala Univ										14	6	10	20	11	26	6		73	71	49
Univ Illinois - Urbana Champaign									10	19	9	26	17	14				72.6	73	50
Univ Pittsburgh - Pittsburgh							13	9	13	16	19	17	7					72.2	77	51
Purdue Univ - West Lafayette								6	29	26	16	9						72.2	77	51
Univ Maryland - Coll Park								7	13	17	44	16						72.1	79	53
Vanderbilt Univ									17	21	23	13	11					71.9	82	54
Univ Leiden								11	30	11	20	20	6					71.7	84	55
Case Western Reserve Univ								11	9	19	13	14	10	10	7			71.6	85	56
Pennsylvania State Univ - Univ Park											9	6	14	17	19	13	19	70.5	90	57
Tokyo Inst Tech									6	27	41	7	10	7				70.5	90	57
Rice Univ									10	21	16	36	7	6				70.3	92	59
Univ Rochester										11	14	17	26	11		11		69.3	95	60
Univ California - Davis												13	10	6	21	34	13	69.1	96	61
Univ Helsinki								6	27	10	11	19	9	14				68.2	100	62
Tohoku Univ											9	19	24	21	13	6		68	102	63
Lund Univ									6	9	16	17	27	13	9			66.9	106	64
Univ Colorado - Boulder											6		16	11	14	16	31	66.8	107	65
McMaster Univ										9			11	17	24	31		66.6	108	66
Nagoya Univ												14	10	20	20	21	9	66.1	112	67
Univ Basel					6		7	7	7	6	9	7	14	13	9	7	7	65.6	114	68
Univ California - Santa Barbara										7				9	6	23	47	65.5	117	69
Univ Southern California									10	9	14	14	17	14	9	7	6	65.4	119	70
Ohio State Univ - Columbus									7	7	11	19	19	17	10	10		65.3	120	71
Texas A&M Univ - Coll Station												9	10	23	38	11	7	64.9	122	72
Hebrew Univ Jerusalem												6		9	11	7	54	64	128	73
Univ Paris 06											7	7	9	14		7	47	63.7	132	74
Univ Arizona													10	26	39	19		63.1	134	75
Univ Florida														13	24	27	33	63	135	76
Indiana Univ - Bloomington												10	13	26	21	19	9	62.7	137	77
Univ California - Irvine															14	16	66	62.5	140	78
Univ Zurich									10	7	7	9	10		11	13	23	62.5	140	78
Univ Minnesota - Twin Cities														7	9	39	41	62.3	142	80
Univ Freiburg												6	7	17	26	7	34	62.2	144	81
Univ North Carolina - Chapel Hill														6	6	33	56	61.1	151	82
Michigan State Univ									6	9	7	19	20	10	7	10	11	60.6	159	83
Washington Univ - St. Louis										9		9	6	6		7	53	59.9	161	84
Univ Goettingen															6	11	74	58.8	168	85
Rutgers State Univ - New Brunswick															7	23	69	57.8	177	86
Univ Strasbourg 1												7	7	27	13	13	24	57.1	184	87
Univ Oslo														7	17	21	50	56.6	188	88

Legend:

Frequency lower 15%

Frequency between 15 and 30%

Frequency between 30 and 50%

Frequency greater than 50%

Note: Frequencies lower than 4% are not shown

4.2.2 Sensitivity of university ranks to the methodological assumptions

Following the distinction we made earlier on the degree of sensitivity of the universities ranks to the methodological assumptions, there are 10 universities whose rank has low degree of sensitivity to the methodological assumptions in the THES, 19 universities with medium sensitivity and 59 universities that are highly sensitive.

Table 15 lists the universities whose simulated rank (90% confidence interval) is highly sensitive. For those universities, the space of inference of a university's rank is too wide to draw any meaningful conclusions. These universities are spread in the entire classification ladder, except for the Top9 positions in THES, in fact they are classified between the 10th position (Massachusetts Inst Tech) and the 88th position (University of Oslo).

The Massachusetts Inst Tech, which performs very well according to the THES (rank 10) is ranked between the 2nd and the 25th position, when the methodological assumptions are changed. On the other hand, the University of Oslo in Norway, which is ranked in the 88th position in THES, is situated between the 58th and the 88th position, when the methodological assumptions are changed.

What is common to most of the "highly sensitive" universities is the fact that they are neither very good nor very bad in the majority of the THES indicators but somewhere in between (see Table 26 and Table 27 in Section 5 for information on the underlying indicators).

The only country whose universities are not highly sensitive to the methodological assumptions is Australia. The two Australian universities, the Australian National University or the University of Melbourne have a simulated rank range lower than 22 positions, yet close to that.

Table 15. THES Index: 59 universities whose simulated rank is highly sensitive to the methodological assumptions

University	Country	THES rank	Range of ranks
Univ British Columbia	Canada	30	[26 50]
Univ Toronto	Canada	35	[11 58]
McMaster Univ	Canada	64	[49 80]
Univ Helsinki	Finland	62	[37 70]
Univ Paris 06	France	74	[44 87]
Univ Strasbourg 1	France	87	[47 88]
Univ Heidelberg	Germany	43	[33 55]
Univ Munich	Germany	44	[32 62]
Tech Univ Munich	Germany	45	[27 67]
Univ Freiburg	Germany	81	[57 88]
Univ Goettingen	Germany	85	[54 88]
Hebrew Univ Jerusalem	Israel	73	[46 88]
Tokyo Univ	Japan	17	[15 46]
Kyoto Univ	Japan	24	[22 63]
Osaka Univ	Japan	37	[34 74]
Tokyo Inst Tech	Japan	58	[44 70]
Tohoku Univ	Japan	63	[40 79]
Nagoya Univ	Japan	67	[49 83]
Univ Leiden	Netherlands	56	[38 63]
Univ Oslo	Norway	88	[58 88]
Uppsala Univ	Sweden	48	[42 78]
Lund Univ	Sweden	64	[42 74]
Swiss Fed Inst Tech - Zurich	Switzerland	34	[13 41]
Univ Basel	Switzerland	68	[25 83]
Univ Zurich	Switzerland	80	[36 86]
King's Coll London	UK	23	[9 35]
Univ Manchester	UK	28	[13 38]
Univ Bristol	UK	31	[13 40]
Univ Birmingham	UK	45	[23 58]
Univ Sheffield	UK	47	[23 65]
Univ Nottingham	UK	49	[20 67]
Massachusetts Inst Tech (MIT)	US	10	[2 25]
Duke Univ	US	13	[8 60]
Johns Hopkins Univ	US	15	[12 36]
Stanford Univ	US	18	[4 37]
Univ Michigan - Ann Arbor	US	32	[22 46]
Univ California - Los Angeles	US	33	[23 70]
New York Univ	US	38	[37 68]
Univ Texas - Austin	US	39	[23 63]
Univ Wisconsin - Madison	US	40	[28 60]
Univ Washington - Seattle	US	41	[35 62]
Univ California - San Diego	US	42	[31 78]
Univ Illinois - Urbana Champaign	US	50	[41 71]
Univ Pittsburgh - Pittsburgh	US	51	[32 70]
Purdue Univ - West Lafayette	US	51	[27 68]
Vanderbilt Univ	US	53	[29 74]
Case Western Reserve Univ	US	55	[36 76]
Pennsylvania State Univ - Univ Park	US	57	[50 87]
Rice Univ	US	59	[41 68]
Univ Rochester	US	60	[44 78]
Univ California - Davis	US	61	[57 82]
Univ Colorado - Boulder	US	66	[54 86]
Univ California - Santa Barbara	US	69	[47 87]
Ohio State Univ - Columbus	US	70	[45 78]
Univ Southern California	US	71	[43 81]
Texas A&M Univ - Coll Station	US	72	[55 81]
Indiana Univ - Bloomington	US	78	[56 83]
Michigan State Univ	US	83	[41 86]
Washington Univ - St. Louis	US	84	[43 88]

4.2.3 Non-representative university ranks of the plurality of the scenarios

The second objective of the analysis is to identify those universities whose proposed THES rank may not be reliable or not representative of the plurality of the methodological scenarios. There is a very high degree of correlation between the common THES rank and the simulated median, $r^2 = .907$ ($p < 0.001$, $n = 88$), which produces a high degree of confidence that for most of the universities studied, the THES rank is reliable and no deliberate bias was introduced in the THES.

However, caution is needed in the case of few universities whose performance based on the THES deviates from the simulated median. According to the criterion we set earlier, the THES rank for 10 universities is likely to be unreliable or non-representative. For the remaining universities, we can have high confidence that they are roughly placed, on average, in the correct place and this classification may be used for policy-making or for benchmarking purposes. However, the degree of volatility discussed previously needs to be taken into consideration prior to drawing final conclusions.

Table 16 lists the 10 universities whose simulated median rank differs from the THES rank by more than 13 positions. These universities are ranked between the 17th position (Tokyo University) and the 87th position (University of Strasbourg) in the THES index. There is only one French university (University of Strasbourg), three Japanese universities (Tokyo, Kyoto, Osaka), one Swedish university (Uppsala), one university in the UK (Nottingham) and four universities in the US (California - San Diego, Pennsylvania State University – Park, California – Davis, Michigan State).

Table 16. THES: 10 universities whose THES rank is non-representative of the simulated scenarios

University	*Country*	*THES rank*	*Median rank*
Univ Strasbourg 1	France	87	71
Tokyo Univ	Japan	17	31
Kyoto Univ	Japan	24	42
Osaka Univ	Japan	37	53
Uppsala Univ	Sweden	48	64
Univ Nottingham	UK	49	33
Univ California - San Diego	US	42	66
Pennsylvania State Univ - Univ Park	US	57	71
Univ California - Davis	US	61	75
Michigan State Univ	US	83	62

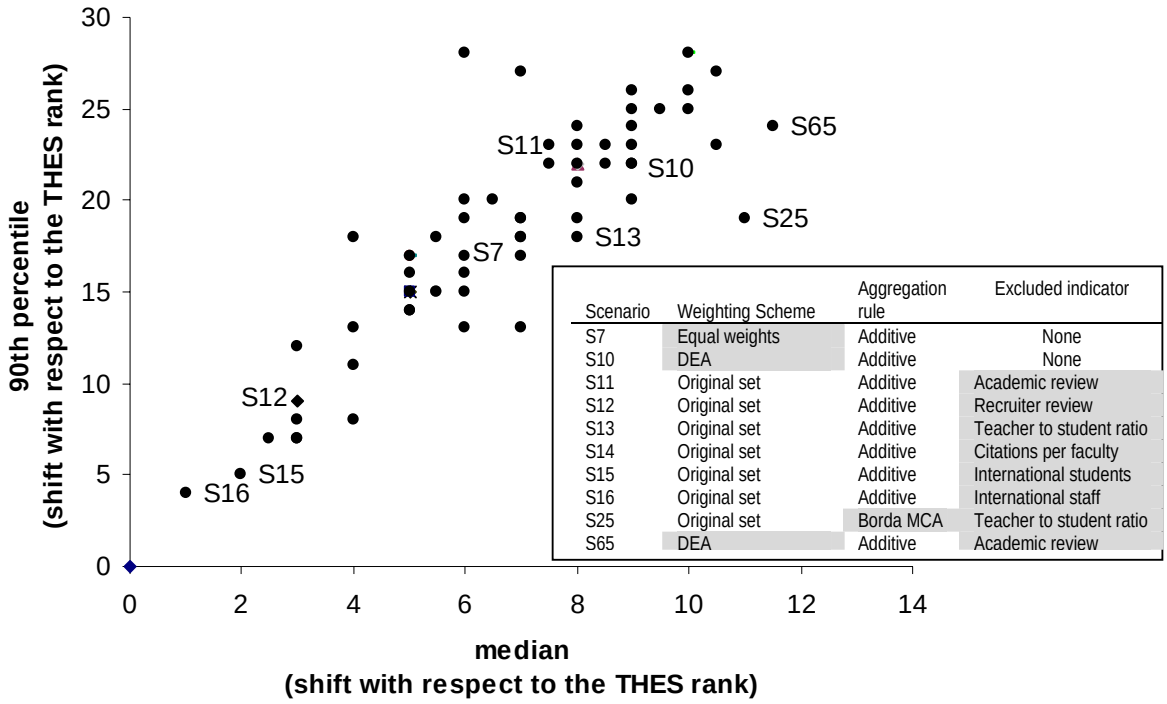
4.2.4 Impact of the assumptions on the university ranks

Complementary to the two objectives discussed above, it would be informative to study the impact of the methodological assumptions on the university rank compared to the THES rank. To this end, we consider the two sensitivity measures described previously for each scenario and plot them in Figure 2.

The more distant from the origin a methodological scenario is, the more it influences the THES rank. Ten of the 70 simulated scenarios have an impact of 10 positions or more on the ranks of at least half of the universities, as shown by the first sensitivity measure. Detailed description of the scenarios and the corresponding values for the two sensitivity measures are provided in the Appendix in Table A 3. These ten scenarios employ either DEA or Equal weighing or Factor Analysis for the determination of the weights. All three aggregation rules are employed, although the Borda MCA appears only once. All scenarios are characterised by the exclusion of one indicator, either the academic review, or the teacher to student ratio, or the recruiter review in the majority of the cases.

The combination of the DEA approach and upon exclusion of the academic review indicator has the highest impact on the universities ranks overall (Scenario 65). In this scenario, half of the universities shift less than 11.5 positions with respect to the THES rank, and the 10% most affected universities shift more than 24 positions. Follows the impact due to the combination of the original weights used in THES, under a Borda MCA aggregation rule and upon eliminating the teacher to student ratio indicator (Scenario 25). In this scenario, half of the universities shift less than 11.0 positions with respect to the THES rank, and the 10% most affected universities shift more than 19 positions. Even when all six indicators are included in the simulations but the assumptions regarding the weighting or aggregation rule are challenged, the impact on half of the universities is 9 positions or less and 10% of the universities shift more than 22 positions (Scenario 10: ‘worst case’ scenario based on all six indicators). Had an equal weighting been used for the THES indicators (Scenario 7), the impact would have been translated to a shift of 6 positions or less for half of the universities and a shift of 17 positions or more for the 10% most affected universities.

Figure 2. Sensitivity analysis: impact of the assumptions on the THES ranks



Note: median versus 90th percentile of the absolute differences between a simulated and the THES rank in a given scenario (over 88 universities). Detailed results are provided in the Appendix, Table A 3)

If we keep the assumptions on the weighting and the aggregation as in the original THES, but we eliminate one indicator at a time, it is interesting to notice in Table 17 that it is the exclusion of the teacher to student ratio that bears the highest impact on the university ranks (8 positions or less for half of the universities), followed by the impact due to the exclusion of the academic review (7 positions or less for half of the universities). The impact of the exclusion of the other four indicators is significantly lower.

Table 17. Impact of the exclusion of a single indicator from THES

Scenario	Excluded indicator	First Sensitivity Measure: 50th percentile	Second Sensitivity Measure: 90th percentile
S13	Teacher to student ratio	8.0	18.0
S11	Academic review	7.5	23.0
S12	Recruiter review	3.0	9.0
S14	Citations per faculty	3.0	7.0
S15	International students	2.0	5.0
S16	International staff	1.0	4.0

4.3 Robustness assessment of the SJTU ranking system

4.3.1 Simulated ranks

The frequency matrix of the university ranks based on the SJTU framework is shown in **Error! Reference source not found.**. Harvard, Stanford, Berkley, Cambridge and the M.I.T. are Top5 universities according to the original SJTU. When we acknowledge the methodological assumptions in the development of the SJTU, it is evident that all five universities remain indeed in the Top5 with a frequency higher than 80%. To be more specific, Harvard and Berkeley are Top5 in 100% of the simulations, followed by Cambridge (96% of the times), Stanford (87% frequency) and by the M.I.T. (81% frequency). The universities that are ranked between the 6^{th} and 10^{th} positions when acknowledging the methodological uncertainties are also found in similar positions in the SJTU Index: these are the California Inst Tech, the Columbia University, the Princeton University, the University of Chicago and the University of Oxford. The Yale University has a 50% chance to perform between the 6^{th}-10^{th} position and 50% chance to perform between the 11^{th}-15^{th} position. As we move towards the middle ranked universities, the impact of the assumptions on the rank becomes stronger. For example, the University of Ann Arbor in Michigan presents a bimodal distribution in its rank: it could be ranked somewhere between the 11^{th}-30^{th} position (63% frequency) or between the 66^{th}-70^{th} position (20% frequency). Universities that are ranked in lower positions with a certain degree of confidence are the Boston University, the Texas A&M University-College Station, the Nagoya University, the Lund University and the Tokyo Institute of Technology. These universities are ranked in the Top81-88 universities more than 50% of the times.

Table 18. Frequency matrix of the Universities ranks according to the SJTU

SJTU (70 scenarios)

	Rank 1-5	Rank 6-10	Rank 11-15	Rank 16-20	Rank 21-25	Rank 26-30	Rank 31-35	Rank 36-40	Rank 41-45	Rank 46-50	Rank 51-55	Rank 56-60	Rank 61-65	Rank 66-70	Rank 71-75	Rank 76-80	Rank 81-88	Original SJTU score	Original SJTU rank	Common SJTU rank
Harvard Univ	100																	100	1	1
Stanford Univ	87	13																73.7	2	2
Univ California - Berkeley	100																	71.9	3	3
Univ Cambridge	96																	71.6	4	4
Massachusetts Inst Tech (MIT)	81	19																70	5	5
California Inst Tech	27	61	11															66.4	6	6
Columbia Univ	11	89																63.2	7	7
Princeton Univ		71	19	10														59.5	8	8
Univ Chicago		56	40															58.4	9	9
Univ Oxford		87	13															56.4	10	10
Yale Univ		50	50															55.9	11	11
Cornell Univ		27	73															54.3	12	12
Univ California - Los Angeles		11	86															52.6	13	13
Univ California - San Diego			43	51														50.4	14	14
Univ Pennsylvania			83	16														49	15	15
Univ Washington - Seattle				89	9													48.2	16	16
Univ Wisconsin - Madison			43	56														48	17	17
Johns Hopkins Univ			13	59	26													46.1	19	18
Tokyo Univ			11	30	46	13												45.9	20	19
Univ Michigan - Ann Arbor			11	16	20	16					6			20				44	21	20
Kyoto Univ				39	59													43.1	22	21
Imperial Coll London				26	61	13												43	23	22
Univ Toronto				19	60	21												43	23	22
Univ Coll London				23	74													42.8	25	24
Univ Illinois - Urbana Champaign				34	44	21												42.7	26	25
Swiss Fed Inst Tech - Zurich			6	29	27	34												39.9	27	26
Washington Univ - St. Louis					33	63												39.7	28	27
Northwestern Univ						89	11											38.2	29	28
New York Univ					10	66	24											38	30	29
Duke Univ					14	20	30							21	11			37.4	32	30
Univ Minnesota - Twin Cities					9	19	24	11						19	14			37	33	31
Univ Colorado - Boulder						31	47	19										36.6	34	32
Univ California - Santa Barbara						19	16	21	10					19				35.8	35	33
Univ British Columbia						6	80	14										35.4	36	34
Univ Maryland - Coll Park						9	49	43										35	37	35
Univ Texas - Austin							20	47	19	7								34.4	38	36
Univ Paris 06						33	37	21	9									33.8	39	37
Vanderbilt Univ							51	29	11	6								33.6	41	38
Univ California - Davis							14	13	10	9					6		31	32.7	43	39
Pennsylvania State Univ - Univ Park							9	21	7	9	6		6	10	17	11		32.7	43	39
Univ California - Irvine							6	11	23	14	10			13	11		6	32.5	45	41
Rutgers State Univ - New Brunswick							6	19	27	19	17		7					32.1	47	42
Univ Manchester							9	41	37	13								32	48	43
Univ Pittsburgh - Pittsburgh							13	13	26		6			7	23			31.9	49	44
Univ Southern California									11	13	13		11	20	14		6	31.4	50	45
Univ Florida								13	17	19				13	13	14		31.1	51	46
Univ Edinburgh								24	34	24	10	6						30.8	53	47
Univ Munich						9	6	27	21	20	14							30.8	53	47
Tech Univ Munich							23	23	19	11	7	11						30.6	56	49
Australian Natl Univ									39	26	21	9						30.4	57	50
Univ North Carolina - Chapel Hill									10	14	17		7	13		30		30.2	58	51
Univ Zurich								13	20	24	20	11	10					30.2	58	51
Carnegie Mellon Univ						6	10	16	17	17	7	16	6					30	60	53
Ohio State Univ - Columbus										11	14	11		10	9	26	11	29.5	61	54
Univ Bristol										14	23	44	10					29.4	62	55
McGill Univ								14	14	13	11	9		9	14	10		29.3	63	56
Hebrew Univ Jerusalem								20	9	24	16	14	13					29	64	57
Univ Heidelberg									26	14		30	10	6	7			28.9	65	58
Uppsala Univ									10	7	24	11	20		6	10		28.8	66	59
Osaka Univ										10	14	13		13		29	10	28.6	67	60
Purdue Univ - West Lafayette										9	10	20	19	16	20	6		28.3	68	61
Univ Oslo								7	14	17	13		17	7	10	9		28.2	69	62
Brown Univ										21	21	20	16	11	6			28.1	70	63
Univ Leiden									10	16	13	10	20	26	6			28	71	64
Univ Sheffield										20	23	19	27	9				27.9	72	65
Univ Helsinki										7	17	17	19	29	11			27.8	73	66
Univ Arizona											13	7		9	13	9	43	27.6	74	67
Univ Rochester										20	20	19	19	13	6			27.4	75	68
Tohoku Univ											9	10	13	20	11	20	17	27.1	76	69
Case Western Reserve Univ										19	20	16	11	11	7	7		26.8	78	70
Univ Melbourne													27	14	24	20	11	26.6	79	71
Michigan State Univ													9	14	11	13	49	26.1	80	72
Univ Nottingham												14	16		31	20	11	26	81	73
Univ Basel									11	13	17	11	10	11	13	13		25.9	82	74
Boston Univ													6	10	17	14	50	25.5	83	75
Ecole Normale Super Paris								7	13	7	7	13	11		10	11	14	25.5	83	75
King's Coll London												19	14		16	26	23	25.5	83	75
McMaster Univ													20	11		21	41	25	87	78
Rice Univ										11	10	23		11	16	19	6	25	87	78
Univ Goettingen										7	6	10	7	19	16	14	19	25	87	78
Indiana Univ - Bloomington												6	14	14		14	47	24.9	90	81
Texas A&M Univ - Coll Station														6	13	7	70	24.8	91	82
Univ Birmingham													16	19	10	26	30	24.5	92	83
Nagoya Univ													6		9	17	69	24.1	94	84
Univ Freiburg												11	14	7	7	14	43	24.1	94	84
Lund Univ														6	11	21	57	23.9	97	86
Univ Strasbourg 1												16	6	13	10	16	34	23.8	99	87
Tokyo Inst Tech													6	11	13	17	51	23.8	99	87

Legend:
Frequency lower 15%
Frequency between 15 and 30%
Frequency between 30 and 50%
Frequency greater than 50%
Note: Frequencies lower than 4% are not shown

4.3.2 Sensitivity of university ranks to the methodological assumptions

The university ranks in the SJTU classification have different degrees of sensitivity to the methodological assumptions. There are 31 universities whose rank has low degree of sensitivity to the methodological assumptions in the SJTU, 5 universities with medium sensitivity and 52 universities that are highly sensitive.

Table 19 lists the universities whose simulated rank (90% confidence interval) is highly sensitive and thus the space of inference of those ranks is too wide to draw meaningful conclusions. These universities are ranked between the 20th position (University of Michigan - Ann Arbor) and the 88th position (University of Strasbourg 1). The University of Michigan - Ann Arbor, which performs relatively well according to the SJTU Index (rank 20) can perform anywhere between the 11th and the 70th position, when uncertainties are acknowledged. On the other hand, the University of Strasbourg 1, which is originally ranked in the 88th position is actually performing somewhere between the 52nd and the 88th position.

Justification of the high volatility of those universities can be found by going back to the information provided by the underlying indicators in the SJTU and presented in Table 26 and Table 27. Just to anticipate few examples here, the McGill University in Canada whose simulated SJTU rank ranges between the 36th and the 80th position, has no indicator above the US average, three SJTU indicators close to the US average (*alumni winning Nobel prizes and field metals, articles in Science and Social Citation Index* and *academic performance with respect to size*) and three indicators below the US average (*staff winning Nobel prizes an field medals, highly cited researchers* and *articles published in Nature and Science*). Given these scores in the underlying indicators, it is also apparent that no matter how the data are treated and combined, the McGill University will not perform among the Top20 universities (due to lack of very indicators scores).

All of the 13 countries included in the analysis (namely Australia, Canada, Finland, France, Germany, Israel, Japan, Netherlands, Norway, Sweden, Switzerland, UK and US) have universities that are highly sensitive to the methodological assumptions.

Table 19. SJTU Index: 52 universities whose simulated rank is highly sensitive to the methodological assumptions

University	Country	SJTU rank	Range of ranks
Univ Melbourne	Australia	71	[61 87]
McMaster Univ	Canada	81	[62 86]
McGill Univ	Canada	55	[36 80]
Univ Helsinki	Finland	67	[49 72]
Univ Strasbourg 1	France	88	[52 88]
Ecole Normale Super Paris	France	77	[39 87]
Univ Freiburg	Germany	85	[56 88]
Univ Goettingen	Germany	79	[46 87]
Univ Heidelberg	Germany	58	[41 73]
Tech Univ Munich	Germany	49	[31 60]
Univ Munich	Germany	48	[29 55]
Hebrew Univ Jerusalem	Israel	57	[36 65]
Tokyo Inst Tech	Japan	86	[62 87]
Nagoya Univ	Japan	84	[62 88]
Tohoku Univ	Japan	69	[52 84]
Osaka Univ	Japan	60	[48 81]
Univ Leiden	Netherlands	64	[42 71]
Univ Oslo	Norway	62	[38 80]
Lund Univ	Sweden	86	[65 88]
Uppsala Univ	Sweden	59	[37 77]
Univ Basel	Switzerland	74	[43 78]
Univ Zurich	Switzerland	52	[39 65]
Univ Birmingham	UK	83	[63 86]
King's Coll London	UK	76	[58 86]
Univ Nottingham	UK	73	[55 82]
Univ Sheffield	UK	65	[47 70]
Univ Bristol	UK	56	[41 65]
Indiana Univ - Bloomington	US	82	[59 88]
Rice Univ	US	80	[46 83]
Texas A&M Univ - Coll Station	US	78	[65 88]
Boston Univ	US	75	[61 86]
Michigan State Univ	US	72	[58 87]
Case Western Reserve Univ	US	70	[45 81]
Univ Rochester	US	68	[45 72]
Univ Arizona	US	66	[50 87]
Brown Univ	US	63	[45 71]
Purdue Univ - West Lafayette	US	61	[47 78]
Ohio State Univ - Columbus	US	54	[46 83]
Carnegie Mellon Univ	US	53	[28 69]
Univ North Carolina - Chapel Hill	US	51	[41 79]
Univ Florida	US	46	[37 78]
Univ Southern California	US	45	[35 85]
Univ Pittsburgh - Pittsburgh	US	42	[33 77]
Rutgers State Univ - New Brunswick	US	42	[35 65]
Univ California - Irvine	US	41	[31 83]
Pennsylvania State Univ - Univ Park	US	40	[32 77]
Univ California - Davis	US	39	[33 86]
Univ Texas - Austin	US	36	[32 57]
Univ California - Santa Barbara	US	33	[27 82]
Univ Minnesota - Twin Cities	US	31	[23 73]
Duke Univ	US	30	[22 72]
Univ Michigan - Ann Arbor	US	20	[11 70]

4.3.3 Non-representative university ranks of the plurality of the scenarios

The second objective of the analysis is to identify those universities whose proposed SJTU rank may not be reliable or not representative of the plurality of the methodological scenarios. There is a very high degree of correlation between the common SJTU rank and the simulated median, $r^2 = .94$ ($p < 0.001$, $n = 88$), which produces a high degree of confidence that for most of the universities studied, the SJTU rank is reliable, on average, and no deliberate bias was introduced in the SJTU. Similar results were found previously for the THES index.

However, caution is needed in the case of few universities whose performance based on the SJTU deviates from the simulated median. According to the criterion we set earlier, the SJTU rank for 6 universities is likely to be unreliable or non-representative. For the remaining universities, we can have high confidence that they are roughly placed, on average, in the correct place and this classification may be used for policy-making or for benchmarking purposes. However, the degree of volatility discussed previously needs to be taken into consideration prior to drawing final conclusions.

Table 20 lists the 6 universities whose simulated median rank differs from the SJTU rank by more than 13 positions. These universities are ranked between the 39th position (Davis University of California) and the 80th position (Rice University) in the SJTU index. There is one French university (Ecole Normale Super Paris), one Swiss university (University of Basel) and four universities in the US (Rice University, University of Southern California, Pennsylvania State University-Park and University of California – Davis).

Table 20. SJTU Index: 6 universities whose SJTU rank is not representative of the simulated scenarios

University	Country	SJTU rank	Median rank
Ecole Normale Super Paris	France	77	62
Univ Basel	Switzerland	74	59
Rice Univ	US	80	66
Univ Southern California	US	45	62
Pennsylvania State Univ - Univ Park	US	40	54
Univ California - Davis	US	39	56

Compared to the relevant THES results discussed previously, there are two universities, both in the US, namely the Pennsylvania State University-Park and the University of California – Davis, whose THES or SJTU rank is not representative of the plurality of the methodological scenarios. Instead, given either the THES or the

SJTU set of indicators, these universities ranks are biased towards the methodological assumptions made in the development of the Indices.

4.3.4 Impact of the assumptions on the university ranks

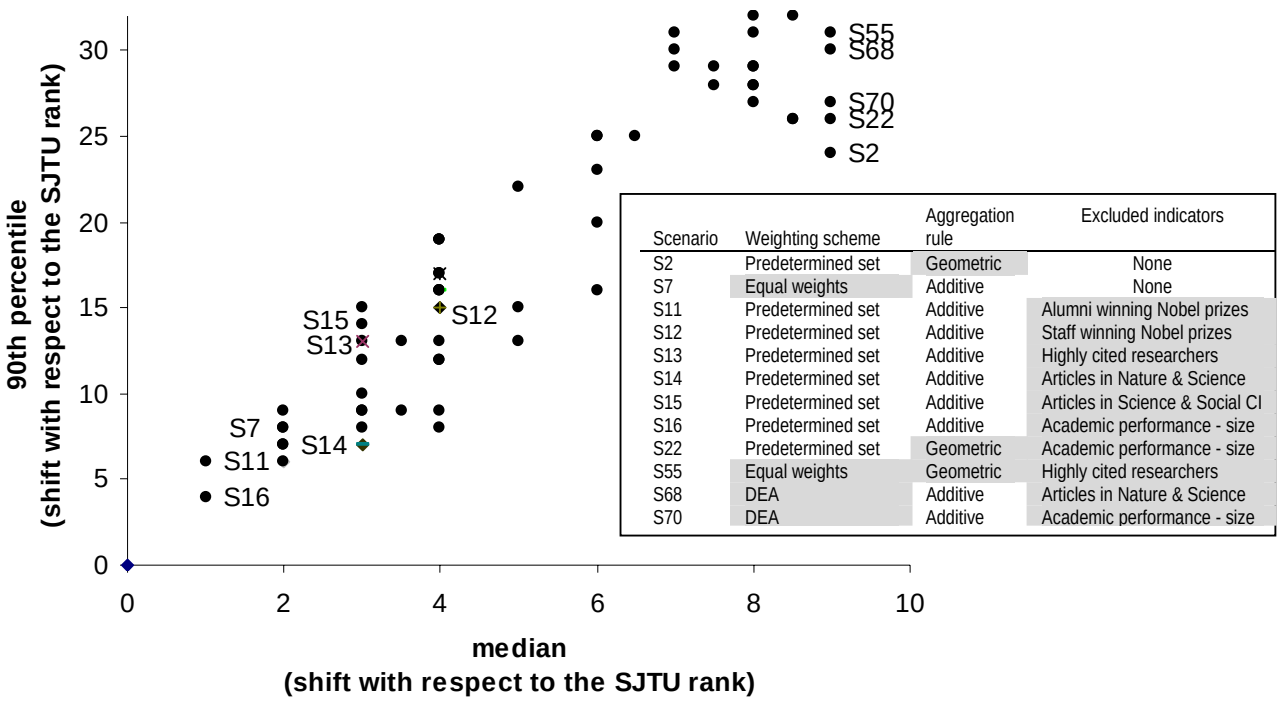
We discuss below the impact of the methodological assumptions on the university rank compared to the SJTU rank. To this end, for each scenario we consider the two sensitivity measures described previously and plot them in Figure 3.

The more distant from the origin a methodological scenario is, the more it influences the SJTU rank. Five out the 70 simulated scenarios have an impact of 9 positions or less on the ranks of at least half of the universities, as shown by the first sensitivity measure. Detailed description of the scenarios and the corresponding values for the two sensitivity measures are provided in the Appendix in Table A 4. These five scenarios employ either DEA or Equal weighing or the original weighting scheme for the determination of the weights. Only two aggregation rules appear to have a strong impact, the additive or the geometric aggregation. The Borda MCA rule, even when combined to the other assumptions, such as the weighting scheme or the exclusion of an indicator, is not particularly influential on the ranks.

The combination of the geometric aggregation with an equal weighting scheme and upon exclusion of the *highly cited researchers* indicator has the highest impact on the universities ranks overall (Scenario 55). In this scenario, half of the universities shift less than 9 positions with respect to the SJTU rank, and the 10% most affected universities shift more than 31 positions. Follows the impact due to the combination of the DEA approach and the exclusion of the *articles in Nature and Science* (Scenario 68). In this scenario, half of the universities shift less than 9 positions with respect to the SJTU rank, and the 10% most affected universities shift more than 30 positions.

Even when all six indicators are included in the simulations but the assumptions regarding the weighting or aggregation rule are challenged, the impact on half of the universities is less than 9 positions and 10% of the universities shift more than 24 positions (Scenario 2: ‘worst case’ scenario based on all six indicators). Had an equal weighting been used for the SJTU indicators (Scenario 7), the impact would have been translated to a shift of only 2 positions or less for half of the universities and a shift of 8 positions or more for the 10% most affected universities.

Figure 3. Sensitivity analysis: impact of the assumptions on the SJTU ranks



Scenario	Weighting scheme	Aggregation rule	Excluded indicators
S2	Predetermined set	Geometric	None
S7	Equal weights	Additive	None
S11	Predetermined set	Additive	Alumni winning Nobel prizes
S12	Predetermined set	Additive	Staff winning Nobel prizes
S13	Predetermined set	Additive	Highly cited researchers
S14	Predetermined set	Additive	Articles in Nature & Science
S15	Predetermined set	Additive	Articles in Science & Social CI
S16	Predetermined set	Additive	Academic performance - size
S22	Predetermined set	Geometric	Academic performance - size
S55	Equal weights	Geometric	Highly cited researchers
S68	DEA	Additive	Articles in Nature & Science
S70	DEA	Additive	Academic performance - size

Note: median versus 90th percentile of the absolute differences between a simulated and the SJTU rank in a given scenario (over 88 universities). Detailed results are provided in the Appendix, Table A 4)

If we keep the assumptions on the weighting and the aggregation as in the original SJTU, but we eliminate one indicator at a time, it is interesting to notice in Table 21 that it is the exclusion of the staff winning Nobel prizes and field medals that bears the highest impact on the university ranks (4 positions or less for half of the universities), followed by the impact due to the exclusion of the articles in science and social citation index or the exclusion of the highly cited researchers (3 positions or less for half of the universities). When excluding either of those three SJTU indicators, the 10% most affected universities shift 13 positions or more. The impact of the exclusion of the other three indicators is significantly lower, both at the median level and for the 10% most affected universities.

Table 21. Impact of the exclusion of a single indicator from SJTU

Scenario	Excluded indicator	First Sensitivity Measure: 50th percentile	Second Sensitivity Measure: 90th percentile
S12	Staff winning Nobel prizes	4.0	15.0
S15	Articles in Science & Social CI	3.0	14.0
S13	Highly cited researchers	3.0	13.0
S14	Articles in Nature & Science	2.0	7.0
S11	Alumni winning Nobel prizes	1.0	6.0
S16	Academic performance - size	1.0	4.0

4.4 Hybrid framework, multi-modelling approach

An integrated approach that uses a hybrid framework and a multi-modelling principle might be better able to represent the multidimensional dimensions of a university. In this approach we consider:

- ***Hybrid conceptual framework***
  - the twelve indicators of the THES and SJTU together.
- ***three main sources of methodological uncertainties***
  - inclusion of all indicators or eliminating one indicator at a time (13 alternatives),
  - weights attached to the indicators (four alternatives as in Table 13),
  - aggregation rule (three alternatives as in Table 13).

In total, $13 \times 4 + 13 \times 3 \times 2 = 130$ simulations are built for the hybrid framework of the twelve indicators of the THES and SJTU.

4.4.1 Sensitivity of university ranks to the methodological assumptions

The frequency matrix of the university ranks across the 130 simulations are shown in Table 22. Harvard, Cambridge, California Inst Tech, California – Berkeley and the M.I.T are placed without doubt in the Top5 positions. The Stanford university follows with a frequency of 45% in the Top5 and of 43% in the positions from 6th to 10th. Bimodal distributions are evident for several universities. Pronounced examples are those of the university of Michigan - Ann Arbor, the Duke university, the McGill

university, the Washington university of St. Louis, the university of Oslo and the Indiana University-Bloomington.

If we look at the median rank across the 130 simulations, the Top20 universities are mainly from the UK (Cambridge, Oxford, Imperial College of London and University College of London) or from the US (Harvard, California Inst Tech, California – Berkeley, MIT, Stanford, Columbia, Princeton, Chicago, Yale, Cornell, Pennsylvania, Johns Hopkins, California - Los Angeles and Wisconsin - Madison). It is interesting to note that some of those universities were not included among the Top20 universities in either THES or the SJTU. To be more specific, the university of California-Berkley was ranked 21st in THES, but according to the 130 simulations that employ the full set of twelve indicators, this university is indeed a Top5 university (72% frequency). Similarly, the university of California-Los Angeles and the university of Wisconsin-Madison were ranked below the 30th position in the THES, but they should be considered as Top20 universities, accordingly to our simulations. Also, the Imperial College of London and the University College of London which were ranked 22nd and 24th in the SJTU, they should be considered Top20 universities based on the simulations. The Swiss Fed Inst Tech of Zurich was ranked 26th in the SJTU and only 40th in the THES, but when considering both sets of indicators and the plurality of methodologies in combining the underlying information, this university is definitely a Top20 university.

Table 22. Frequency matrix of university ranks - hybrid THES & SJTU

THES & SJTU (130 scenarios)

	Rank 1-5	Rank 6-10	Rank 11-15	Rank 16-20	Rank 21-25	Rank 26-30	Rank 31-35	Rank 36-40	Rank 41-45	Rank 46-50	Rank 51-55	Rank 56-60	Rank 61-65	Rank 66-70	Rank 71-75	Rank 76-80	Rank 81-88	Median rank	Common THES rank	Common SJTU rank
Harvard Univ	100																	1	1	1
Univ Cambridge	98																	2	4	4
California Inst Tech	59	41																4	7	6
Univ California - Berkeley	72	28																4	21	3
Massachusetts Inst Tech (MIT)	81	19																5	10	5
Stanford Univ	45	43	12															6	18	2
Columbia Univ		90	8															8	11	7
Univ Oxford	22	76																8	2	10
Princeton Univ	7	78	15															8	5	8
Univ Chicago		72	28															9	8	9
Yale Univ	18	42	41															10	3	11
Cornell Univ		6	94															12	20	12
Univ Pennsylvania			100															13	13	15
Imperial Coll London			79	20														14	6	22
Univ Coll London			75	24														15	9	24
Johns Hopkins Univ			22	75														17	15	18
Univ California - Los Angeles			15	70	7	8												18	33	13
Swiss Fed Inst Tech - Zurich			8	49	21	18	5											19	34	26
Tokyo Univ				66	28													19	17	19
Univ Wisconsin - Madison				50	35	8	6											20	40	17
Univ Toronto				28	49	20												22	35	23
Kyoto Univ				22	47	22	9											23	24	21
Univ Michigan - Ann Arbor				14	42	5			8	11	7		8					24	32	20
Univ California - San Diego				13	42	17	12	8	7									25	42	14
Duke Univ				22	35	7				6	7	8	8	7				25	13	30
Univ Washington - Seattle				17	42	29	6	6										25	41	16
Northwestern Univ					44	54												26	27	28
Univ Illinois - Urbana Champaign					22	35	23	10	5	5								29	50	25
McGill Univ				11	16	30	6			7		6	16	5				29	12	55
Carnegie Mellon Univ				9	15	32	35	8										30	19	53
Univ Manchester				5	20	28	37	10										30	28	44
Australian Natl Univ					14	32	42	8										31	16	50
Univ British Columbia						40	48	10										31	30	34
Univ Edinburgh					11	37	42	10										31	22	47
New York Univ					5	27	40	20	7									33	38	29
Univ Bristol						6	27	42	18									37	31	56
Brown Univ							10	53	28	8								40	28	63
Ecole Normale Super Paris						5	18	34	22	12	5							40	25	77
Tech Univ Munich						12	17	22	18	13	5							40	45	49
Univ Texas - Austin							12	36	36	12								40	39	36
King's Coll London						5	14	27	21	12	12							41	23	76
Univ Maryland - Coll Park							29	17	28	19	5							41	54	35
Vanderbilt Univ							8	26	32	22	7							42	53	38
Univ Melbourne							9	28	31	19	8							43	26	71
Univ Munich							10	22	28	26	11							44	44	48
Univ Sheffield								26	22	20	14	7		5				46	47	65
Univ Heidelberg								12	25	27	20	12						48	43	58
Washington Univ - St. Louis						8	12	16	8	5		6	8	16	8	7		49	84	27
Univ Nottingham								8	21	22	13	12	8		7	5		50	49	73
Univ Colorado - Boulder								9	22	16	22	13	8	5				51	66	32
Univ Pittsburgh - Pittsburgh								5	11	25	12	6	6	22	10			53	51	42
Univ Leiden									8	26	24	28	11					55	56	64
Purdue Univ - West Lafayette									5	15	32	23	14	5				55	51	61
Univ Paris 06								7	10	12	19	16	22	7				55	74	37
Univ Birmingham									12	18	18	19	12		8	9		56	45	83
Case Western Reserve Univ										6	36	25	12	5	8	5		57	55	70
Univ Basel										16	22	14	11	11	6	8	8	58	68	74
Univ Helsinki										5	22	35	25	8	5			59	62	67
Univ Rochester										5	12	40	23	11	6			59	60	68
Univ Zurich										14	10	34	18	6	8			59	80	52
Uppsala Univ										12	17	16	17	16	12	7		61	48	59
Boston Univ									7	18	15	8	8	15	20	8		62	36	75
Univ California - Santa Barbara								10	8	8	12	8	9	18	12	10	5	62	69	33
Osaka Univ								5	12	5	9	10	12	12	10	14	10	64	37	60
Rice Univ										9	21	7	18	22	14	5		64	59	80
Univ Southern California											5	17	27	32	11			65	71	45
Univ Minnesota - Twin Cities								5	9	11	8	9	8	20	22	5		65	79	31
Hebrew Univ Jerusalem												13	28	20	18	9	6	66	73	57
Pennsylvania State Univ - Univ Park										8	14	8	9	19	21	12	5	68	57	40
Rutgers State Univ - New Brunswick											10	13	12	14	10	20	19	70	86	42
Ohio State Univ - Columbus											5	7	9	32	24	20		70	70	54
Univ California - Davis									5	10	8	8	6	11	8	6	37	71	61	39
Univ California - Irvine											8	8	10	23	15	15	20	71	77	41
Tokyo Inst Tech													9	32	25	22	12	72	58	86
Univ Florida												5	8	22	27	30	7	73	75	46
Tohoku Univ												5	8	28	22	20	15	73	63	69
McMaster Univ													11	25	12	10	39	75	64	81
Univ Strasbourg 1												7	13	9	14	15	38	77	87	88
Lund Univ														15	19	31	31	78	64	86
Univ Oslo											9	12	13		7	16	39	78	88	62
Michigan State Univ														13	18	23	42	79	83	72
Indiana Univ - Bloomington												7	13	18		12	48	79	78	82
Univ North Carolina - Chapel Hill													7	9	15	22	48	80	82	51
Univ Freiburg												6	13	15	5	12	48	80	81	85
Univ Arizona													5	6	15	21	52	81	76	66
Nagoya Univ														5	5	22	65	83	67	84
Texas A&M Univ - Coll Station																18	77	84	72	78
Univ Goettingen																13	78	85	85	79

Legend:
Frequency lower 15%
Frequency between 15 and 30%
Frequency between 30 and 50%
Frequency greater than 50%
Note: Frequencies lower than 4% are not shown

4.4.2 Sensitivity of university ranks to the methodological assumptions

We next discuss the degree of sensitivity of university ranks, as we did previously when dealing with either the THES or the SJTU frameworks. In this case, the hybrid framework and the relevant simulation are used. There are 22 universities whose rank has low degree of sensitivity to the methodological assumptions, 16 universities with medium sensitivity and 50 universities that are highly sensitive.

Table 23 lists the universities whose simulated rank (90% confidence interval) is highly sensitive and thus no meaningful conclusions can be drawn. These universities were ranked below the 10th position under either THES or SJTU. For example, the McGill University in Canada, which performs very well according to the THES Index (rank 12) and has a mid performance in the SJTU Index (rank 55) is expected to perform anywhere between the 18th and the 66th position, when the methodological assumptions in the hybrid Index are challenged. Three French Universities, the Ecole Normale Super of Paris and the University of Paris 06 and the University of Paris 1, are included in this list. The Ecole Normale Super of Paris is ranked 25th in the THES framework and in a much lower position in the SJTU framework (77th), which is reflected in the impact on the simulated ranks ranging between 30th and 57th position. Note that the common THES or SJTU ranks are outside the expected range when acknowledging the uncertainties in the twelve indicator framework. This result provides a further confirmation that both the THES and the SJTU ranks for that university were biased versus the indicators used in the respective frameworks. Similarly, the University of Paris 06 is ranked 74th in the THES framework and in a much higher position in the SJTU framework (37th), which is also reflected in the impact on the simulated ranks ranging between 35th and 71st position. All countries included in the analysis have at least one university with high sensitivity to the methodological assumptions even when the hybrid framework of indicators is used.

Table 23. Hybrid THES&SJTU Index: 50 universities whose simulated rank is highly sensitive to the conceptual and methodological assumptions

University	Country	Median rank	THES rank	SJTU rank	Range of ranks	
Univ Melbourne	Australia	43	26	71	[31	60]
McGill Univ	Canada	29	12	55	[18	66]
McMaster Univ	Canada	75	64	81	[60	88]
Univ Helsinki	Finland	59	62	67	[49	73]
Ecole Normale Super Paris	France	40	25	77	[30	57]
Univ Paris 06	France	55	74	37	[35	71]
Univ Strasbourg 1	France	77	87	88	[54	88]
Tech Univ Munich	Germany	40	45	49	[25	63]
Univ Heidelberg	Germany	48	43	58	[37	62]
Univ Freiburg	Germany	80	81	85	[59	88]
Univ Goettingen	Germany	85	85	79	[66	88]
Hebrew Univ Jerusalem	Israel	66	73	57	[53	84]
Osaka Univ	Japan	64	37	60	[39	82]
Tohoku Univ	Japan	73	63	69	[57	83]
Nagoya Univ	Japan	83	67	84	[66	88]
Univ Leiden	Netherlands	55	56	64	[44	66]
Univ Oslo	Norway	78	88	62	[51	86]
Uppsala Univ	Sweden	61	48	59	[46	79]
Univ Basel	Switzerland	58	68	74	[45	83]
Univ Zurich	Switzerland	59	80	52	[45	79]
King's Coll London	UK	41	23	76	[25	63]
Univ Sheffield	UK	46	47	65	[36	69]
Univ Nottingham	UK	50	49	73	[37	78]
Univ Birmingham	UK	56	45	83	[41	80]
Univ Michigan - Ann Arbor	US	24	32	20	[18	65]
Univ California - San Diego	US	25	42	14	[18	45]
Duke Univ	US	25	13	30	[18	67]
Univ Illinois - Urbana Champaign	US	29	50	25	[21	46]
Vanderbilt Univ	US	42	53	38	[32	55]
Washington Univ - St. Louis	US	49	84	27	[28	80]
Univ Colorado - Boulder	US	51	66	32	[35	69]
Univ Pittsburgh - Pittsburgh	US	53	51	42	[38	73]
Purdue Univ - West Lafayette	US	55	51	61	[43	72]
Case Western Reserve Univ	US	57	55	70	[46	79]
Univ Rochester	US	59	60	68	[49	76]
Boston Univ	US	62	36	75	[43	78]
Univ California - Santa Barbara	US	62	69	33	[37	81]
Rice Univ	US	64	59	80	[45	77]
Univ Southern California	US	65	71	45	[45	78]
Univ Minnesota - Twin Cities	US	65	79	31	[39	77]
Pennsylvania State Univ - Univ Park	US	68	57	40	[45	81]
Rutgers State Univ - New Brunswick	US	70	86	42	[52	86]
Ohio State Univ - Columbus	US	70	70	54	[54	81]
Univ California - Davis	US	71	61	39	[44	86]
Univ California - Irvine	US	71	77	41	[52	87]
Univ Florida	US	73	75	46	[57	82]
Michigan State Univ	US	79	83	72	[65	87]
Indiana Univ - Bloomington	US	79	78	82	[60	87]
Univ North Carolina - Chapel Hill	US	80	82	51	[62	85]
Univ Arizona	US	81	76	66	[63	87]

4.4.3 Non-representative university ranks of the plurality of the scenarios

According to the criterion we set previously, we will discuss non-representative rankings (biased rankings) in either the THES or the SJTU. In this discussion, the bias will be assessed towards the hybrid framework. Recall that in Sections 4.2.3 and 4.3.3 the bias was assessed towards either the THES or the SJTU frameworks respectively.

There is a fair degree of correlation between the common THES rank and the simulated median, $r^2 = .710$ ($p < 0.001$, $n = 88$), which produces a high degree of

confidence that for several universities, the THES rank is reliable and no deliberate bias was introduced in the THES. However, caution is needed in the case of few universities whose performance based on the THES deviates from the simulated median. According to the criterion we set earlier, the THES rank for 26 universities is likely to be unreliable or non-representative. For the remaining universities, we can have high confidence that they are roughly placed, on average, in the correct place and this classification may be used for policy-making or for benchmarking purposes. However, the degree of volatility discussed previously needs to be taken into consideration prior to drawing final conclusions.

There is a low degree of correlation between the common SJTU rank and the simulated median, $r^2 = .484$ ($p < 0.001$, $n = 88$), which casts doubts on the ranking of several universities. In fact, there are 33 universities whose SJTU rank is very likely to be unreliable or non-representative given the set of twelve indicators and the plurality of the methodological assumptions. For the remaining universities, we can have confidence that they are roughly placed, on average, in the correct place and this classification may be used for policy-making or for benchmarking purposes. Again, the degree of volatility discussed previously needs to be taken into consideration prior to drawing final conclusions.

Table 24 lists the universities whose simulated median rank differs by more than 13 positions from the THES rank (26 universities) or from the SJTU rank (33 universities). For all those universities, neither the THES rank nor the SJTU can be reliably used to classify the universities in an ordinal scale.

The 26 universities whose THES rank is biased with respect to the THES set of indicators are ranked between the 12th position (McGill University in Canada) and the 86th position (Rutgers State University - New Brunswick) in the THES index. This list includes two Australian universities (Australian National University, University of Melbourne), two Canadian universities (University of Toronto, McGill University), two French universities (Ecole Normale Super Paris, University of Paris 06), University of Strasbourg), three Japanese universities (Osaka, Tokyo Inst Tech, Nagoya), two Swedish universities (Uppsala, Lund), two Swiss universities (Swiss Fed Inst Tech – Zurich, University of Zurich), one university in the UK (King's College of London) and twelve universities in the US (California – Berkeley, California - Los Angeles, Wisconsin – Madison, California - San Diego, Washington–Seattle, Illinois-Urbana Champaign, Maryland-College Park, Washington - St. Louis,

Colorado – Boulder, Boston, Minnesota - Twin Cities, and Rutgers State - New Brunswick).

Table 24. Hybrid Index: universities whose THES or SJTU rank is non-representative of the simulated scenarios

University	*Country*	*THES rank*	*Median rank*
Australian Natl Univ	Australia	16	31
Univ Melbourne	Australia	26	43
Univ Toronto	Canada	35	22
McGill Univ	Canada	12	29
Ecole Normale Super Paris	France	25	40
Univ Paris 06	France	74	55
Osaka Univ	Japan	37	64
Tokyo Inst Tech	Japan	58	72
Nagoya Univ	Japan	67	83
Uppsala Univ	Sweden	48	61
Lund Univ	Sweden	64	78
Swiss Fed Inst Tech - Zurich	Switzerland	34	19
Univ Zurich	Switzerland	80	59
King's Coll London	UK	23	41
Univ California - Berkeley	US	21	4
Univ California - Los Angeles	US	33	18
Univ Wisconsin - Madison	US	40	20
Univ California - San Diego	US	42	25
Univ Washington - Seattle	US	41	25
Univ Illinois - Urbana Champaign	US	50	29
Univ Maryland - Coll Park	US	54	41
Washington Univ - St. Louis	US	84	49
Univ Colorado - Boulder	US	66	51
Boston Univ	US	36	62
Univ Minnesota - Twin Cities	US	79	65
Rutgers State Univ - New Brunswick	US	86	70
University	*Country*	*SJTU rank*	*Median rank*
Australian Natl Univ	Australia	50	31
Univ Melbourne	Australia	71	43
McGill Univ	Canada	55	29
Ecole Normale Super Paris	France	77	40
Univ Paris 06	France	37	55
Tokyo Inst Tech	Japan	86	72
Univ Oslo	Norway	62	78
Univ Basel	Switzerland	74	58
Univ Manchester	UK	44	30
Univ Edinburgh	UK	47	31
Univ Bristol	UK	56	37
King's Coll London	UK	76	41
Univ Sheffield	UK	65	46
Univ Nottingham	UK	73	50
Univ Birmingham	UK	83	56
Carnegie Mellon Univ	US	53	30
Brown Univ	US	63	40
Washington Univ - St. Louis	US	27	49
Univ Colorado - Boulder	US	32	51
Case Western Reserve Univ	US	70	57
Boston Univ	US	75	62
Univ California - Santa Barbara	US	33	62
Rice Univ	US	80	64
Univ Southern California	US	45	65
Univ Minnesota - Twin Cities	US	31	65
Pennsylvania State Univ - Univ Park	US	40	68
Rutgers State Univ - New Brunswick	US	42	70
Ohio State Univ - Columbus	US	54	70
Univ California - Davis	US	39	71
Univ California - Irvine	US	41	71
Univ Florida	US	46	73
Univ North Carolina - Chapel Hill	US	51	80
Univ Arizona	US	66	81

The 33 universities whose SJTU rank is biased with respect to the SJTU set of indicators are ranked between the 27th position (Washington - St. Louis) and the 86th position (Tokyo Inst. Tech) in the SJTU index. Several universities identified previously as having a biased THES rank are also having a biased SJTU rank. Additionally, there are universities from the UK (e.g., Manchester, Edinburgh, Bristol) and the US (Boston, California – Davis, California – Irvine, Florida).

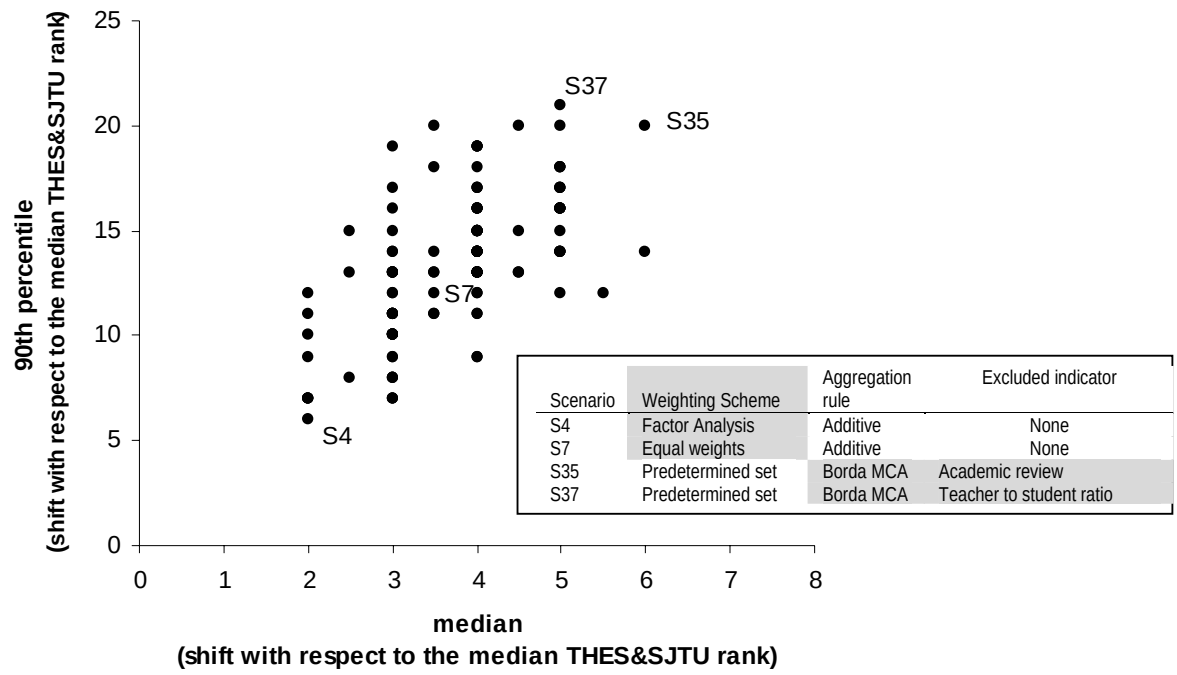
4.2.4 Impact of the assumptions on the university ranks

We study the impact of the assumptions on the university rank compared to the simulated median rank across the 130 simulations and we use for each scenario the two sensitivity measures described previously. Figure 4 plots the two sensitivity measures.

The more distant from the origin a scenario is, the more it deviates from the simulated median ranking. All 130 scenarios have a first sensitivity measure value in the range 2.0 to 6.0 and a second sensitivity measure value in the range 6.0 to just above 20.0. Detailed results per scenario are provided in the Appendix (Table A 5). These values for the two sensitivity measures are much lower than those identified in the previous sections regarding either the THES or the SJTU. This implies that the impact of the methodological assumptions is much lower when the set of twelve indicators is used.

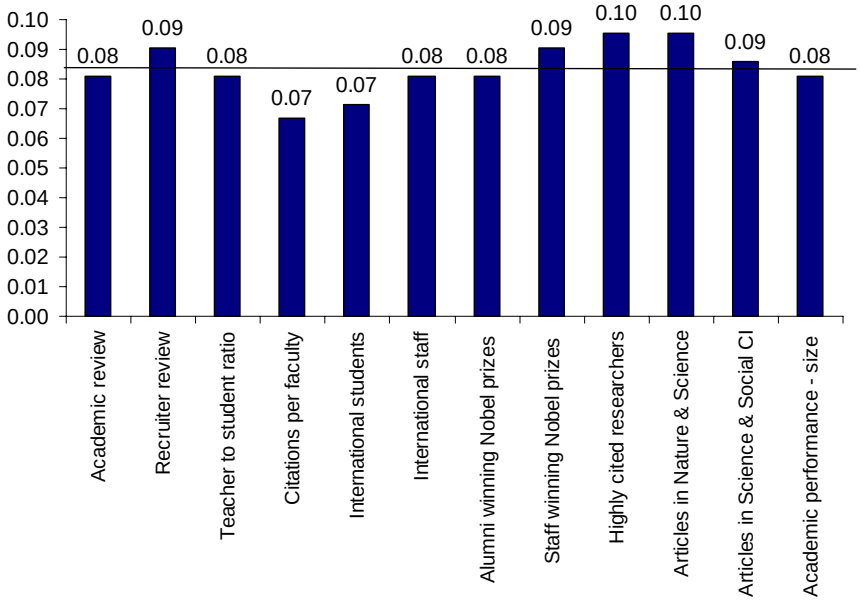
The scenario whose resulting ranking resembles more the simulated median ranking, and it could thus be considered as a representative picture of the classification of the universities performance, employs Factor Analysis to determine the weights for the twelve indicators and a simple weighted average of the indicators (Scenario 2). In fact, in this scenario, half of the universities shift less than 2 positions with respect to the median rank and the 10% most affected universities shift 6.0 positions or more (yet always less than 12 positions). Figure 5 shows the recommended set of weights that derived from factor analysis. An equal weighting scheme would imply assigning 1/12 (=0.0833) to each indicator (Scenario 7). This scenario would have resulted in half of the universities shifting less than 3.5 positions with respect to the median rank, but the 10% most affected universities would shift 12.0 positions or more (worst case: 16 positions).

Figure 4. Sensitivity analysis: impact of the assumptions on the THES ranks



Note: median versus 90th percentile of the absolute differences between a simulated ranking in a given scenario and the median ranking (over 88 universities). For detailed results, see Appendix, Table A5)

Figure 5. Recommended set of weights to be used when combining the twelve indicators of the THES and the SJTU



These results show that an assessment of the universities performance based on the hybrid set of twelve indicators and employing factor analysis to determine the weights can provide a reliable, average, rank for the universities studied.

One issue remains though. The high volatility of more than half of the universities (see results discussed previously in Table 23). This calls for a revision of the set of indicators, either by enriching it with other dimensions that are crucial to assessing university performance or by revising some of the existing indicators in order to remove some bias (e.g., in favour of old and/or big universities). A legitimate question is raised: when will the revision of the dataset of indicators reach a satisfactory level? We would argue that the set of indicators to be used for the classification of the universities will be satisfactory once, upon acknowledging the methodological uncertainties that are intrinsic to the development of a ranking system, the space of inference of the ranks for the majority of the universities is narrow enough to draw meaningful conclusions.

5. Can university rankings be used as a guide for higher education policies?

A nasty question at this point might be: *is all this effort we have done of any use?* Even if we have identified the universities for which the THES or the SJTU are reliable and the cases for which these ranks should be treated with caution, what is the policy utility of knowing that a university is overall better than another university within the same country or internationally? This kind of criticism is often put to composite indicators and ranking systems, and thus it is worthy to be discussed further.

Indeed we have already seen that ranks can provide interesting pieces of information for policy purposes. For example, in the hybrid THES and SJTU framework, we succeeded in finding out clear success stories. However, for more than half of the universities, the range of ranks was too wide to draw any meaningful conclusions. Thus, for the time being, and until a new set of indicators is offered in the relevant literature, one needs to consult Table 25. This table summarises the main conclusions on the reliability of the ranks as discussed in the previous sections. The information provided is grouped in three main categories:

(1) the hybrid THES&SJTU rank is highly sensitive to the methodological assumptions (this is the case for 50 universities),

(2) the THES rank is biased towards the set of the indicators, the set of weights and the aggregation rule (this is the case for 26 universities), and

(3) the SJTU rank is biased towards the set of the indicators, the set of weights and the aggregation rule (this is the case for 33 universities).

Table 25. Conclusions from the robustness assessment of the university ranks

THES & SJTU (130 scenarios)

	Rank 1-5	Rank 6-10	Rank 11-15	Rank 16-20	Rank 21-25	Rank 26-30	Rank 31-35	Rank 36-40	Rank 41-45	Rank 46-50	Rank 51-55	Rank 56-60	Rank 61-65	Rank 66-70	Rank 71-75	Rank 76-80	Rank 81-88	(1) -THES &SJTU	(2) - THES	(3) - SJTU	Median rank
Harvard Univ	100																				1
Univ Cambridge	98																				2
California Inst Tech	59	41																			4
Univ California - Berkeley	72	28																	X		4
Massachusetts Inst Tech (MIT)	81	19																			5
Stanford Univ	45	43	12																		6
Columbia Univ		90	8																		8
Univ Oxford	22	76																			8
Princeton Univ	7	78	15																		8
Univ Chicago		72	28																		9
Yale Univ	18	42	41																		10
Cornell Univ		6	94																		12
Univ Pennsylvania			100																		13
Imperial Coll London			79	20																	14
Univ Coll London			75	24																	15
Johns Hopkins Univ			22	75																	17
Univ California - Los Angeles			15	70	7	8													X		18
Swiss Fed Inst Tech - Zurich			8	49	21	18	5												X		19
Tokyo Univ				66	28																19
Univ Wisconsin - Madison				50	35	8	6												X		20
Univ Toronto				28	49	20													X		22
Kyoto Univ				22	47	22	9														23
Univ Michigan - Ann Arbor				14	42	5			8	11	7		8					X			24
Univ California - San Diego				13	42	17	12	8	7									X	X		25
Duke Univ				22	35	7				6	7	8	8	7				X			25
Univ Washington - Seattle				17	42	29	6	6											X		25
Northwestern Univ					44	54															26
Univ Illinois - Urbana Champaign					22	35	23	10	5	5								X	X		29
McGill Univ				11	16	30	6			7		6	16	5				X	X	X	29
Carnegie Mellon Univ				9	15	32	35	8												X	30
Univ Manchester				5	20	28	37	10												X	30
Australian Natl Univ					14	32	42	8											X	X	31
Univ British Columbia						40	48	10													31
Univ Edinburgh					11	37	42	10												X	31
New York Univ					5	27	40	20	7												33
Univ Bristol						6	27	42	18											X	37
Brown Univ							10	53	28	8										X	40
Ecole Normale Super Paris						5	18	34	22	12	5							X	X	X	40
Tech Univ Munich						12	17	22	18	13	5							X			40
Univ Texas - Austin							12	36	36	12											40
King's Coll London						5	14	27	21	12	12							X	X	X	41
Univ Maryland - Coll Park							29	17	28	19	5								X		41
Vanderbilt Univ							8	26	32	22	7							X			42
Univ Melbourne							9	28	31	19	8							X	X	X	43
Univ Munich							10	22	28	26	11										44
Univ Sheffield								26	22	20	14	7		5				X		X	46
Univ Heidelberg								12	25	27	20	12						X			48
Washington Univ - St. Louis							8	12	16	8	5		6	8	16	8	7	X	X	X	49
Univ Nottingham								8	21	22	13	12	8		7	5		X		X	50
Univ Colorado - Boulder								9	22	16	22	13	8	5				X	X	X	51
Univ Pittsburgh - Pittsburgh								5	11	25	12	6	6	22	10			X			53
Univ Leiden									8	26	24	28	11					X			55
Purdue Univ - West Lafayette									5	15	32	23	14	5				X			55
Univ Paris 06								7	10	12	19	16	22	7				X	X	X	55
Univ Birmingham									12	18	18	19	12		8	9		X		X	56
Case Western Reserve Univ										6	36	25	12	5	8	5		X		X	57
Univ Basel										16	22	14	11	11	6	8	8	X		X	58
Univ Helsinki										5	22	35	25	8	5			X			59
Univ Rochester										5	12	40	23	11	6			X			59
Univ Zurich										14	10	34	18	6	8			X	X		59
Uppsala Univ										12	17	16	17	16	12	7		X	X		61
Boston Univ									7	18	15	8	8	15	20	8		X	X	X	62
Univ California - Santa Barbara								10	8	8	12	8	9	18	12	10	5	X		X	62
Osaka Univ								5	12	5	9	10	12	12	10	14	10	X	X		64
Rice Univ										9	21	7	18	22	14	5		X		X	64
Univ Southern California											5	17	27	32	11			X		X	65
Univ Minnesota - Twin Cities								5	9	11	8	9	8	20	22	5		X	X	X	65
Hebrew Univ Jerusalem												13	28	20	18	9	6	X			66
Pennsylvania State Univ - Univ Park										8	14	8	9	19	21	12	5	X		X	68
Rutgers State Univ - New Brunswick											10	13	12	14	10	20	19	X	X	X	70
Ohio State Univ - Columbus											5	7	9	32	24	20		X		X	70
Univ California - Davis									5	10	8	8	6	11	8	6	37	X		X	71
Univ California - Irvine											8	8	10	23	15	15	20	X		X	71
Tokyo Inst Tech													9	32	25	22	12		X	X	72
Univ Florida												5	8	22	27	30	7	X		X	73
Tohoku Univ												5	8	28	22	20	15	X			73
McMaster Univ													11	25	12	10	39	X			75
Univ Strasbourg 1												7	13	9	14	15	38	X			77
Lund Univ														15	19	31	31		X		78
Univ Oslo											9	12	13		7	16	39	X		X	78
Michigan State Univ														13	18	23	42	X			79
Indiana Univ - Bloomington												7	13	18		12	48	X			79
Univ North Carolina - Chapel Hill													7	9	15	22	48	X		X	80
Univ Freiburg												6	13	15	5	12	48	X			80
Univ Arizona													5	6	15	21	52	X		X	81
Nagoya Univ														5	5	22	65	X	X		83
Texas A&M Univ - Coll Station																18	77				84
Univ Goettingen																13	78	X			85

Legend:
Frequency lower 15%
Frequency between 15 and 30%
Frequency between 30 and 50%
Frequency greater than 50%
Note: Frequencies lower than 4% are not shown

(1) the hybrid THES&SJTU rank is highly sensitive to the methodological assumptions, (2) the THES rank is biased towards the set of the indicators, the set of weights and the aggregation rule, and (3) the SJTU rank is biased towards the set of the indicators, the set of weights and the aggregation rule.

These rankings and the accompanying frequencies in Table 25 aim at answering arguments, well pointed out by Andrews *et al.* (2004: 1323), that many indices "rarely have adequate scientific foundations to support precise rankings: […] typical practice is to acknowledge uncertainty in the text of the report and then to present a table with unambiguous rankings".

The rankings together with the relevant frequencies are only meant to complement the information provided by the underlying indicators. For the majority of indicators studied no clear reference point (in the sense of target) is available. For instance, the indicator *teacher to student ratio* depends strongly on the field of study and there is no established benchmark. Thus, it is quite common to compare performance with other universities either at a national or international level. In order to get a set of reference values to be used as benchmarks for the twelve indicators we have opted for two options:

- To compare the performance of the European universities with the average European university (in our case the average of the 27 European universities included in the dataset we analysed).
- To compare the performance of a university (both European and non-European) with the average US university (in our case the average of the 48 US universities included in the dataset we analysed).

Table 26 provides both synthetic and analytic information on university performance within Europe. The table provides the total number of indicators that are above (+), close (0) or below (-) the European average, thus allowing for a quick assessment of the universities. At the same time, it is also possible to derive policy suggestions since all the indicators are scored. Note that in this comparison, the normalisation, weighting, or aggregation issues are not relevant. The conclusions are drawn exclusively on the comparison of the universities scores with respect to the average university performance within Europe.

Table 26. Benchmarking performance of European universities along the THES and SJTU indicators with respect to the average European University

		Comparison with the European average													
		THES indicators						SJTU indicators							
University	country	Academic review	Recruiter review	Teacher to student ratio	Citations per faculty	International students	International staff	Alumni winning Nobel prizes	Staff winning Nobel prizes	Highly cited researchers	Articles in Nature & Science	Articles in Science & Social Sci	Academic performance - size	Above	Close to	Below
Univ Helsinki	Finland	0	-	-	+	-	0	-	-	-	-	+	0	2	3	7
Ecole Normale Super Paris	France	+	-	+	+	-	0	+	+	-	-	-	-	5	1	6
Univ Strasbourg 1	France	-	0	-	0	-	0	0	-	-	-	-	-	0	4	8
Univ Paris 06	France	-	-	+	0	-	+	+	-	0	0	+	0	4	4	4
Univ Heidelberg	Germany	0	0	-	0	-	+	-	0	-	-	0	0	1	6	5
Univ Munich	Germany	0	-	0	0	-	0	+	-	-	0	0	0	1	7	4
Tech Univ Munich	Germany	-	0	+	0	-	+	+	-	0	-	0	0	3	5	4
Univ Freiburg	Germany	-	-	+	-	-	0	-	-	-	-	-	-	1	1	10
Univ Goettingen	Germany	-	-	0	-	-	-	+	-	-	-	-	-	1	1	10
Univ Leiden	Netherlands	0	0	-	+	0	-	-	-	+	-	0	0	2	5	5
Univ Oslo	Norway	-	-	-	-	-	-	-	+	-	-	-	-	1	0	11
Lund Univ	Sweden	0	-	-	0	0	-	0	-	0	-	0	-	0	6	6
Uppsala Univ	Sweden	0	-	+	0	-	-	-	+	-	-	0	-	2	3	7
Univ Basel	Switzerland	-	-	+	-	+	0	-	-	-	-	-	+	3	1	8
Swiss Fed Inst Tech - Zurich	Switzerland	+	+	-	0	+	+	+	+	+	+	-	+	9	1	2
Univ Zurich	Switzerland	0	-	-	+	+	-	-	0	-	0	0	0	2	5	5
Univ Birmingham	UK	0	+	-	0	+	0	-	-	-	-	0	-	2	4	6
Univ Bristol	UK	0	+	+	0	+	0	-	-	+	0	0	0	4	6	2
Univ Cambridge	UK	+	+	+	0	+	+	+	+	+	+	+	+	11	1	0
Univ Edinburgh	UK	+	+	+	0	0	0	-	-	0	+	0	0	4	6	2
Imperial Coll London	UK	+	+	+	0	+	+	-	+	+	+	+	+	10	1	1
King's Coll London	UK	+	+	+	0	+	+	-	-	-	-	0	-	5	2	5
Univ Manchester	UK	+	+	0	0	+	+	-	-	0	0	+	0	5	5	2
Univ Nottingham	UK	-	+	-	-	+	+	-	-	0	-	0	-	3	2	7
Univ Oxford	UK	+	+	+	0	+	+	+	+	+	+	+	+	11	1	0
Univ Sheffield	UK	-	+	0	0	+	-	-	-	-	0	0	0	2	5	5
Univ Coll London	UK	+	+	+	0	+	+	0	+	+	+	+	0	9	3	0
	Above	9	12	13	4	13	11	8	8	7	6	7	5			
	Close to	9	4	4	18	3	9	3	2	6	6	13	12			
	Below	9	11	10	5	11	7	16	17	14	15	7	10			

Notes:

1. 27 European Universities were considered that are common to the THES and the SJTU rankings.
2. Value "+" indicates that the University score is above the European average by at least 10%; Value "0" indicates that the University score is near the European average (within ± 10%); Value "-" indicates that the University score is below the European average by at least 10%.

To give some illustrative examples, the University of Cambridge, the Imperial College of London and the University of Oxford (all three in the UK) have at least ten indicators above the European average and at least one indicator close to the European average. Of these three universities, only the Imperial college of London has one indicator below the European average, namely the SJTU indicator on *alumni winning Nobel prizes and field medals*. This is the only plausible policy priority for the Imperial College of London, provided that one accepts the importance of this

indicator in the assessment of a university. On the other side, if one examines the performance of the University of Freiburg and the University of Goettingen in Germany and the University of Oslo in Norway, where at least ten indicators are below the European average, it is clear that the space for improvement is enormous. But of course this does not necessarily mean that the respective governments need to pursue the objective of rendering these universities top ranked. However, it flags that there is space for improvement and that these indicators where the universities under-perform[21] need to receive some policy attention.

Table 27 provides both synthetic and analytic information on university performance with respect to the average US University. The table also provides the total number of indicators that are above (+), close (0) or below (-) the US average, thus allowing for a quick assessment of all universities. Only two Universities in Europe, namely the University of Cambridge and the University of Oxford (both in the UK) have at least ten indicators above the US average. On the other hand, the University of McMaster in Canada, the University of Freiburg in Germany, and the Tokyo Inst Tech in Japan have at least ten indicators below the US average. The policy implications of these results are evident.

[21] Performance is discussed in a relative sense since these universities are all in the top 200 world-wide based on either the THES or the SJTU frameworks. Note further that the SJTU index tends to undervalue institutions where a great deal of academic scientific research takes place outside universities (e.g., the Max Planck Institutes in Germany) or in organisations whose researchers are affiliated to several universities (e.g., the CNRS laboratories in France). This partly explains why several European universities and institutions are not included in the Top 100 in the list and therefore were not included in our analysis.

Table 27. Benchmarking universities scores in the six THES and six SJTU indicators with respect to the average US University

		Comparison with the US average														
		THES indicators						SJTU indicators								
University	country	Academic review	Recruiter review	Teacher to student ratio	Citations per faculty	International students	International staff	Alumni winning Nobel prizes	Staff winning Nobel prizes	Highly cited researchers	Articles in Nature & Science	Articles in Science & Social Sci	Academic performance - size	Above	Close to	Below
Australian Natl Univ	Australia	+	+	+	-	+	+	-	-	-	-	-	0	5	1	6
Univ Melbourne	Australia	+	+	0	-	+	+	-	-	-	-	0	-	4	2	6
Univ British Columbia	Canada	+	+	0	-	-	+	-	-	-	-	0	0	3	3	6
McGill Univ	Canada	+	+	+	-	+	+	0	-	-	-	0	0	5	3	4
McMaster Univ	Canada	0	-	-	0	-	-	-	-	-	-	-	-	0	2	10
Univ Toronto	Canada	+	+	-	0	+	-	-	-	-	0	+	+	5	2	5
Univ Helsinki	Finland	0	-	-	0	+	+	-	-	-	-	0	-	2	3	7
Ecole Normale Super Paris	France	0	-	+	+	+	+	+	+	-	-	-	-	6	1	5
Univ Strasbourg 1	France	-	-	-	0	0	+	0	-	-	-	-	-	1	3	8
Univ Paris 06	France	-	-	+	-	-	+	+	-	-	-	0	0	3	2	7
Univ Heidelberg	Germany	0	-	0	0	0	+	-	0	-	-	-	-	1	5	6
Univ Munich	Germany	0	-	0	-	+	+	+	-	-	-	0	-	3	3	6
Tech Univ Munich	Germany	-	0	+	-	+	+	+	-	-	-	-	-	4	1	7
Univ Freiburg	Germany	-	-	+	-	-	+	-	-	-	-	-	-	2	0	10
Univ Goettingen	Germany	-	-	+	-	0	0	+	-	-	-	-	-	2	2	8
Hebrew Univ Jerusalem	Israel	0	-	-	0	+	-	0	-	-	-	-	-	1	3	8
Kyoto Univ	Japan	+	+	+	0	-	-	+	+	-	0	+	-	6	2	4
Nagoya Univ	Japan	-	0	+	0	-	-	-	-	-	-	-	-	1	2	9
Osaka Univ	Japan	0	0	+	0	-	-	-	-	-	-	+	-	2	3	7
Tohoku Univ	Japan	-	-	+	0	-	-	-	-	-	-	0	-	1	2	9
Tokyo Inst Tech	Japan	-	+	-	0	-	-	-	-	-	-	-	-	1	1	10
Tokyo Univ	Japan	+	+	+	0	-	-	+	-	0	+	+	0	6	3	3
Univ Leiden	Netherlands	0	-	-	0	+	-	-	-	-	-	-	-	1	2	9
Univ Oslo	Norway	-	-	-	-	+	0	-	+	-	-	-	-	2	1	9
Lund Univ	Sweden	0	-	-	0	+	0	0	-	-	-	-	-	1	4	7
Uppsala Univ	Sweden	0	-	+	-	+	-	-	+	-	-	-	-	3	1	8
Univ Basel	Switzerland	-	-	+	-	+	+	-	-	-	-	-	0	3	1	8
Swiss Fed Inst Tech - Zurich	Switzerland	+	0	0	-	+	+	+	+	-	0	-	+	6	3	3
Univ Zurich	Switzerland	-	-	-	+	+	0	-	0	-	-	-	-	2	2	8
Univ Birmingham	UK	-	+	0	-	+	+	-	-	-	-	-	-	3	1	8
Univ Bristol	UK	0	+	+	-	+	+	-	-	-	-	-	0	4	2	6
Univ Cambridge	UK	+	+	+	0	+	+	+	+	+	+	+	+	11	1	0
Univ Edinburgh	UK	+	+	+	-	+	+	-	-	-	-	-	-	5	0	7
Imperial Coll London	UK	+	+	+	0	+	+	-	+	-	0	0	+	7	3	2
King's Coll London	UK	0	+	+	-	+	+	-	-	-	-	-	-	4	1	7
Univ Manchester	UK	0	+	+	-	+	+	-	-	-	-	0	-	4	2	6
Univ Nottingham	UK	-	+	0	-	+	+	-	-	-	-	-	-	3	1	8
Univ Oxford	UK	+	+	+	0	+	+	+	+	0	+	+	+	10	2	0
Univ Sheffield	UK	-	+	0	-	+	+	-	-	-	-	-	-	3	1	8
Univ Coll London	UK	+	+	+	0	+	+	0	+	-	+	0	0	7	4	1

(cont.)

		Comparison with the US average														
		THES indicators						SJTU indicators								
University	country	Academic review	Recruiter review	Teacher to student ratio	Citations per faculty	International students	International staff	Alumni winning Nobel prizes	Staff winning Nobel prizes	Highly cited researchers	Articles in Nature & Science	Articles in Science & Social Sci	Academic performance - size	Above	Close to	Below
Boston Univ	US	0	+	-	0	-	+	-	-	-	-	-	-	2	2	8
Brown Univ	US	0	0	+	0	+	0	-	-	-	-	-	-	2	4	6
California Inst Tech	US	+	-	+	+	+	+	+	+	+	+	-	+	10	0	2
Univ California - Berkeley	US	+	+	-	0	+	+	+	+	+	+	+	+	10	1	1
Univ California - Davis	US	0	-	0	0	-	-	-	-	0	-	+	-	1	4	7
Univ California - Irvine	US	0	-	-	0	-	-	-	0	-	-	-	0	0	4	8
Univ California - Los Angeles	US	+	+	-	0	-	-	-	+	+	+	+	0	6	2	4
Univ California - San Diego	US	+	-	-	+	-	-	-	+	+	+	+	+	7	0	5
Univ California - Santa Barbara	US	0	-	-	0	0	-	-	+	0	0	-	0	1	6	5
Carnegie Mellon Univ	US	+	+	+	0	+	+	+	+	-	-	-	-	7	1	4
Case Western Reserve Univ	US	-	-	+	0	-	+	+	-	-	-	-	0	3	2	7
Univ Chicago	US	+	+	+	0	+	+	+	+	+	+	0	+	10	2	0
Columbia Univ	US	+	+	+	0	-	+	+	+	+	+	+	+	10	1	1
Cornell Univ	US	+	+	+	0	-	+	+	+	+	+	+	+	10	1	1
Duke Univ	US	+	+	+	0	-	+	-	-	0	+	0	+	6	3	3
Univ Florida	US	-	-	+	-	-	-	-	-	-	-	+	-	2	0	10
Harvard Univ	US	+	+	+	+	+	+	+	+	+	+	+	+	12	0	0
Univ Illinois - Urbana Champaign	US	+	-	-	0	-	0	+	+	0	0	0	-	3	5	4
Johns Hopkins Univ	US	+	0	+	+	-	+	+	0	0	+	+	-	7	3	2
Massachusetts Inst Tech (MIT)	US	+	+	+	+	-	+	+	+	+	+	0	+	10	1	1
Univ Maryland - Coll Park	US	-	-	0	0	0	-	-	-	-	-	0	-	0	4	8
Univ Michigan - Ann Arbor	US	+	+	-	0	0	0	+	-	+	0	+	-	5	4	3
Michigan State Univ	US	-	0	0	0	0	-	-	-	-	-	-	-	0	4	8
New York Univ	US	+	+	-	-	-	-	+	-	0	-	0	-	3	2	7
Univ North Carolina - Chapel Hill	US	-	+	-	-	-	-	-	-	-	-	0	-	1	1	10
Northwestern Univ	US	0	+	+	0	-	+	-	-	0	-	0	0	3	5	4
Pennsylvania State Univ - Univ Par	US	0	+	0	-	-	-	-	-	0	0	0	-	1	5	6
Univ Pennsylvania	US	+	+	+	0	+	+	+	+	+	0	+	+	10	2	0
Univ Pittsburgh - Pittsburgh	US	-	-	+	0	+	-	-	-	-	-	+	-	3	1	8
Princeton Univ	US	+	+	+	+	+	+	+	+	+	+	-	+	11	0	1
Purdue Univ - West Lafayette	US	0	0	-	0	+	+	-	-	-	-	-	-	2	3	7
Rice Univ	US	-	-	+	0	-	0	-	-	-	-	-	-	1	2	9
Univ Rochester	US	-	-	+	0	-	+	0	-	-	-	-	0	2	3	7
Rutgers State Univ - New Brunswic	US	-	-	-	-	+	-	-	-	-	-	-	-	1	0	11
Univ Southern California	US	-	0	-	0	-	+	-	0	-	-	0	-	1	4	7
Stanford Univ	US	+	+	0	+	-	+	+	+	+	+	+	+	10	1	1
Texas A&M Univ - Coll Station	US	0	0	-	0	0	-	-	-	-	-	0	-	0	5	7
Univ Texas - Austin	US	+	+	-	0	+	-	-	-	0	-	0	-	3	3	6
Univ Colorado - Boulder	US	-	-	+	0	-	-	-	0	-	0	-	-	1	3	8
Univ Minnesota - Twin Cities	US	0	-	-	0	-	-	+	-	0	0	+	-	2	4	6
Univ Arizona	US	-	-	-	0	-	-	-	-	-	0	0	-	0	3	9
Indiana Univ - Bloomington	US	-	0	-	0	0	-	-	-	-	-	-	-	0	3	9
Ohio State Univ - Columbus	US	-	0	-	0	+	-	-	-	0	-	0	-	1	4	7
Vanderbilt Univ	US	-	+	+	0	-	-	-	0	-	-	-	0	2	3	7
Univ Washington - Seattle	US	0	-	+	0	0	-	0	+	+	+	+	-	5	4	3
Washington Univ - St. Louis	US	-	-	+	-	-	-	-	0	-	+	0	+	3	2	7
Univ Wisconsin - Madison	US	+	+	-	+	+	-	+	+	+	+	+	-	9	0	3
Yale Univ	US	+	+	+	0	+	+	+	+	+	+	0	+	10	2	0

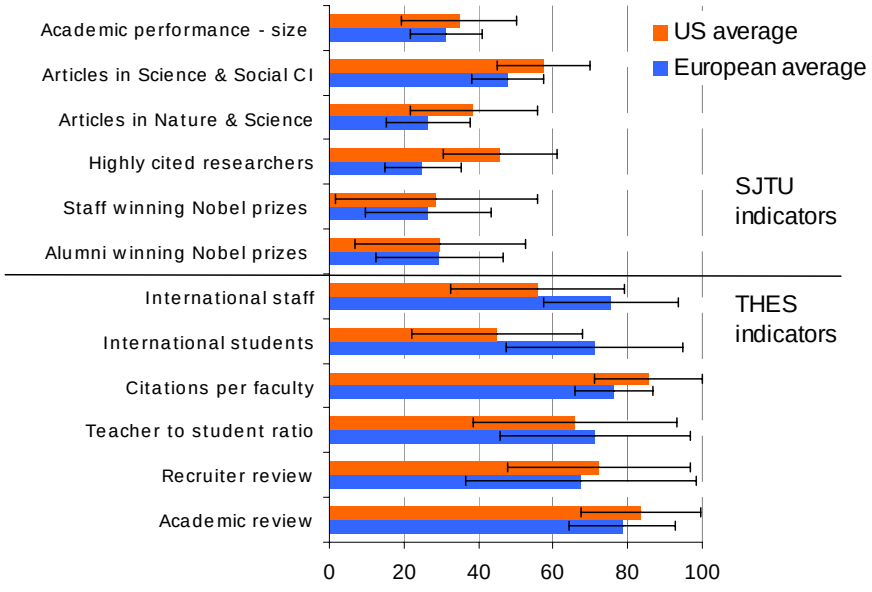
Notes:

1. 48 US Universities were considered in the calculation of the average US score. These US Universities are common to the THES and the SJTU rankings.
2. Value "+" indicates that the University score is above the US average by at least 10%; Value "0" indicates that the University score is near the US average (within ± 10%); Value "-" indicates that the University score is below the US average by at least 10%.

Table 28 presents the average European and the average US university score in the six THES and six SJTU indicators. Again, only the Top 100 universities in the SJTU ranking were considered, that are at the same time ranked by the THES index. Thus, 27 European universities and 48 USA universities are included. The standard deviations are also shown in the table, so as to be used as an indication of the degree of homogeneity in universities performance.

An interesting result is that average US university is not necessarily superior to the average European university unlike most of the current conceptions might suggest. The average US university has a better performance than the average European university in the *number of articles in Science and Social Citation Index*, in the *number of highly cited researchers* (SJTU indicators) and in the *citations per faculty* (THES indicator). Yet, the average European university has a better performance than the average American university in the *proportion of international staff* and the *proportion of international students*. Moreover, it is very interesting to note that the performance of the average European university is comparable to the average US university for the remaining indicators. Teaching and research quality measured through the number of *staff and alumni winning a Nobel prize or a Field Medal* do not significantly differ between Europe and US. Regarding homogeneity issues, the European universities included in our study are more homogenous than their American counterparts in the majority of the indicators, as shown by the more narrow standard deviations in the majority of the indicators for the European universities.

Table 28. Benchmarking average European university scores in the six THES and six SJTU indicators with respect to the average US University



6. Conclusions

When in the early 2000s the Institute of Higher Education of the Shanghai Jiao Tong University released the SJTU classification of international universities, they certainly did not anticipate the upheaval that the release of their ranking would provoke in Europe. The realisation that many of the continent's higher education jewels in fact ranked low – or did not rank at all – was a major shock that forced European governments to confront a reality they had preferred to overlook. Since its first appearance in 2003, the SJTU ranking has set in motion a major re-examination of higher education policies throughout Europe. It has also triggered reform initiatives aimed at fostering excellence and recognition, illustrating again the potency of benchmarking.

Whilst the discussion of the limitations of the indicators underlying the THES and the SJTU rankings have been the object of attention and discussion, there has been no real attempt to examine in depth the impact on the universities ranks of the methodological assumptions made in compiling these two university ranking systems.

The purpose of this report was precisely to take part to the debate by carrying out a robustness analysis of the two ranking systems. After having reviewing the main features and limitations of the 2007 SJTU and THES, we have examined the degree of association between the indicators used to built either SJTU or THES.

The main contribution of this report lies in the uncertainty analysis of the two rankings. We have estimated how much the university ranks depend on the methodology (set of selected indicators, weighting scheme, normalization and aggregation method) chosen for the compilation of the THES or the SJTU. This has allowed us to identify for which universities the THES and SJTU ranking systems can be reliably used to draw conclusions. Finally, in the last part of the report, we offer a discussion on benchmarking average university performance in Europe compared to the United States.

This JRC report has set four main goals:

- To throw light on the methodological issues and eventual limitations of the SJTU and THES rankings;
- To assess the robustness of the two higher education ranking systems with a view to identify for which universities these ranking can be reliably used to draw conclusions;
- To propose, if possible and despite the known limitations of the currently available indicators in the THES and SJTU, an approach that combines these pieces of information in the least biased way;
- To identify whether the average European university lags indeed behind the average US university based on the set of twelve indicators of the THES and SJTU frameworks.

To achieve these goals, besides the classical tools of multivariate analysis, we carry out a thorough uncertainty and sensitivity analysis of the 2007 SJTU and THES rankings under a plurality of scenarios in which we activate simultaneously different sources of uncertainty. The sources cover a wide and versatile spectrum of methodological assumptions (all with their advantages and implications). Subsequently, a frequency matrix of the university ranks is calculated across the different simulations (triggering the exclusion of an indicator, the weighting and the aggregation rule). Such a multi-modeling approach allows one to deal with the criticism, often made to league tables and rankings systems, that ranks are presented as if they were calculated under conditions of certainty while this is rarely the case. Thus we deviate from the classic approach – also taken in the two university ranking systems - to build a composite indicator by a simple weighted summation of indicators.

Upon propagating uncertainties a university rank is no longer a simple number, but a distribution of values. Thus, the ranking system might be seen to lose relevance if a high fraction of universities were to overlap with one another. In fact if a high number of universities overlap (wide range of ranks), this casts doubts on the relative position of the universities. In general, there is a trade-off between the level of uncertainty that is included in the ranking system and its worthiness, which is herein considered as the capacity of the system to discriminate effectively between universities.

The combined and iterative use of uncertainty and sensitivity analysis during the preparation of a ranking system for universities could therefore contribute to a balanced structure, could provide information on whether the universities rankings measure anything meaningful and could reduce the possibility that the classification may send misleading or non-robust policy messages.

The verification that is offered in the present work is nevertheless partial. We have considered the 88 universities that are ranked Top100 in the SJTU and Top200 in the THES and which are common to both rankings. Furthermore we have implicitly assumed that all the plurality of the debate (i.e. the sources of uncertainty) is captured by the variability in the weights (equal weights, original set of weights, factor analysis-derived weights, data envelopment analysis), the aggregation rule (additive, geometric, Borda multi-criteria analysis) and the exclusion of one indicator from the dataset. Even if the dataset we have used is restricted to 88 universities and even if we have not propagated other eventual sources of uncertainty (e.g., data uncertainty which was not available) in our study, it should be clear to the reader that this can be done in principle without difficulty, following a similar approach to that presented herein.

This report puts forward three findings and recommendations.

The three findings are that:

- While indicators and league tables are enough to start a discussion on higher education issues in Europe and benchmark it worldwide, they are not sufficient to conclude it. As already widely discussed in the literature, the choice of the indicators reflects more the league tables compilers' opinion and the availability of internationally comparable data than the result of a consensus from the academic community. Both rankings rely highly on bibliometric indicators and thus they tend to be biased towards English-speaking and hard sciences intensive institutions, leaving aside social and human sciences.

- The THES and SJTU rankings should not be used to discuss about the determinants of university performance (Aghion *et al.*, 2008) or to deliver policy messages on educational issues. Indeed, for the majority of the universities we analysed, the THES or SJTU rank have proven impossible to capture with adequate statistical robustness.

The assigned university rank largely depends on the methodological assumptions made in compiling the two rankings. For instance, we cannot conclude that Paris VI University performs significantly better than McGill University though the difference in positions suggests a disparity in quality or performance. It implies that no conclusive inference regarding the relative performance of the majority of the universities can be drawn from either ranking.

- The average US university is not necessarily superior to the average European university unlike most of the current conceptions might suggest. An analysis of the 27 European universities and 48 USA universities that are ranked Top100 in the SJTU and Top200 in the THES shows that the average US university is not necessarily superior to the average European university. The average US university has a better performance than the average European university in the *number of articles in Science and Social Citation Index*, in the *number of highly cited researchers* (SJTU indicators) and in the *citations per faculty* (THES indicator). Yet, the average European university has a better performance than the average American university in the *proportion of international staff* and the *proportion of international students*. For the remaining seven indicators analysed (in particular the two indicators related to the number of *Alumni* or *Staff winning Nobel prizes and field medal*), the performance of the average European university is comparable to the average US university. Regarding homogeneity issues, the European universities analysed have a more homogenous performance than their American counterparts.

We recommend that the university ranking systems can and should be improved as follows:

- First, the indicators should be revisited along the lines of the recommendations and suggestions already provided by the Berlin Principles (see Box 1 in Appendix). The Berlin principles place emphasis on league tables for universities that recognize the diversity of institutions, provide clear information about the indicators and target groups. The principles also provide recommendations on the way data should be gathered, processed in a transparent way and how final rankings should be presented. The THES

ranking fails to comply with the Berlin principles as, for instance, there is clearly a lack of information surrounding the construction of the two expert driven indicators.

- Second, the compilation of university rankings should always be accompanied by a robustness analysis. We believe that this could constitute an additional recommendation to be added to the already 16 existing Berlin principles. The multi-modeling approach adopted in this report allowed us to show that the rank of most of the 88 institutions is highly dependant on the methodology chosen for the compilation of both rankings. In our study we have selected numerous scenarios that represent distinct, diverse and at times contradicting approaches in order to aggregate information on university performance. The multi modeling approach employed, has already proven to be useful in the development and validation of several composite indicators (e.g., Environmental Performance Index, Composite Learning Index, Alcohol Policy Index, Knowledge Economy Index) and was also included in the JRC/OECD Handbook on Composite Indicators. A comparative advantage of the multi-modeling approach is that it can offer a representative picture of the classification of university performances. While university rankings can not inform us about the real position of most of universities, given the statistical uncertainty associated with the ranks, a multi-modeling approach, like the one implemented in this report, allows to rank institutions in a range bracket. The upshot is that this way of doing is probably better than assigning a specific rank which is not representative of the real performance of the university.

- Third, the assessment of the universities performance based on the hybrid set of the twelve indicators used in the THES and SJTU rankings provides a more reliable average rank of the institutions. The two sensitivity measures we used showed that the impact of the methodological assumptions is much lower when using the set of twelve indicators as opposed to either the THES or the SJTU indicators alone. Given the diversity of the indicators, as confirmed by correlation analysis, and the fact that the number of statistical dimensions in the combined THES&SJTU framework is twice the number of statistical dimensions for either the THES or the SJTU (result of factor analysis), more diverse aspects of universities are captured if all twelve indicators are

considered. The linkages between the THES indicators on one side and the SJTU indicators on the other are positive and significant, yet fair ($r \leq 0.58$). This result evidences the relatively low degree of overlap of information between the two sets and suggests that an eventual merging of the twelve indicators may provide a more holistic picture of the universities performance.

Even if all three previous recommendations are taken into consideration, one further issue remains: the high volatility of more than half of the universities we analysed. We recall the reader that these universities are considered the "elite" of the thousands of universities world-wide. If the ranks of those universities are full of uncertainty, let alone the ranks of the universities further down the classification ladder. This high volatility calls for a revision of the set of indicators, either by enriching it with other dimensions that are crucial to assessing university performance or by revising some of the existing indicators in order to remove some of the bias present (e.g., eliminate bias in favour of old and/or big universities and/or hard sciences). A legitimate question is raised: when will the revision of the dataset of indicators reach a satisfactory level? Uncertainty and sensitivity analysis should be employed as a guide to determine when the revision process of the indicators has reached a satisfactory level. We would argue that the stopping criterion for the revision is reached when, upon acknowledging the methodological uncertainties that are intrinsic to the development of a ranking system, the space of inference of the ranks for the majority of the universities is narrow enough to justify a meaningful classification.

We hope that the debate of this study will

- lead to improvements to league tables methodologies;
- enable users to better understand the complexities of the league tables, and avoid misunderstanding them; and
- help higher education institutions develop approaches that help them satisfy the legitimate information needs of their stakeholders.

7. References

Acemoglu, D., Aghion, P., F. Zilibotti, 2006, Distance to Frontier, Selection, and Economic Growth. *Journal of the European Economic Association* 4(1):37-74.

Aghion, P., Dewatripont . M. Hoxby, C., Sapir, A., 2008, *Higher aspirations: An agenda for reforming European universities*. Bluegel Blueprint Series N. 5.

Andrews, C. J., Hassenzahl, D.M., Johnson B.B, 2004, Accommodating uncertainty in comparative risk, *Risk Analysis* 24(5):1323-1335.

Booysen, F., 2002, An overview and evaluation of composite indices of development, *Social Indicators Research* 59(2):115-151.

Borda, J.C. de, 1784, *Mémoire sur les élections au scrutin*, in Histoire de l'Académie Royale des Sciences, Paris.

Brand, D. A., Saisana, M., Rynn, L. A., Pennoni, F., Lowenfels, A. B., 2007, Comparative Analysis of Alcohol Control Policies in 30 Countries, *PLoS Medicine* 4(4):752-759.

Brown, R., 2006, League tables–do we have to live with them? *Perspectives*10:33-38.

Cherchye, L., Moesen, W., Rogge, N., van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S., 2008, Creating Composite Indicators with DEA and Robustness Analysis: the case of the Technology Achievement Index, *Journal of Operational Research Society* 59:239-251.

Cherchye, L., Moesen, W., van Puyenbroeck, T., 2004, Legitimately Diverse, Yet Comparable: on Synthesising Social Inclusion Performance in the EU, *Journal of Common Market Studies*, 42:919-955.

CHERI/HEFCE, 2008, *Counting what is measured or measuring what counts?* League tables and their impact on higher education institutions in England, Report to HEFCE by the Centre for Higher Education Research and Information (CHERI), Open University, and Hobsons Research, April 2008/14.

Condorcet, M. de, 1785, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la probabilité des voix*, De l' Imprimerie Royale, Paris.

European Commission (2008) Progress towards the Lisbon Objectives in Education and Training, Indicators and Benchmarks, Commission Staff Working Document.

Hoskins, B., Jesinghaus, J., Mascherini, M., Munda, G., Nardo, M., Saisana, M., van Nijlen, D., Vidoni, D., Villalba, E., 2006, *Measuring Active Citizenship in Europe*. CRELL Research Paper 4. EUR 22530 EN.

IREG, 2006, *Berlin Principles on Ranking Higher Education Institutions.*

JRC/OECD, 2008, *Handbook on Constructing Composite Indicators*. Methodology and user Guide, OECD Publishing, ISBN 978-92-64-04345-9.

Keeney, R., Raiffa, H., 1976, *Decision with multiple objectives: preferences and value trade-offs*, Wiley, New York.

Kemeny, J., 1959, Mathematics without numbers, *Daedalus* 88:577–591.

Liu, N. C., Cheng, Y., 2007, Academic Ranking of World Universities: Methodologies and Problems. In Sadlak, J., Liu Nian Cai, eds. *The World-Class University and Ranking: Aiming Beyond Status.* UNESCO-CEPES, Shanghai Jiao Tong University, Cluj University Press, 2007.

Marginson, S., 2007, Global University Rankings: Implications in general and for Australia . *Journal of Higher Education Policy and Management* 29(2):131-142

Melyn, W., Moesen, W.W., 1991, Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available, *Public Economic research Paper* 17, CES, KU Leuven.

Munda, G., Nardo, M., Saisana, M., Srebotnjak, T., 2008, Measuring uncertainties in composite indicators of sustainability, *Int. J. Environmental Technology and Management*, forthcoming.

Munda, G., Saisana, M., 2008, Methodological Considerations on Regional Sustainability Assessment based on Multicriteria and Sensitivity Analysis, *Regional Studies*, forthcoming.

Podinovskii, V.V., 1994, Criteria importance theory, *Mathematical Social Sciences* 27: 237-252.

Saisana, M., 2008, *The 2007 Composite Learning Index: Robustness Issues and Critical Assessment*, EUR Report 23274 EN, European Commission, JRC-IPSC, Italy.

Saisana, M., Munda, G., 2008, *Knowledge Economy: measures and drivers*, EUR Report number pending EN, European Commission, JRC-IPSC.

Saisana, M., Nardo, M., Saltelli, A., 2005, Uncertainty and Sensitivity Analysis of the 2005 ESI, in "Esty, Daniel C., Marc Levy, Tanja Srebotnjak, and Alexander de Sherbinin (2005) 2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship. New Haven: Yale Center for Environmental Law & Policy", pp. 75-87.

Saisana, M., Saltelli, A., 2006, Uncertainty and Sensitivity Analysis of the 2006 Environmental Performance Index, in "Esty, Daniel C., Marc A. Levy, Tanja Srebotnjak, Alexander de Sherbinin, Christine H. Kim, and Bridget Anderson (2006) Pilot 2006 Environmental Performance Index. New Haven: Yale Center for Environmental Law & Policy", pp.291-306.

Saisana, M., Saltelli, A., 2008, *Sensitivity Analysis for the 2008 Environmental Performance Index*, EUR Report number pending EN, European Commission, JRC-IPSC.

Saisana, M., Saltelli, A., Tarantola, S., 2005, Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators, *Journal of the Royal Statistical Society A* 168(2):307-323.

Saisana, M., Tarantola S., 2002, *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*, EUR Report 20408 EN, European Commission, JRC-IPSC, Italy.

Salmi, J., Saroyan, A., 2007, League Tables as Policy Instruments. *Higher Education Management and Policy* 19:24-62.

Sapir, A., Aghion, P., Bertola, G., Hellwig, M., Pisani-Ferry, Rosati, D. , Jean, Viñals, J., Wallace, H., Buti, M., Nava, M., Smith, P.M., 2004, *An Agenda for a Growing Europe: The Sapir Report*. Oxford University Press. ISBN-13: 978-0-19-927148-1

Taylor, P., Braddock, R., 2007, International University Ranking systems and the Idea of University Excellence, *Journal of Higher Education Policy and Management* 29(3):245-260.

Thanassoulis, E., Portela, M.C., Allen, R., 2004, Incorporating value judgments in DEA, in W.W. Cooper, L.M. Seiford and J. Zhu (eds.), Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, Boston.

Usher, A., Savino, M., 2007, A Global Survey of University Ranking and League Tables. *Higher Education in Europe*, 32(1):5-15.

van Raan, F.J., 2007, Challenges in the Ranking of Universities. In Sadlak, J., Liu Nian Cai, eds. *The World-Class University and Ranking: Aiming Beyond Status.* UNESCO-CEPES, Shanghai Jiao Tong University, Cluj University Press.

Vansnick, J. C., 1990, Measurement theory and decision aid - in Bana e Costa C.A. (ed.) - Readings in multiple criteria decision aid, Springer-Verlag, Berlin, pp. 81-100.

Williams, R., 2007, Broadening the Criteria: Lessons from the Australian Rankings. In Sadlak, J., Liu Nian Cai, eds. *The World-Class University and Ranking: Aiming Beyond Status.* UNESCO-CEPES, Shanghai Jiao Tong University, Cluj University Press.

Wong, Y-H B., Beasley, J.E., 1990, Restricting weight flexibility in data envelopment analysis, *Journal of the Operational Research Society* 47:136-150.

Yorke, M., Longden, B., 2005, *Significant Figures – Performance Indicators and 'League Tables'*, London, Standing Conference of Principals (now GuildHE).

Young, H. P., Levenglick, A., 1978, A Consistent Extension of Condorcet's Election Principle, *SIAM Journal on Applied Mathematics* 35:285–300.

Zitt, M., and Filliatreau, G., 2007, Big Is (Made) Beautiful – Some comments about Shanghai-Ranking of World-Class Universities. In Sadlak, J., Liu Nian Cai, eds. *The World-Class University and Ranking: Aiming Beyond Status.* UNESCO-CEPES, Shanghai Jiao Tong University, Cluj University Press.

ANNEX: Methodological boxes and additional information

Box 1. The Berlin Principles 93
Box 2. Multi-criteria Analysis and its role in the THES & SJTU analysis 94
Box 3. Data Envelopment Analysis and its role in the THES & SJTU analysis 95

Table A 1. Scaled data for THES and SJTU used in the analysis 96
Table A 2. Summary statistics for the THES and SJTU indicators used in the analysis (88 universities that are ranked both in THES and SJTU) 97
Table A 3. Sensitivity measures: impact of the assumptions on the THES rank 98
Table A 4. Sensitivity measures: impact of the assumptions on the SJTU rank 99
Table A 5. Sensitivity measures: impact of the assumptions on the median rank (THES& SJTU, 130 scenarios) 100

Box 1. The Berlin Principles

The Berlin Principles outline good practice in the compiling of higher education rankings. Prompted by the increased publication and use of higher education league tables, two organisations, the Institute for Higher Education Policy in Washington and the UNESCO European Centre for Higher Education (UNESCO-CEPES) in Bucharest founded the International Ranking Expert Group (IREG) in 2004. At its second meeting in 2006, the group developed a set of principles of quality and good practice for higher education rankings – referred to as the Berlin Principles on Ranking of Higher Education Institutions (IREG, 2006). The principles, mainly aimed at compilers of league tables, outline guidelines on four main areas: (a) purposes and goals, (b) methodologies (design and weighting of indicators), (c) collection and processing of data, and (d) presentation of ranking results.

(a) Purposes and goals

- Rankings should not be the only way that higher education institutions are assessed, but rather they should complement the work of the Government and other bodies overseeing higher education.
- Compilers should bear their target group and the purpose of the league table in mind and develop their ranking system accordingly – there is no 'one size fits all'. In this perspective, compilers should recognise diversity, taking the different missions and goals of higher education institutions into account and consult frequently with experts and the institutions themselves.
- Compilers should state which sources of data are used in the rankings and be clear about the message that each source communicates. It is regarded as good practice to combine the different perspectives provided by several sources in order to get a more complete view of each institution.
- International rankings in particular should take into account the specific features of different higher education systems, and allow for the fact that notions of quality are not necessarily shared by different nations or systems.

(b) Methodologies (design and weighting of indicators):

- Compilers need to be transparent about their methodologies and choosing the indicators according to their relevance and validity (i.e. rather than counting what is measured, measure what counts).
- Compilers need to measure outcomes rather than inputs wherever possible, make the weights assigned to different indicators (if used) prominent and limit any changes to these.

(c) Collection and processing of data

- It is recommended the use of audited and verifiable data whenever possible, including data that are collected according to recognised procedures for scientific data collection (to avoid bias), the application of quality assurance measures to the ranking processes themselves and organisational measures which enhance the credibility of rankings, such as advisory boards.

(d) Presentation of ranking results

- Compilers are recommended to provide users with a clear understanding of all factors used to develop a league table and allow users to choose how the ranking is displayed (and ideally how the variables should be weighted).
- Rankings should be compiled in a way that eliminates or reduces errors in the original data and be organised or published in a format that allows compilers to make corrections if necessary and for users of the rankings to be made aware of these.

Box 2. Multi-criteria Analysis and its role in the THES & SJTU analysis

The two international university rankings THES and SJTU are based essentially on an additive (and linear) model. Some policy analysts challenge aggregations based on additive models, inter alia, because of the undesired, at times, property of compensability. Compensability refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some indicators by a sufficiently large advantage on another indicator, whereas smaller advantages would not do the same. Thus a preference relation is non-compensatory if no trade-off occurs and is compensatory otherwise. The use of weights, to be attached to the indicators, with intensity of preference originates compensatory multi-criteria methods and gives the meaning of trade-offs to the weights. On the contrary, the use of weights with ordinal criterion scores originates non-compensatory aggregation procedures and gives the weights the meaning of importance coefficients (Keeney and Raiffa, 1976; Podinovskii, 1994). Vansnick (1990) showed that the two main approaches in multi-criteria decision theory i.e., the compensatory and non-compensatory ones can be directly derived from the seminal work of Borda (1784) and Condorcet (1785). Indeed, looking at social choice literature, one can realize that various ranking procedures used in multi-criterion methods have their origins in social choice.

To deal with the issue of eventual compensability among the indicators values, we built 21 scenarios (out of 70 for either THES or SJTU) that employ a multicriteria method, proposed by Brand *et al.* 2007, and which is essentially a combination of the Borda and the Condorcet-Kemeny-Young-Levenglick approaches (Kemeny, 1959; Young and Levenglick, 1978). Specifically, the algorithm computes scores for a university i as follows:

$$Y_i = \sum_{j=1}^{6} (n_{ij} + \frac{k_{ij}}{2}) \cdot w_j \,,\ 1 \le i \le 88,\ 1 \le j \le 6$$

where

$n_{ij} \equiv$ number of universities that have weaker performance than university i relative to indicator j, $0 \le n_{ij} \le 87$

$k_{ij} \equiv$ number of universities with equivalent performance to university i relative to indicator j, $0 \le k_{ij} \le 87$

$w_j \equiv$ weight assigned to indicator j

In brief, when university *a* performs better than university *b* for a given indicator, then university *a* gets all the credit (= indicator's weight), whilst university *b* gets zero credit. In case two universities have equal values in a given indicator, the credit (weight) for that indicator is split equally between the two universities. This way, a university cannot "compensate" for a preponderance of weak performance in few indicators with a small number of exceptionally high values in few indicators. In other words, to attain a reasonably good score under this approach, a university must devote a reasonable amount of attention to the majority of indicators. This is not true under additive models, which are fully compensatory.

This approach was applied to calculate the simulated THES rank (in 21 scenarios), the simulated SJTU ranks (in 21 scenarios) and the hybrid THES&SJTU ranks (in 39 scenarios).

Box 3. Data Envelopment Analysis and its role in the THES & SJTU analysis

In absence of reliable information about the true weights to be attached to the 6 selected indicators in THES or in the SJTU, we endogenously selected those university-specific weights that maximize a university's rank in the overall classification ($n = 88$) using the Data Envelopment Analysis (DEA) method (Melyn & Moesen, 1991; Cherchye et al., 2004). This gives the following linear programming problem for each university i:

$$Y_i = \max_{wij} \frac{\sum_{j=1}^{6} y_{ij} w_{ij}}{\max_{y_c \in \{dataset\}} \sum_{j=1}^{6} y_{cj} w_{ij}} \qquad \text{(\textit{bounding constraint})}$$

Subject to

$w_{ij} \geq 0$ (*non-negativity constraint*)

where $j = 1,...,6$, $i = 1,...,88$

In this basic programming problem, the weights are non-negative and a university's score is between 0 (worst) and 1 (best). The DEA-based ranking system meets the important property of 'units invariance', which makes the normalisation stage for the underlying indicators redundant.

The non-negativity restriction on the weights, however, allows for extreme scenarios. If a university has a value in a given indicator that dominates the values of other universities, this university would always obtain a score of 1.0 even if it has very low values in many other indicators. Furthermore, it may lead to a situation where a large number of universities have top score equal to 1.0, rendering a further assessment between universities impossible. Therefore, some additional constraints on the weights were introduced, as recommended by several DEA supporters (see Thanassoulis *et al*. (2004) for a survey). We preferred to attach restrictions on the shares (instead of the weights), because shares (i) do not depend on measurement units and (ii) directly reveal the contribution of an indicator to the overall score (Cherchye et al., 2008). Formally, the j-th share for a university i is given as the product $y_{ij} w_{ij}$. Clearly, the sum of the shares equals the overall DEA-derived score. In what follows, we focus on *share constraints* (for each indicator *i*) of the type

$$L_i \leq \frac{y_{ij} w_{ij}}{\sum_{i=1}^{m} y_{ij} w_{ij}} \leq U_i \qquad \text{(\textit{share constraint})}$$

with L_i and U_i the respective lower and upper bounds (Wong and Beasley, 1990).

In our case the shares of the indicators in either the THES or the SJTU were constructed as equal to 25% below the equal weight equivalent ($L_i = 0.75 * 1/6 = 0.125$) or 25% above ($U_i = 1.25 * 1/6 = 0.208$). In others words, the share of each of the six indicators in THES (or in the SJTU) contributes to the DEA-derived score between the bounds: 12.5% and 20.8%.

In the case of the hybrid THES & SJTU set of twelve indicators the lower and upper bounds are 6.25% and 10.4% respectively ($L_i = 0.75 * 1/12 = 0.0625$; $U_i = 1.25 * 1/12 = 0.104$).

Table A 1. Scaled data for THES and SJTU used in the analysis

		THES indicators						SJTU indicators					
University	Country	AC-PEER	R-PEER	STAFF/S	CITATION	INT-S	INT-F	Alumni	Award	HiCi	N&S	SCI	Size
Australian Natl Univ	Australia	100.0	91.0	100.0	66.0	68.0	91.0	16.6	12.6	36.3	30.0	44.7	32.2
Univ Melbourne	Australia	100.0	99.0	64.0	70.0	64.0	95.0	14.4	14.1	22.2	18.4	55.1	25.0
Univ British Columbia	Canada	100.0	91.0	70.0	74.0	35.0	63.0	19.5	18.9	31.4	31.0	63.1	36.3
McGill Univ	Canada	100.0	97.0	99.0	72.0	73.0	96.0	27.0	0.0	29.6	24.2	58.7	32.7
McMaster Univ	Canada	84.0	49.0	31.0	89.0	28.0	37.0	15.6	18.9	22.2	14.3	45.4	25.7
Univ Toronto	Canada	100.0	96.0	21.0	93.0	86.0	50.0	26.3	19.3	39.2	37.7	77.6	44.4
Univ Helsinki	Finland	79.0	45.0	28.0	93.0	56.0	78.0	17.6	17.9	19.6	21.7	52.7	28.9
Ecole Normale Super Paris	France	91.0	60.0	83.0	98.0	61.0	81.0	48.5	31.6	12.8	16.8	25.8	25.8
Univ Strasbourg 1	France	58.0	64.0	19.0	82.0	45.0	77.0	27.6	22.5	16.6	18.5	32.8	22.9
Univ Paris 06	France	60.0	5.0	90.0	73.0	20.0	92.0	38.4	23.6	23.4	27.2	54.2	33.5
Univ Heidelberg	Germany	84.0	63.0	61.0	78.0	42.0	87.0	18.6	27.2	18.2	22.8	48.7	29.0
Univ Munich	Germany	80.0	60.0	70.0	72.0	58.0	71.0	34.8	22.9	14.8	27.1	51.8	31.1
Tech Univ Munich	Germany	68.0	71.0	88.0	69.0	59.0	83.0	40.3	23.6	25.7	20.2	44.4	29.9
Univ Freiburg	Germany	57.0	23.0	93.0	61.0	24.0	76.0	23.5	20.9	16.6	18.9	37.4	23.6
Univ Goettingen	Germany	66.0		73.0	59.0	41.0	54.0	36.3	20.0	14.8	16.3	39.7	25.2
Hebrew Univ Jerusalem	Israel	86.0	20.0	18.0	91.0	83.0	14.0	31.1	20.0	24.6	23.3	43.6	28.2
Kyoto Univ	Japan	99.0	89.0	83.0	90.0	29.0	24.0	37.2	33.4	38.5	35.1	68.6	30.6
Nagoya Univ	Japan	53.0	74.0	83.0	85.0	25.0	30.0	0.0	14.1	18.2	22.3	50.0	25.0
Osaka Univ	Japan	83.0	75.0	86.0	91.0	17.0	29.0	11.8	0.0	24.6	30.4	64.0	28.5
Tohoku Univ	Japan	53.0	59.0	96.0	84.0	40.0	32.0	17.6	0.0	22.2	23.7	63.0	28.3
Tokyo Inst Tech	Japan	67.0	86.0	59.0	91.0	34.0	42.0	15.6	0.0	22.2	23.9	46.9	30.4
Tokyo Univ	Japan	100.0	92.0	96.0	88.0	25.0	44.0	33.8	14.1	41.9	52.7	80.9	34.0
Univ Leiden	Netherlands	81.0	63.0	35.0	93.0	78.0	40.0	23.5	15.5	28.7	20.9	45.2	28.2
Univ Oslo	Norway	61.0	25.0	54.0	62.0	51.0	55.0	24.3	33.4	18.2	16.8	42.5	27.9
Lund Univ	Sweden	76.0	41.0	43.0	81.0	77.0	52.0	27.6	0.0	23.4	19.1	50.7	18.0
Uppsala Univ	Sweden	85.0	34.0	81.0	72.0	51.0	26.0	24.3	32.2	12.8	23.6	49.1	21.0
Univ Basel	Switzerland	52.0	20.0	99.0	67.0	89.0	81.0	24.3	17.1	21.0	22.9	35.6	35.1
Swiss Fed Inst Tech - Zurich	Switzerland	92.0	75.0	61.0	74.0	100.0	92.0	37.7	36.3	35.5	39.9	38.4	50.5
Univ Zurich	Switzerland	71.0	32.0	13.0	95.0	99.0	58.0	11.8	26.8	22.2	28.3	48.4	31.1
Univ Birmingham	UK	71.0	93.0	62.0	75.0	84.0	70.0	23.5	10.9	22.2	15.4	45.3	26.9
Univ Bristol	UK	81.0	98.0	85.0	77.0	88.0	72.0	10.2	17.9	29.6	26.7	47.3	32.8
Univ Cambridge	UK	100.0	100.0	99.0	83.0	98.0	91.0	93.6	91.5	54.0	58.2	65.4	65.1
Univ Edinburgh	UK	96.0	98.0	82.0	76.0	71.0	80.0	21.2	16.7	26.7	34.2	47.0	29.3
Imperial Coll London	UK	99.0	99.0	100.0	81.0	98.0	100.0	19.5	37.4	40.6	39.7	62.2	39.4
King's Coll London	UK	90.0	95.0	91.0	70.0	93.0	84.0	15.6	23.1	19.6	15.4	45.1	26.6
Univ Manchester	UK	88.0	99.0	77.0	70.0	84.0	85.0	25.6	18.9	24.6	28.3	56.9	28.4
Univ Nottingham	UK	69.0	98.0	64.0	65.0	84.0	88.0	14.4	20.0	24.6	16.2	44.8	27.3
Univ Oxford	UK	100.0	100.0	100.0	82.0	97.0	96.0	60.3	57.9	46.3	52.3	65.4	44.7
Univ Sheffield	UK	69.0	96.0	71.0	69.0	81.0	67.0	22.0	14.1	22.2	27.8	46.0	29.5
Univ Coll London	UK	96.0	97.0	100.0	82.0	91.0	98.0	28.8	32.2	38.5	42.9	63.2	33.8
Boston Univ	US	91.0	89.0	49.0	88.0	29.0	88.0	14.4	0.0	30.6	27.1	50.4	17.3
Brown Univ	US	90.0	77.0	74.0	89.0	75.0	58.0	17.6	13.6	31.4	28.2	39.9	29.6
California Inst Tech	US	100.0	55.0	100.0	100.0	100.0	91.0	55.5	69.1	58.4	67.6	50.3	100.0
Univ California - Berkeley	US	100.0	98.0	59.0	92.0	73.0	88.0	72.5	77.1	67.9	72.9	69.2	52.6
Univ California - Davis	US	83.0	34.0	60.0	82.0	31.0	26.0	0.0	0.0	46.9	33.1	64.2	30.0
Univ California - Irvine	US	79.0	30.0	38.0	82.0	27.0	30.0	0.0	29.4	35.5	28.0	48.9	32.1
Univ California - Los Angeles	US	100.0	92.0	56.0	91.0	20.0	36.0	25.6	42.8	57.4	49.1	75.9	35.5
Univ California - San Diego	US	98.0	39.0	51.0	95.0	23.0	30.0	16.6	34.0	59.3	55.5	64.6	46.6
Univ California - Santa Barbara	US	88.0	31.0	25.0	89.0	43.0	21.0	0.0	35.3	42.6	36.2	42.7	35.1
Carnegie Mellon Univ	US	96.0	94.0	76.0	87.0	67.0	96.0	32.8	32.8	32.3	12.0	36.7	31.0
Case Western Reserve Univ	US	59.0	50.0	99.0	85.0	18.0	80.0	38.1	11.5	21.0	19.6	43.0	32.2
Univ Chicago	US	100.0	97.0	100.0	86.0	71.0	90.0	70.8	80.2	50.8	42.8	54.1	41.3
Columbia Univ	US	100.0	96.0	94.0	91.0	35.0	89.0	76.0	65.7	56.5	54.3	69.6	46.4
Cornell Univ	US	100.0	98.0	74.0	93.0	36.0	69.0	43.6	51.3	54.5	51.4	65.1	39.9
Duke Univ	US	98.0	97.0	100.0	92.0	16.0	74.0	19.5	0.0	46.9	43.6	62.0	39.2
Univ Florida	US	62.0	41.0	77.0	74.0	27.0	32.0	21.2	0.0	37.1	24.8	65.5	26.5
Harvard Univ	US	100.0	100.0	100.0	96.0	93.0	91.0	100.0	100.0	100.0	100.0	100.0	73.0
Univ Illinois - Urbana Champaign	US	94.0	64.0	29.0	86.0	34.0	51.0	39.0	36.6	44.5	36.4	57.6	26.2
Johns Hopkins Univ	US	99.0	77.0	98.0	96.0	35.0	69.0	48.1	27.8	41.3	50.9	67.9	24.7
Massachusetts Inst Tech (MIT)	US	100.0	99.0	85.0	98.0	34.0	94.0	74.6	80.6	65.9	68.4	61.7	53.4
Univ Maryland - Coll Park	US	71.0	62.0	71.0	85.0	48.0	45.0	24.3	20.0	40.6	31.2	53.3	25.9
Univ Michigan - Ann Arbor	US	99.0	96.0	53.0	89.0	41.0	52.0	40.3	0.0	60.7	40.8	77.1	30.7
Michigan State Univ	US	42.0	78.0	61.0	90.0	46.0	44.0	11.8	0.0	36.3	24.6	51.0	18.4
New York Univ	US	95.0	93.0	48.0	77.0	29.0	49.0	35.8	24.5	41.3	34.4	53.9	25.9
Univ North Carolina - Chapel Hill	US	72.0	86.0	28.0	74.0	22.0	20.0	11.8	0.0	37.8	29.3	60.1	27.7
Northwestern Univ	US	88.0	97.0	77.0	91.0	35.0	68.0	20.4	18.9	46.9	34.2	57.0	36.9
Pennsylvania State Univ - Univ Park	US	84.0	82.0	63.0	64.0	18.0	24.0	13.2	0.0	45.1	37.7	58.0	23.7
Univ Pennsylvania	US	97.0	96.0	88.0	92.0	83.0	65.0	33.3	34.4	56.9	40.3	70.8	38.7
Univ Pittsburgh - Pittsburgh	US	62.0	45.0	94.0	85.0	78.0	38.0	23.5	0.0	39.9	23.6	65.6	28.5
Princeton Univ	US	100.0	94.0	95.0	97.0	83.0	75.0	62.3	80.4	59.3	42.9	46.5	58.9
Purdue Univ - West Lafayette	US	87.0	79.0	24.0	82.0	76.0	64.0	17.6	16.7	29.6	20.9	51.3	20.2
Rice Univ	US	65.0	45.0	80.0	88.0	39.0	58.0	20.4	21.9	22.2	22.3	30.0	29.8
Univ Rochester	US	58.0	38.0	100.0	84.0	24.0	63.0	31.1	8.9	26.7	21.5	43.1	35.2
Rutgers State Univ - New Brunswick	US	72.0	33.0	30.0	73.0	60.0	24.0	14.4	20.0	39.9	32.1	44.8	24.2
Univ Southern California	US	62.0	77.0	52.0	80.0	25.0	78.0	0.0	26.8	37.1	23.4	52.7	25.9
Stanford Univ	US	100.0	99.0	66.0	100.0	25.0	94.0	42.0	78.7	86.1	69.6	70.3	65.7
Texas A&M Univ - Coll Station	US	76.0	75.0	27.0	81.0	41.0	39.0	0.0	0.0	33.1	23.6	53.7	20.3
Univ Texas - Austin	US	95.0	94.0	22.0	92.0	66.0	47.0	20.4	16.7	46.9	28.0	54.8	21.3
Univ Colorado - Boulder	US	60.0	15.0	100.0	85.0	33.0	31.0	15.6	30.8	39.9	38.8	45.7	30.0
Univ Minnesota - Twin Cities	US	76.0	43.0	28.0	89.0	26.0	35.0	33.8	0.0	48.6	35.9	67.0	23.5
Univ Arizona	US	69.0	57.0	37.0	88.0	28.0	45.0	0.0	0.0	29.6	37.1	54.6	25.7
Indiana Univ - Bloomington	US	67.0	79.0	28.0	85.0	45.0	37.0	13.2	17.9	27.7	20.0	39.9	18.0
Ohio State Univ - Columbus	US	69.0	77.0	39.0	79.0	67.0	38.0	16.6	0.0	41.9	22.0	61.2	19.8
Vanderbilt Univ	US	55.0	81.0	99.0	87.0	28.0	41.0	19.5	29.6	31.4	23.8	51.0	36.0
Univ Washington - Seattle	US	84.0	50.0	73.0	92.0	44.0	33.0	27.0	31.8	52.4	49.0	74.1	27.4
Washington Univ - St. Louis	US	72.0	62.0	100.0	1.0	27.0	48.0	23.5	26.0	39.2	43.2	53.4	39.3
Univ Wisconsin - Madison	US	94.0	81.0	31.0	95.0	50.0	44.0	40.3	35.5	52.9	43.1	67.2	28.6
Yale Univ	US	100.0	98.0	100.0	91.0	84.0	75.0	50.9	43.6	57.9	57.2	63.2	48.9

Table A 2. Summary statistics for the THES and SJTU indicators used in the analysis (88 universities that are ranked both in THES and SJTU)

	Indicator	Mean	St.dev	CV	Max	Min
THES indicators	Academic review	82.4	16.0	0.2	100	42.0
	Recruiter review	71.7	26.3	0.4	100	5.0
	Teacher to student ratio	68.0	27.0	0.4	100	13.0
	Citations per faculty	82.4	13.2	0.2	100	1.0
	International students	53.2	26.0	0.5	100	16.0
	International staff	60.9	24.5	0.4	100	14.0
SJTU indicators	Alumni winning Nobel prizes	28.3	19.8	0.7	100	0.0
	Staff winning Nobel prizes	25.6	22.9	0.9	100	0.0
	Highly cited researchers	36.8	16.2	0.4	100	12.8
	Articles in Nature & Science	33.4	15.6	0.5	100	12.0
	Articles in Science & Social CI	54.6	12.5	0.2	100	25.8
	Academic performance - size	33.1	12.9	0.4	100	17.3

Table A 3. Sensitivity measures: impact of the assumptions on the THES rank

	Assumptions			First Sensitivity Measure	Second Sensitivity Measure
Scenario	Weighting	Aggregation	Excluded indicator	50th percentile	90th percentile
S65	DEA	Additive	Academic review	11.5	24
S25	Original set	Borda MCA	Teacher to student ratio	11	19
S67	DEA	Additive	Teacher to student ratio	10.5	27
S53	Equal weights	Geometric	Academic review	10.5	23
S66	DEA	Additive	Recruiter review	10	28
S36	Factor Analysis	Geometric	Recruiter review	10	28
S54	Equal weights	Geometric	Recruiter review	10	26
S68	DEA	Additive	Citations per faculty	10	25
S55	Equal weights	Geometric	Teacher to student ratio	9.5	25
S29	Factor Analysis	Additive	Academic review	9	26
S35	Factor Analysis	Geometric	Academic review	9	25
S37	Factor Analysis	Geometric	Teacher to student ratio	9	24
S56	Equal weights	Geometric	Citations per faculty	9	23
S70	DEA	Additive	International staff	9	22
S10	DEA	Additive	None	9	22
S8	Equal weights	Geometric	None	9	20
S48	Equal weights	Additive	Recruiter review	8.5	23
S49	Equal weights	Additive	Teacher to student ratio	8.5	22
S23	Original set	Borda MCA	Academic review	8	24
S30	Factor Analysis	Additive	Recruiter review	8	23
S47	Equal weights	Additive	Academic review	8	22
S31	Factor Analysis	Additive	Teacher to student ratio	8	22
S61	Equal weights	Borda MCA	Teacher to student ratio	8	22
S17	Original set	Geometric	Academic review	8	21
S58	Equal weights	Geometric	International staff	8	19
S13	Original set	Additive	Teacher to student ratio	8	18
S11	Original set	Additive	Academic review	7.5	23
S38	Factor Analysis	Geometric	Citations per faculty	7.5	22
S59	Equal weights	Borda MCA	Academic review	7	27
S44	Factor Analysis	Borda MCA	Citations per faculty	7	19
S5	Factor Analysis	Geometric	None	7	19
S50	Equal weights	Additive	Citations per faculty	7	18
S19	Original set	Geometric	Teacher to student ratio	7	18
S62	Equal weights	Borda MCA	Citations per faculty	7	17
S52	Equal weights	Additive	International staff	7	13
S43	Factor Analysis	Borda MCA	Teacher to student ratio	6.5	20
S41	Factor Analysis	Borda MCA	Academic review	6	28
S42	Factor Analysis	Borda MCA	Recruiter review	6	20
S32	Factor Analysis	Additive	Citations per faculty	6	19
S7	Equal weights	Additive	None	6	17
S69	DEA	Additive	International students	6	16
S39	Factor Analysis	Geometric	International students	6	15
S40	Factor Analysis	Geometric	International staff	6	13
S4	Factor Analysis	Additive	None	5.5	18
S27	Original set	Borda MCA	International students	5.5	15
S24	Original set	Borda MCA	Recruiter review	5.5	15
S46	Factor Analysis	Borda MCA	International staff	5	17
S6	Factor Analysis	Borda MCA	None	5	17
S60	Equal weights	Borda MCA	Recruiter review	5	17
S64	Equal weights	Borda MCA	International staff	5	16
S28	Original set	Borda MCA	International staff	5	15
S33	Factor Analysis	Additive	International students	5	15
S63	Equal weights	Borda MCA	International students	5	15
S57	Equal weights	Geometric	International students	5	15
S51	Equal weights	Additive	International students	5	14
S3	Original set	Borda MCA	None	5	14
S34	Factor Analysis	Additive	International staff	4.5	11
S9	Equal weights	Borda MCA	None	4	18
S45	Factor Analysis	Borda MCA	International students	4	13
S20	Original set	Geometric	Citations per faculty	4	11
S26	Original set	Borda MCA	Citations per faculty	4	8
S18	Original set	Geometric	Recruiter review	3	12
S12	Original set	Additive	Recruiter review	3	9
S21	Original set	Geometric	International students	3	8
S14	Original set	Additive	Citations per faculty	3	7
S22	Original set	Geometric	International staff	3	7
S2	Original set	Geometric	None	2.5	7
S15	Original set	Additive	International students	2	5
S16	Original set	Additive	International staff	1	4

Note: Assumptions different than those used originally in building THES by its developers are shaded in grey

Table A 4. Sensitivity measures: impact of the assumptions on the SJTU rank

	Assumptions			First sensitivity measure	Second sensitivity measure
Scenario	Weighting	Aggregation	Excluded indicator	50th percentile	90th percentile
S55	Equal weights	Geometric	Highly cited researchers	9	31
S68	DEA	Additive	Articles in Nature & Science	9	30
S70	DEA	Additive	Academic performance - size	9	27
S22	Original set	Geometric	Academic performance - size	9	26
S2	Original set	Geometric	None	9	24
S37	Factor Analysis	Geometric	Highly cited researchers	8.5	32
S20	Original set	Geometric	Articles in Nature & Science	8.5	26
S10	DEA	Additive	None	8.5	26
S69	DEA	Additive	Articles in Science & Social CI	8	32
S67	DEA	Additive	Highly cited researchers	8	31
S40	Factor Analysis	Geometric	Academic performance - size	8	29
S19	Original set	Geometric	Highly cited researchers	8	29
S8	Equal weights	Geometric	None	8	28
S5	Factor Analysis	Geometric	None	8	28
S21	Original set	Geometric	Articles in Science & Social CI	8	27
S38	Factor Analysis	Geometric	Articles in Nature & Science	7.5	29
S58	Equal weights	Geometric	Academic performance - size	7.5	28
S57	Equal weights	Geometric	Articles in Science & Social CI	7	31
S39	Factor Analysis	Geometric	Articles in Science & Social CI	7	30
S56	Equal weights	Geometric	Articles in Nature & Science	7	29
S35	Factor Analysis	Geometric	Alumni winning Nobel prizes	6.5	25
S53	Equal weights	Geometric	Alumni winning Nobel prizes	6	25
S17	Original set	Geometric	Alumni winning Nobel prizes	6	25
S65	DEA	Additive	Alumni winning Nobel prizes	6	23
S43	Factor Analysis	Borda MCA	Highly cited researchers	6	20
S61	Equal weights	Borda MCA	Highly cited researchers	6	16
S31	Factor Analysis	Additive	Highly cited researchers	5	22
S24	Original set	Borda MCA	Staff winning Nobel prizes	5	15
S62	Equal weights	Borda MCA	Articles in Nature & Science	5	13
S51	Equal weights	Additive	Articles in Science & Social CI	4	19
S66	DEA	Additive	Staff winning Nobel prizes	4	19
S54	Equal weights	Geometric	Staff winning Nobel prizes	4	19
S63	Equal weights	Borda MCA	Articles in Science & Social CI	4	17
S33	Factor Analysis	Additive	Articles in Science & Social CI	4	17
S49	Equal weights	Additive	Highly cited researchers	4	17
S32	Factor Analysis	Additive	Articles in Nature & Science	4	16
S45	Factor Analysis	Borda MCA	Articles in Science & Social CI	4	16
S36	Factor Analysis	Geometric	Staff winning Nobel prizes	4	16
S18	Original set	Geometric	Staff winning Nobel prizes	4	16
S44	Factor Analysis	Borda MCA	Articles in Nature & Science	4	15
S12	Original set	Additive	Staff winning Nobel prizes	4	15
S60	Equal weights	Borda MCA	Staff winning Nobel prizes	4	13
S25	Original set	Borda MCA	Highly cited researchers	4	12
S42	Factor Analysis	Borda MCA	Staff winning Nobel prizes	4	12
S59	Equal weights	Borda MCA	Alumni winning Nobel prizes	4	9
S41	Factor Analysis	Borda MCA	Alumni winning Nobel prizes	4	8
S30	Factor Analysis	Additive	Staff winning Nobel prizes	3.5	13
S64	Equal weights	Borda MCA	Academic performance - size	3.5	9
S48	Equal weights	Additive	Staff winning Nobel prizes	3	15
S15	Original set	Additive	Articles in Science & Social CI	3	14
S50	Equal weights	Additive	Articles in Nature & Science	3	13
S13	Original set	Additive	Highly cited researchers	3	13
S27	Original set	Borda MCA	Articles in Science & Social CI	3	12
S6	Factor Analysis	Borda MCA	None	3	10
S23	Original set	Borda MCA	Alumni winning Nobel prizes	3	9
S9	Equal weights	Borda MCA	None	3	9
S26	Original set	Borda MCA	Articles in Nature & Science	3	8
S46	Factor Analysis	Borda MCA	Academic performance - size	3	7
S28	Original set	Borda MCA	Academic performance - size	3	7
S4	Factor Analysis	Additive	None	2	9
S52	Equal weights	Additive	Academic performance - size	2	8
S7	Equal weights	Additive	None	2	8
S14	Original set	Additive	Articles in Nature & Science	2	7
S3	Original set	Borda MCA	None	2	7
S34	Factor Analysis	Additive	Academic performance - size	2	6
S47	Equal weights	Additive	Alumni winning Nobel prizes	2	6
S29	Factor Analysis	Additive	Alumni winning Nobel prizes	2	6
S11	Original set	Additive	Alumni winning Nobel prizes	1	6
S16	Original set	Additive	Academic performance - size	1	4

Note: Assumptions different than those used originally in building SJTU by its developers are shaded in grey

Table A 5. Sensitivity measures: impact of the assumptions on the median rank (THES& SJTU, 130 scenarios)

	Assumptions			First Sensitivity Measure	Second Sensitivity Measure
Scenario	Weighting	Aggregation	Excluded indicator	50th percentile	90th percentile
S35	Original set	Borda MCA	Academic review	6	20
S85	Equal weights	Additive	Teacher to student ratio	6	14
S93	Equal weights	Additive	Articles in Science & Social CI	5.5	12
S37	Original set	Borda MCA	Teacher to student ratio	5	21
S42	Original set	Borda MCA	Staff winning Nobel prizes	5	20
S119	DEA	Additive	Academic review	5	18
S127	DEA	Additive	Highly cited researchers	5	18
S129	DEA	Additive	Articles in Science & Social CI	5	18
S124	DEA	Additive	International staff	5	17
S125	DEA	Additive	Alumni winning Nobel prizes	5	17
S128	DEA	Additive	Articles in Nature & Science	5	17
S121	DEA	Additive	Teacher to student ratio	5	16
S126	DEA	Additive	Staff winning Nobel prizes	5	16
S130	DEA	Additive	Academic performance - size	5	16
S103	Equal weights	Geometric	Highly cited researchers	5	16
S60	Factor Analysis	Geometric	Recruiter review	5	16
S105	Equal weights	Geometric	Articles in Science & Social CI	5	15
S90	Equal weights	Additive	Staff winning Nobel prizes	5	14
S91	Equal weights	Additive	Highly cited researchers	5	14
S101	Equal weights	Geometric	Alumni winning Nobel prizes	5	14
S25	Original set	Geometric	Teacher to student ratio	5	14
S89	Equal weights	Additive	Alumni winning Nobel prizes	5	12
S13	Original set	Additive	Teacher to student ratio	4.5	20
S38	Original set	Borda MCA	Citations per faculty	4.5	15
S94	Equal weights	Additive	Academic performance - size	4.5	13
S18	Original set	Additive	Staff winning Nobel prizes	4.5	13
S120	DEA	Additive	Recruiter review	4	19
S36	Original set	Borda MCA	Recruiter review	4	19
S40	Original set	Borda MCA	International staff	4	19
S41	Original set	Borda MCA	Alumni winning Nobel prizes	4	19
S39	Original set	Borda MCA	International students	4	18
S76	Factor Analysis	Borda MCA	International staff	4	17
S77	Factor Analysis	Borda MCA	Alumni winning Nobel prizes	4	17
S24	Original set	Geometric	Recruiter review	4	17
S96	Equal weights	Geometric	Recruiter review	4	16
S75	Factor Analysis	Borda MCA	International students	4	16
S67	Factor Analysis	Geometric	Highly cited researchers	4	16
S23	Original set	Geometric	Academic review	4	16
S30	Original set	Geometric	Staff winning Nobel prizes	4	16
S10	DEA	Additive	None	4	15
S123	DEA	Additive	International students	4	15
S95	Equal weights	Geometric	Academic review	4	15
S102	Equal weights	Geometric	Staff winning Nobel prizes	4	15
S104	Equal weights	Geometric	Articles in Nature & Science	4	15
S106	Equal weights	Geometric	Academic performance - size	4	15
S78	Factor Analysis	Borda MCA	Staff winning Nobel prizes	4	15
S64	Factor Analysis	Geometric	International staff	4	15
S65	Factor Analysis	Geometric	Alumni winning Nobel prizes	4	15
S45	Original set	Borda MCA	Articles in Science & Social CI	4	15
S83	Equal weights	Additive	Academic review	4	14
S92	Equal weights	Additive	Articles in Nature & Science	4	14
S97	Equal weights	Geometric	Teacher to student ratio	4	14
S6	Factor Analysis	Borda MCA	None	4	14
S73	Factor Analysis	Borda MCA	Teacher to student ratio	4	14
S59	Factor Analysis	Geometric	Academic review	4	14
S61	Factor Analysis	Geometric	Teacher to student ratio	4	14
S69	Factor Analysis	Geometric	Articles in Science & Social CI	4	14
S31	Original set	Geometric	Highly cited researchers	4	14
S33	Original set	Geometric	Articles in Science & Social CI	4	14
S86	Equal weights	Additive	Citations per faculty	4	13
S111	Equal weights	Borda MCA	International students	4	13
S82	Factor Analysis	Borda MCA	Academic performance - size	4	13
S68	Factor Analysis	Geometric	Articles in Nature & Science	4	13
S44	Original set	Borda MCA	Articles in Nature & Science	4	13
S27	Original set	Geometric	International students	4	13
S28	Original set	Geometric	International staff	4	13
S32	Original set	Geometric	Articles in Nature & Science	4	13
S34	Original set	Geometric	Academic performance - size	4	13

	Assumptions			First Sensitivity Measure	Second Sensitivity Measure
Scenario	Weighting	Aggregation	Excluded indicator	50th percentile	90th percentile
S54	Factor Analysis	Additive	Staff winning Nobel prizes	4	12
S2	Original set	Geometric	None	4	12
S87	Equal weights	Additive	International students	4	11
S114	Equal weights	Borda MCA	Staff winning Nobel prizes	4	11
S113	Equal weights	Borda MCA	Alumni winning Nobel prizes	4	9
S55	Factor Analysis	Additive	Highly cited researchers	4	9
S46	Original set	Borda MCA	Academic performance - size	3.5	20
S3	Original set	Borda MCA	None	3.5	18
S29	Original set	Geometric	Alumni winning Nobel prizes	3.5	14
S5	Factor Analysis	Geometric	None	3.5	13
S63	Factor Analysis	Geometric	International students	3.5	13
S7	Equal weights	Additive	None	3.5	12
S16	Original set	Additive	International staff	3.5	11
S17	Original set	Additive	Alumni winning Nobel prizes	3.5	11
S72	Factor Analysis	Borda MCA	Recruiter review	3	19
S71	Factor Analysis	Borda MCA	Academic review	3	17
S122	DEA	Additive	Citations per faculty	3	16
S11	Original set	Additive	Academic review	3	15
S8	Equal weights	Geometric	None	3	14
S66	Factor Analysis	Geometric	Staff winning Nobel prizes	3	14
S108	Equal weights	Borda MCA	Recruiter review	3	13
S98	Equal weights	Geometric	Citations per faculty	3	13
S74	Factor Analysis	Borda MCA	Citations per faculty	3	13
S62	Factor Analysis	Geometric	Citations per faculty	3	13
S70	Factor Analysis	Geometric	Academic performance - size	3	13
S26	Original set	Geometric	Citations per faculty	3	13
S100	Equal weights	Geometric	International staff	3	12
S43	Original set	Borda MCA	Highly cited researchers	3	12
S84	Equal weights	Additive	Recruiter review	3	11
S109	Equal weights	Borda MCA	Teacher to student ratio	3	11
S12	Original set	Additive	Recruiter review	3	11
S15	Original set	Additive	International students	3	11
S19	Original set	Additive	Highly cited researchers	3	11
S20	Original set	Additive	Articles in Nature & Science	3	11
S88	Equal weights	Additive	International staff	3	10
S107	Equal weights	Borda MCA	Academic review	3	10
S53	Factor Analysis	Additive	Alumni winning Nobel prizes	3	10
S81	Factor Analysis	Borda MCA	Articles in Science & Social Cl	3	10
S1	Original set	Additive	None	3	10
S14	Original set	Additive	Citations per faculty	3	10
S21	Original set	Additive	Articles in Science & Social Cl	3	10
S22	Original set	Additive	Academic performance - size	3	10
S56	Factor Analysis	Additive	Articles in Nature & Science	3	9
S79	Factor Analysis	Borda MCA	Highly cited researchers	3	9
S9	Equal weights	Borda MCA	None	3	8
S110	Equal weights	Borda MCA	Citations per faculty	3	8
S118	Equal weights	Borda MCA	Academic performance - size	3	8
S115	Equal weights	Borda MCA	Highly cited researchers	3	7
S117	Equal weights	Borda MCA	Articles in Science & Social Cl	3	7
S57	Factor Analysis	Additive	Articles in Science & Social Cl	3	7
S99	Equal weights	Geometric	International students	2.5	15
S112	Equal weights	Borda MCA	International staff	2.5	13
S47	Factor Analysis	Additive	Academic review	2.5	8
S49	Factor Analysis	Additive	Teacher to student ratio	2	12
S52	Factor Analysis	Additive	International staff	2	11
S58	Factor Analysis	Additive	Academic performance - size	2	10
S51	Factor Analysis	Additive	International students	2	9
S116	Equal weights	Borda MCA	Articles in Nature & Science	2	7
S48	Factor Analysis	Additive	Recruiter review	2	7
S50	Factor Analysis	Additive	Citations per faculty	2	7
S80	Factor Analysis	Borda MCA	Articles in Nature & Science	2	7
S4	Factor Analysis	Additive	None	2	6

European Commission

EUR 23487 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: Higher Education Rankings: Robustness Issues and Critical Assessment. How much confidence can we have in Higher Education Rankings?
Author(s): Michaela Saisana and Beatrice D'Hombres
Luxembourg: Office for Official Publications of the European Communities
2008 – 106 pp. – 21 x 29.70 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-09704-1
DOI 10.2788/92295

Abstract

University rankings are very appealing, in that they provide one with a single number that allows one, at a glance, to situate a given university in the worldwide context. However, this very simplicity of use can be highly misleading in that most rankings are based on an extremely simple formula that arbitrarily aggregates subjectively chosen indicators. This JRC report has set four main goals:

1. To throw a considerable amount of light on the methodological issues and eventual limitations of the SJTU and THES rankings;
2. To assess the robustness of the two higher education ranking systems with a view to identify for which universities these ranking can be reliably used to draw conclusions;
3. To propose, if possible and despite the known limitations of the currently available indicators in the THES and SJTU, an approach that combines these pieces of information in the least biased way;
4. To identify whether the average European university lags indeed behind the average US university based on the set of twelve indicators of the THES and SJTU frameworks.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-23487-EN-C

ISBN 978-92-79-09704-1

Publications Office
Publications.eu.int