How can differences in international university rankings be explained?

Fredrik Niclas Piro* and Gunnar Sivertsen

Nordic Institute for Studies in Innovation, Research and Education (NIFU)

P.O. Box 2815 Tøyen, N-0608 Oslo, Norway

Phone: + 47 22 59 51 00 / 57

Fax: +47 22 59 51 01

E-mail addresses: Fredrik.piro@nifu.no, gunnar.sivertsen@nifu.no

***Corresponding author**

Abstract

University rankings are typically presenting their results as league tables with more emphasis on final scores and positions, than on the clarification of why the universities are ranked as they are. Finding out the latter is often not possible, because final scores are based on weighted indicators where raw data and the processing of these are not publically available. In this study we use a sample of Scandinavian universities, explaining what is causing differences between them in the two most influential university rankings: Times Higher Education and the Shanghai-ranking. The results show that differences may be attributed to both small variations on what we believe are not important indicators, as well as substantial variations on what we believe are important indicators. The overall aim of this paper is to provide a methodology that can be used in understanding universities' different ranks in global university rankings.

Introduction

Aggregating data on institutional performance into one score, thereby creating a league table for universities is the core idea of most university rankings. The rankings are supposedly making the world of universities more transparent. However, the rankings themselves are strongly lacking transparency with regard to the underlying data and how the methods are actually applied.

Ranking universities has been a phenomenon in rapid growth over the last fifteen years with an increased number of rankings being introduced and with increased attention given to these (Kehm & Stensaker, 2009). As well as much attention, the rankings have received much criticism (see e.g. Bookstein et al., 2010, Harvey, 2008, Liu & Cheng, 2005, Waltman et al., 2012). Among many problems highlighted in the critical literature on university rankings, we would like to address four: First, the rankings are heavily research-oriented, lacking data on the universities' teaching activities. Second, in their use of composite indexes there is little justification for both the selection of indicators, and the weights given to each of them, which seems arbitrary and not underpinned by theory in any way. Third, formula changes in the rankings may have strong influence on the results, which is not well communicated when results are presented. Finally, the huge emphasis on rank position, where the universities' positions may differ greatly based on statistical insignificant scores.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Nevertheless, there is a clear tendency that universities that perform at the top correlates between the rankings. It therefore seems appropriate to suggest that the rankings are able to identify *the* best universities in the world, although they use different indicators and calculation methods. But how much lower is really the quality of a university ranked 150th compared to a university ranked 175th? Whether a university ranked 70th is much better, or only marginally better, compared to number 75, is difficult to assess in most rankings.

Our study is about two of the best known and most used rankings (e.g. Harvey 2008, Waltman et al., 2012): Academy Ranking of World Universities (ARWU) conducted by the Center for World-Class Universities of Shanghai Jiao Tong University (and often referred to as the Shanghai-ranking), and the Times Higher Education World University Rankings (THE). In addition to the above mentioned criticism against rankings, these two are in particular being accused of their bias towards size dependency in ARWU (Waltman et al. 2012) and "take-for-granted reputational hierarchies" in THE (Harvey, 2008). Using a sample of Scandinavian universities where comparable background information is available to us, and where the universities are found in reasonably comparable research policy climates, making the comparison of the universities meaningful, we decompose the results of the rankings, in order to detect how the differences in the ranking positions are actually constructed. The novelty of our study is, firstly, that we set up an easy interpretable model for studying the relative importance of each indicator when comparing two or more universities' rank positions, representing a methodology that is useful regardless of which countries the universities are selected from. Secondly, using in-depth information from Nordic universities, we present to the best of our knowledge, the first analysis of university rankings that go beyond the ranking scores and illustrate how differences in the raw-data are translated into differences in ranking scores. The context of, and the data used in our study, are from the Nordic region, but we believe our findings may help understanding rank differences between universities in other countries as well.

The ARWU and THE rankings are based on different compositions of indicators. In THE, there were five indicator groups for the 2013 edition (Table 1):

Table 1: THE indicators and weights, 2013

30 %	**Research indicators (RES)**
18 %	Results from reputation survey
6 %	Research income (scaled against staff numbers and normalized for purchasing-power parity)
6 %	Research productivity (research output scaled against staff numbers)
30 %	**Research influence: Field normalized citation rates (CIT)**
30 %	**The learning environment: Teaching (TEA)**
15 %	Results from reputation survey
4,5 %	Staff-to-student ratio (as a proxy for teaching quality)
2,25 %	Ratio of doctoral to bachelor's degrees
6 %	Number of doctorates awarded scaled against number of academic staff
2,25 %	Institutional income scaled against academic staff numbers
7,5 %	**International outlook (INT)**
2,5 %	Ratio of international to domestic students
2,5 %	Ratio of international to domestic staff
2,5 %	Proportion of publications with international co-authors
2,5 %	**Research income from industry scaled against number of academic staff (INC)**

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Unlike THE, ARWU has no indicators that are composed of sub-indicators. In 2013, six indicators were included (Table 2). Not all Scandinavian universities are included in or given a rank position in THE and ARWU. We have selected a sample of the largest universities in Scandinavia for this study. Their ranks and scores in THE and ARWU are shown in Tables 3 and 4.

Table 2: ARWU indicators and weights, 2013

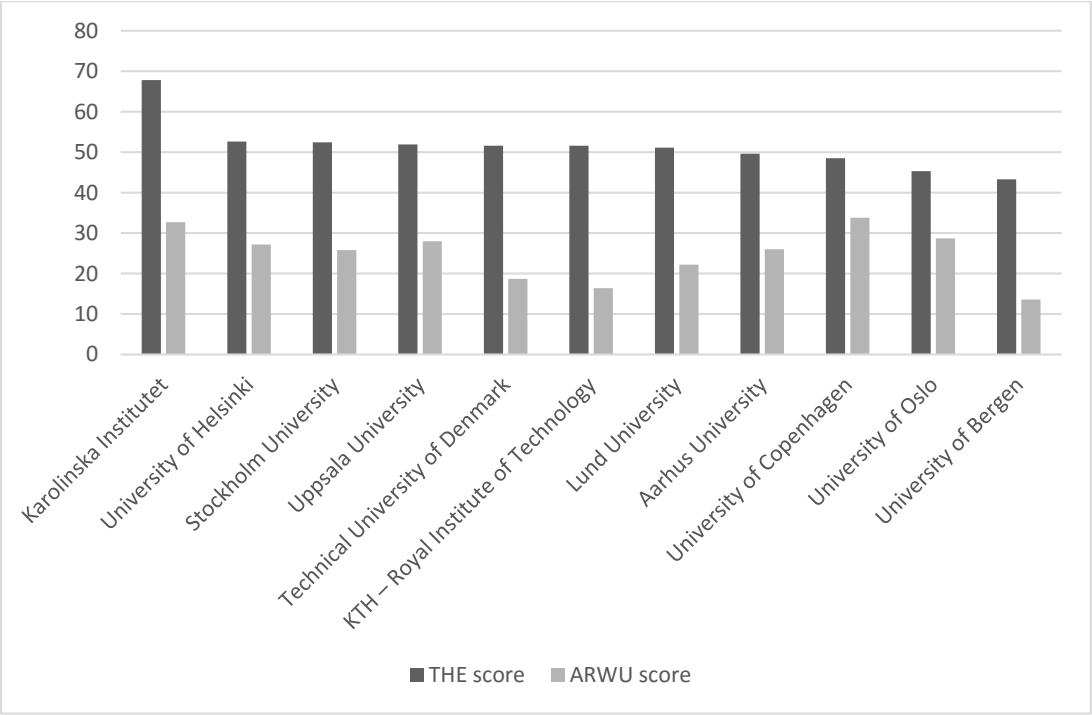
20 %	**AWA:** Number of researchers who earned a **Nobel Prize** in physics, chemistry, medicine or economics and/or the **Fields Medal** in Mathematics since 1911
20 %	**HiCi:** Number of **highly cited** researchers in the fields of life sciences, medicine, physics, engineering and social sciences
20 %	**N&S:** Number of articles published in **Nature and Science** during the last five years
20 %	**PUB:** Number of publications in **Web of Science**
10 %	**ALU:** Number of alumni who earned a **Nobel prize or Fields Medal** since 1901
10 %	**PCP:** The weighted score of the above five indicators divided by the number of full-time equivalent academic staff[1]

The focal point of our decomposition will be two Norwegian universities: the University of Oslo (UiO) and the University of Bergen (UiB). We ask why these two universities were *lower* ranked than many other Scandinavian universities in 2013. Hence, we have only included those universities that were in general higher ranked in 2013 than UiO and UiB: five Swedish universities, three Danish universities and one Finnish university. In THE, the Scandinavian differences in total scores are strong (Figure 1), with Karolinska Institutet being in a class of its own, but the difference between University of Helsinki (second highest ranked Scandinavian university) and the University of Copenhagen (ranked 9th) is not very big – just 4.1 points separating a university ranked 100th and a university 150th.

[1] It is not clear to us how ARWU has identified the Nordic data on FTEs, as only Sweden is specifically listed as a country where ARWU has obtained such information. If the number of FTEs cannot be obtained, the weighted scores of the other five indicators are used.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Figure 1: Eleven Scandinavian universities' scores in THE and ARWU, 2013



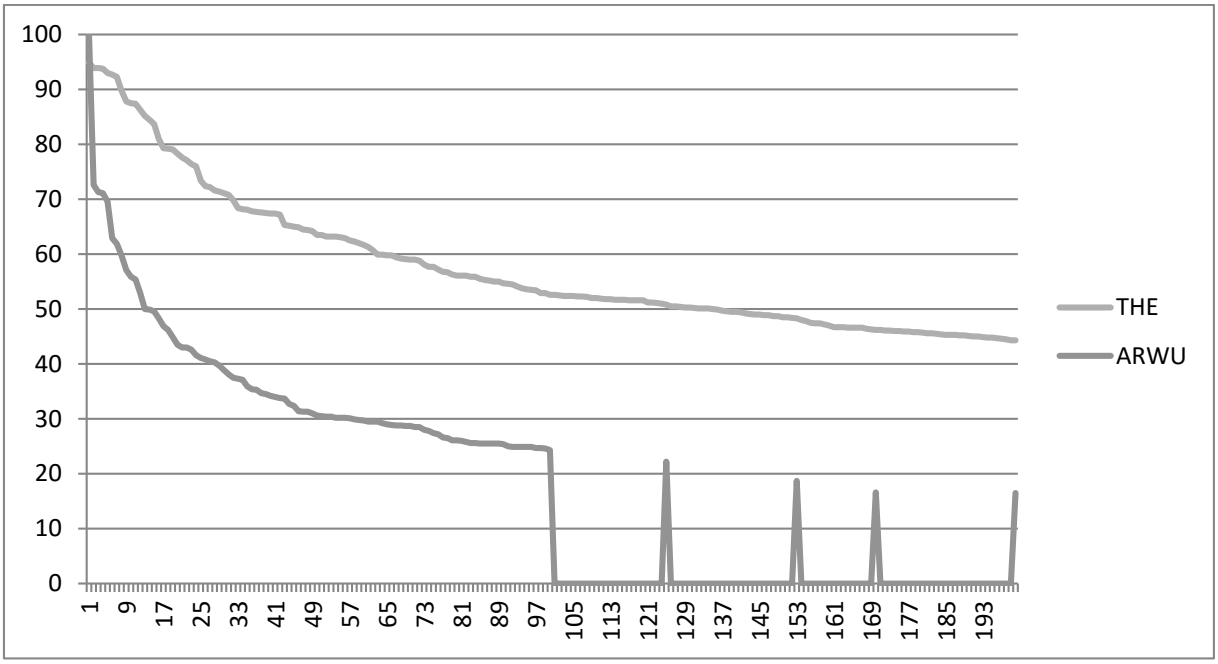
The differences in ARWU scores appear to be much smaller than in THE. The ranking order is also different, with the University of Copenhagen ranked first among the Scandinavian universities (only 9th in THE), and UiO now ranked third (while 10th in THE). However, only looking at the rank numbers hide a lot of information. Both rankings are highly skewed, i.e. the farther down a ranking a university appears, the smaller the difference contra surrounding universities (Figure 2).

In ARWU, total scores are only available for the first one hundred universities, but based on a previous analysis on the background data on these indicators for the same universities (Piro et al., 2014), we may plot the Scandinavian data and identify a trend line below position 100 even here (KTH – Royal Institute of Technology was ranked 201-300 in ARWU with 16.4 points, so that we know that the university ranked 200th has at least 16.4 points).

In THE, 30.7 points separate Caltech at the top and Pennsylvania State University, ranked 49th. Between Boston University (50) and University of Helsinki (100) only 10.9 points differ. After that, there is 3.8 points between rank 101 and 149, and 4.2 points between rank 150 and 199. While there is a gradually declining curve in THE, there is an immediate drop in ARWU after Harvard at the top (100 points) and Stanford University ranked second with 72.6 points. The distance from Harvard to Vanderbilt University (49th) is 69 points in ARWU, while the distance from the Technical University of München (50) to University of Freiburg (100) is just 6.3 points. The trend is obvious in both rankings; the distance between the universities gets smaller the farther from the top we get. We therefore find it useful to try identify exactly what separates universities typically ranked lower down the hierarchies of these rankings.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Figure 2: Relative differences in scores between universities in THE and ARWU, 2013



Data and methods

Our method is to decompose the rankings, i.e. explain the relative importance of each indicator and to analyze size differences in the background data that cause these effects in the rankings. We use Nordic universities to exemplify, because we have the needed data for these institutions to conduct the decomposition.

Table 3: Scandinavian universities in ARWU, 2013

No.	University	SUM*	ALU	AWA	HiCi	N&S	PUB	PCP
42	University of Copenhagen	33.8	22.7	19.0	25.9	30.3	60.2	35.0
44	Karolinska Institutet	32.7	23.3	26.8	32.4	20.9	49.2	35.9
69	University of Oslo	28.7	18.5	32.8	17.7	16.5	50.6	25.6
73	Uppsala University	28.0	18.5	27.6	14.5	24.7	47.9	24.7
76	University of Helsinki	27.2	13.1	16.4	24.0	19.9	51.1	28.8
81	Aarhus University	26.0	12.0	22.2	7.2	25.6	50.9	28.5
82	Stockholm University	25.8	25.1	27.6	16.2	19.1	38.3	23.5
101-150	Lund University	22.2	22.0	0.0	25.1	14.9	49.7	20.3
151-200	Technical University of Denmark	18.7	5.3	11.6	17.6	14.0	37.4	20.8
201-300	KTH – Royal Institute of Technology	16.4	0.0	15.0	7.2	9.2	39.6	22.1
201-300	University of Bergen	13.6	0.0	0.0	7.2	14.6	36.8	18.4

*Final scores in ARWU (SUM) is only publically available for the top 100 universities. We have estimated the final score for universities ranked 100th and onwards, based on ARWU's description on how the final points are summed.

The decomposition follows some basic principles. We are explaining why a university is ranked above or below another university. The different rank positions between UiO and UiB and other universities

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

can be explained by indicators causing surplus or deficit in rank. For example, UiO scores better than the University in Copenhagen on *one* indicator in ARWU: the Award category (AWA), measuring Nobel prizes/Fields medals (Table 3). University of Copenhagen scores better on all other indicators, and these indicators explain why University of Copenhagen is above UiO (by 5.1 points). Without the AWA indicator, i.e. if the University of Copenhagen would have had an equal score to UiO on AWA, then it would have been 6.3 points above UiO. Hence, a difference of 6.3 points should be explained by analyzing the five other indicators (ALU, HiCi, N&S, PUB and PCP).

The first step of the decomposition is to weight the differences in scores on these indicators, according to their percentage weights of the total score in ARWU (where ALU and PCP only is weighted half as much as HiCi, N&S and PUB). Thereafter, we calculate how much of the 6.3 points difference these indicators represent. The final step is to calculate how much of the initial difference (5.1 points) that each of these indicators actually explain. An illustration is given in Table 4.

Table 4: Illustrating the decomposition of ARWU: UiO versus University of Copenhagen

	SUM	ALU	HiCi	N&S	PUB	PCP
Weight of indicator	**0.8**	**0.1**	**0.2**	**0.2**	**0.2**	**0.1**
Difference in points: UiO vs. University of Copenhagen	**5.1**	**4.2**	**1.3**	**13.8**	**9.6**	**9.4**
Weighted difference: UiO vs. University of Copenhagen	6.3	0.42	0.26	2.76	1.92	0.94
Percentage of weighted difference	100	6.67	4.13	43.81	30.48	14.92
Points of initial difference explained	5.1	0.34	0.21	2.23	1.55	0.76

In this example, 2.2 (out of 5.1) points separating UiO and University of Copenhagen is explained by Copenhagen's higher score on publications in Nature and Science, while 1.55 points are explained by more publications in Web of Science (WoS) from Copenhagen. The remaining three indicators explain much less of the difference (in the range 0.21 to 0.76 points). The same procedure is then applied on the results from THE (Table 5). It is not possible to decompose THE as easily as ARWU, because most indicators in THE are aggregated scores from several sub-indicators – whose scores are unknown to us. This seems to fit the disproportional distributions of studies we have been able to find about the two rankings, where ARWU is far more often scrutinized, arguably because THE appears to be impervious. Some background data do exist, however, and in light of this it is possible to produce estimates about which sub-indicators matter – and especially whether it is THE's reputation survey that makes the difference between Scandinavian universities (see Piro et al., 2014 for a presentation of the raw data being used).

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Table 5: Scandinavian universities in THE, 2013

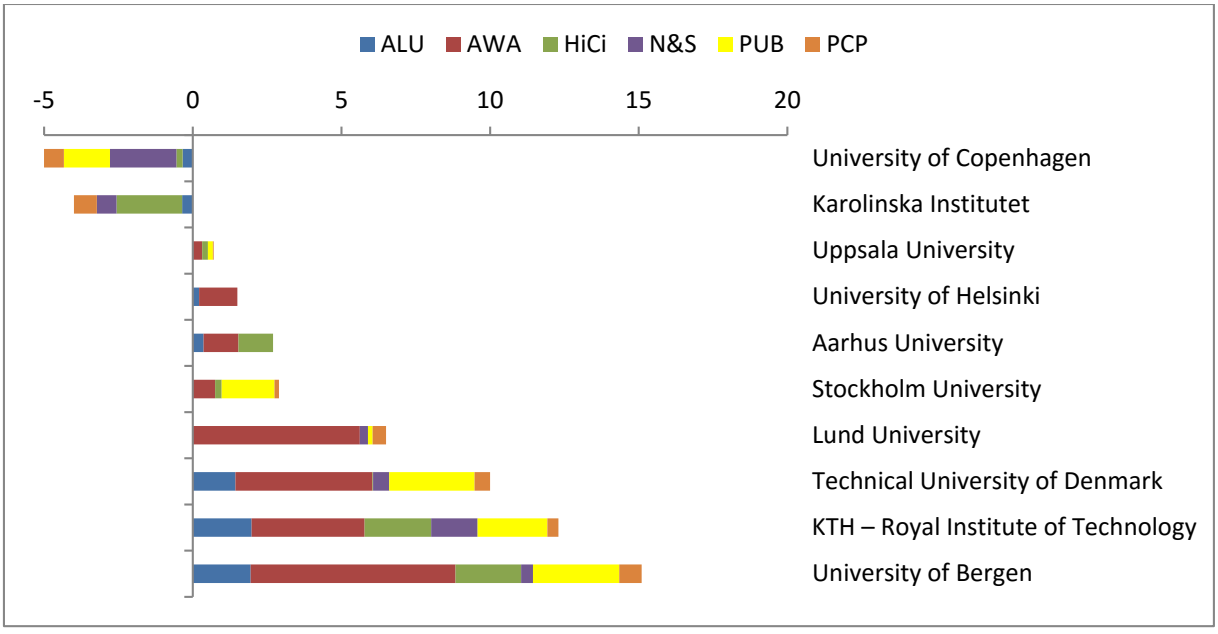
No.	University	SUM	TEA	INT	INC	RES	CIT
36	Karolinska Institutet	67.8	58.1	73.2	68.7	67.7	76.0
100	University of Helsinki	52.6	35.3	51.3	30.7	46.5	77.8
103	Stockholm University	52.4	28.6	53.4	31.5	41.4	88.7
111	Uppsala University	51.9	38.9	57.8	40.5	49.7	66.4
117	Technical University of Denmark	51.6	38.9	77.4	98.7	26.9	78.7
117	KTH –Royal Institute of Technology	51.6	45.0	86.2	100.0	41.6	55.6
123	Lund University	51.1	30.1	67.4	33.5	48.6	72.0
138	Aarhus University	49.6	30.7	67.2	67.5	46.4	65.8
150	University of Copenhagen	48.5	33.9	76.0	43.0	32.0	73.1
185	University of Oslo	45.3	32.4	70.3	32.5	33.0	65.4
201-225	University of Bergen	43.3	30.8	62.9	34.7	28.2	66.6

In the next sections we will use the universities in Oslo and Bergen to demonstrate how in-depth insights into the rank differences can be created along with a skeptical attitude to the use of the rankings as strategic information for the institutions.

Results

We begin decomposing ARWU. In Figure 3 we show how relative differences in the ranking can be explained by the decomposition methodology from the point of view of one institution – in this case the *University of Oslo* (UiO) who is ranked relatively high in ARWU (69th) – only the University of Copenhagen (*Copenhagen*) and Karolinska Institutet (*KI*) are ranked higher in the Scandinavian countries. They have higher scores than UiO on all indicators except AWA, measuring Nobel prizes and Fields medals.

Figure 3: University of Oslo's distance to Scandinavian universities in ARWU, 2013



Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Overall, AWA is the most important explanation for UiO's high Scandinavian position in ARWU, although UiO only has three awards (two in economics (1969, 1989) and one in chemistry (1969). These three awards is the only reason why UiO is ranked above University of Helsinki, and the most important explanation related to all other universities' lower rank (except Stockholm University). The other reason why UiO is ranked third among Scandinavian universities, is its high score on publications in WoS (PUB), but this is a size-effect, as UiO is the third largest Scandinavian university, measured by number of academic staff.

UiO is 5.1 points behind Copenhagen and 4 points behind KI. UiO's lower score on the alumni indicator and the weighted size indicator PCP are not important in explaining this, as both indicators explain less than ten per cent of the difference. But Copenhagen scores substantially better than UiO on two indicators: N&S (Copenhagen has 30.3 points and UiO 16.5 points) and PUB (Copenhagen has 60.2 points, while UiO has 50.6). These two indicators explain about 75 per cent of the difference between Copenhagen and UiO in ARWU. How can we explain these differences? In 2012, Copenhagen had 114 publications in Nature and Science, while UiO had 39. In ARWU, the publications are weighted by number of authors, listing of authors etc., and our reconstruction of this method gives Copenhagen 51.5 N&S-points (which is not the same as the 30.3 points they get on the indicator in the ranking, as this is standardized by Harvard's results), while UiO gets 14.8 N&S-points. Hence, Copenhagen has 2.9 times the publications UiO has in Nature and Science, resulting in 3.5 times as many N&S-points. Using the more intuitive *number of publications* there is a difference of 75 publications (114 vs. 39) that explains 43.8 per cent of the difference in total score between UiO and Copenhagen.

About 30 per cent (i.e. 1.6 points) of the difference between UiO and Copenhagen (5.1 points) is explained by the PUB indicator, where Copenhagen had 5267 and UiO 3435 publications in 2012 (please note that ARWU weights the publications by number of authors etc., but for simplicity reasons we use the raw numbers here). This means that Copenhagen had 1.53 times as many publications as UiO did – but also 1.50 times the number of scientific staff (5022 vs. 3344), so that the difference on this indicator is almost 100 per cent explained by a size difference – not productivity differences. If UiO was to equal Copenhagen on this indicator, the scientific productivity per staff member would have to increase far above the level in Copenhagen or UiO would have to appoint more staff members; approximately 700-800 man-years, as this would have halved the difference on the PUB-indicator and most likely have led to more N&S-publications at UiO (thereby possibly also improving the HiCi-score).

There are different reasons for why Karolinska Institutet (KI) is ranked above UiO. KI has a slightly higher score at the alumni indicator, but overall this is not an important explanation. KI also has a higher score at the PCP-indicator, which must be seen as a multiplicative effect from a higher score on several indicators with fewer staff members (1952) compared to UiO (3344). This is reflected in UiO's better score at the PUB-indicator (although only marginally). Thus the researcher productivity is higher at KI compared to UiO. This is hardly a surprise as KI almost exclusively publish in journals indexed in WoS, while UiO's portfolio includes many fields with low coverage in WoS (e.g. social sciences, arts and humanities). In addition, KI has a much higher number of highly cited researchers and publications in Nature and Science compared to UiO.

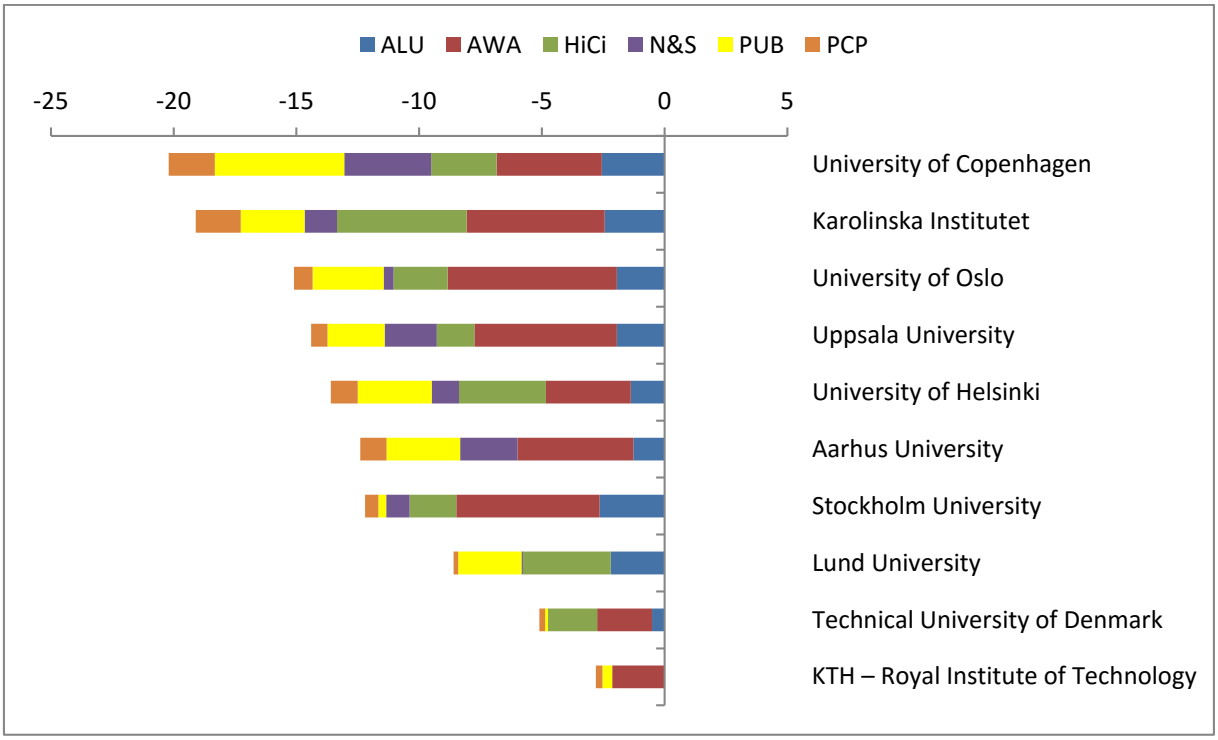
The difference between UiO and KI in ARWU is primarily explained by one indicator, HiCi, where KI had 19 researchers at Thomson Reuters' list, while UiO had six. This means that an additional 13

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

persons at this list explains 55 per cent of the difference in ARWU between these universities. 16.5 per cent is explained by an additional 11 publications in Nature and Science from KI in the period 2008-2012. While institutional size mattered in the case UiO vs. Copenhagen, it is a small number of highly cited researchers and an extra two publications per year in Nature and Science that explains why KI is ranked above UiO. Cf. the Leiden-ranking where Copenhagen has a higher field normalized citation index than UiO, these differences may be seen as a more structural phenomenon, i.e. that KI's research is more cited than UiO's. The common denominator for all universities ranked below UiO is UiO's higher score on AWA and the size-related indicator PUB (except for the universities in Helsinki and Lund), but AWA explains much more of the variation compared to PUB.

University of Bergen (UiB) is ranked in the field 201-300 in ARWU in 2013. While 5.1 points differed between Copenhagen and UiO, there are 20.2 points separating Copenhagen and UiB (Figure 4).

Figure 4: University of Bergen's distance to Scandinavian universities in ARWU, 2013



Most universities perform better than UiB on most indicators, and especially the AWA-indicator, which explains most of UiO and Stockholm University's better total score compared to UiB. 60 per cent (8.8 out of 15.1 points) of the difference between UiO and UiB is due to UiO's three Nobel prizes, and the university's alumni score. Keeping AWA and ALU aside (since UiB does not have any points on any of these), we may try to explain the other indicators in sheer numbers. On the three indicators HiCi, N&S and PUB – UiB receives 11.5 points less than Copenhagen (representing 57 per cent of the total difference). This is due to an additional eleven highly cited researchers in Copenhagen (12 vs. UiB's one person), 92 more publications in Nature and Science (114 against UiB's 22; equaling 18.4 per year), and a substantially higher number of WoS publications: Copenhagen had 5267 publications – UiB had 2008. The researcher productivity, however is identical, i.e. the difference at the PUB-indicator is completely a matter of size – not productivity.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Copenhagen has a higher share of publications in Nature and Science per staff member, and more highly cited researchers than UiB. Together these two indicators explain 30 per cent of the variation to UiB. This means that 70 per cent of total difference between UiB and Copenhagen is due to a higher number of staff members in Copenhagen and the fact that UiB (or its former students) has no awards. However, the last Nobel prize awarded to Copenhagen was in 1975, while former students won in 1997, 1984 and 1975.

For all universities ranked above UiB; AWA and ALU explains much of the difference (from 76 per cent for Royal Institute of Technology (KTH) to 26 per cent for Lund University). HiCi, N&S and PUB do not explain as much of the variation as the former two. What we find most interesting about these indicators, is where the score differences more or less can be completely ascribed to size. Not only Copenhagen's higher score on PUB is a matter of size; we find the same for Stockholm University and UiO (where 19 per cent of the variation is due to PUB). These universities have exactly the same researcher productivity as UiB. We then find two universities with a higher PUB score than UiB, despite a lower researcher productivity – which is truly a size effect. University of Helsinki's PUB score is 51.1 (UiB's is 36.8). In absolute numbers, University of Helsinki had 3456 publications to be divided by 4190 staff members (ratio of 0.8), while UiB with 2008 publications and 2064 staff members had a ratio of 1.0. Despite University of Helsinki's lower productivity, this university gets 14.3 points more than UiB at the PUB-indicator, explaining 22 per cent of their higher total score compared to UiB. Aarhus University also has a lower productivity rate than UiB, but nevertheless a higher PUB-score, explaining 24 per cent of the variation to UiB. There are, however, four Swedish universities with higher PUB-scores that are not attributable to a size effect: KI (ratio 2.1 per staff member), Uppsala University (1.3), Lund University (1.3) and KTH (1.4). These differences explain between 30 per cent (for Lund University) to 14 per cent (for KI) of the variation. These Swedish universities also have higher scores, both in total and per staff member, on the HiCi and N&S indicators.

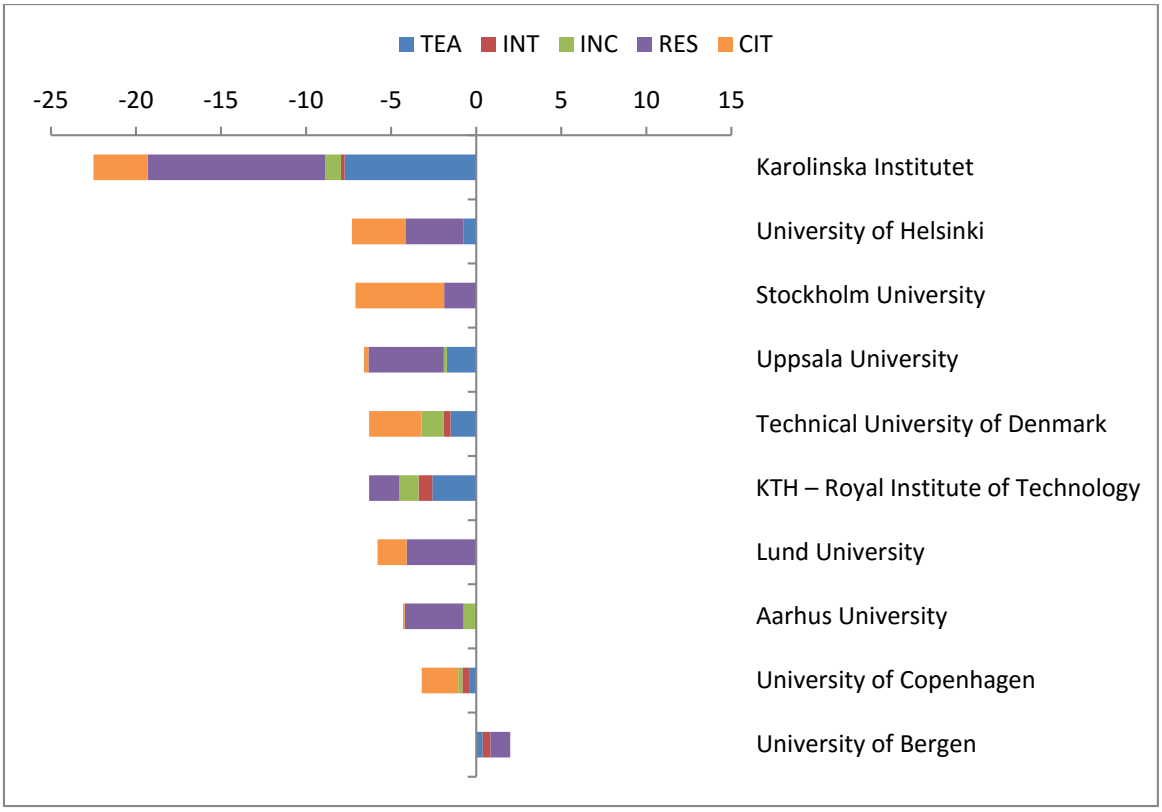
Compared to UiB, KI has 2.5 times as many publications in Nature and Science per staff member (which may be partly explained by different scientific profiles of the two universities) and almost ten times as many highly cited researchers per staff member (19 vs. one at UiB). These two factors – together with PUB – explains 9.2 of 19.1 extra points at KI – and must be considered independent of size. Similar size-independent results are also found for *Uppsala University* with 2.7 times as many publications in Nature and Science per staff member and 1.5 times as many highly cited researchers per staff member (four against one at UiB) (these two factors (and PUB) explain 5.9 of 14.4 extra points at Uppsala University); also at *Lund University* with 1.8 times as many publications in Nature and Science per staff member and 4.3 times as many highly cited researchers, explaining 6.3 out of 8.6 extra points on UiB. The main reasons for UiB's relative low rank in ARWU thus seems to be low scores on ALU and AWA, but also fewer highly cited researchers and publications in Nature and Science compared to most Scandinavian universities analyzed here. UiB has a lower PUB-score than all universities, but as we have argued; in many cases this is a matter of size.

In *Times Higher Education (THE)* UiO was ranked 185th in 2013. Scandinavian universities ranked above UiO were in general - KI aside - not many points above. KI was ranked 36th with 67.8 points, followed by eight universities ranked between 100th and 150th, from University of Helsinki (52.6 points) to University of Copenhagen (48.5); only 4.1 points separated two universities with fifty positions between the two in the ranking.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

There are no indicators in THE where UiO's scores are especially high in a Scandinavian context. Its lowest score is on the indicator measuring industry income, but this indicator only decides 2.5 per cent in total and hardly explains anything of the Scandinavian differences (Figure 5). The internationalization indicator is also of little significance, partly due to small Scandinavian differences on the indicator, and partly because the relative weight of the indicator is very low. UiO's score on this indicator is relatively high, it is only KTH with a higher Scandinavian score (which explains 0.8 out of 6.3 points separating the two universities).

Figure 5: University of Oslo's distance to Scandinavian universities in THE, 2013



Except for KI, UiO's distances (positive and negative) to other universities are relatively small (Figure 5). The variation is primarily found in the three main indicators TEA, RES and CIT (which constitute 90 per cent of the total score in THE). The citation indicator is not important in explaining why Uppsala University and Aarhus University are above UiO, but UiO's citation index is significantly lower than that of the Technical University of Denmark (DTU) and Stockholm University (which has the highest Scandinavian CIT score in THE). For these two universities (49 and 74 per cent respectively), and Copenhagen (67 per cent), CIT is by far the most important explanation why UiO is ranked below these universities. CIT is not equally important in explaining KI's higher rank. Only 14 per cent of total score difference between UiO and KI is explain by CIT. On average, KI's publications are cited 26 per cent above the world average, while UiO's are cited 12 per cent above world average.

Most universities that rank above UiO, do so because they score better on two or three of the three main indicators: research, teaching and citations. The only two exceptions are the technical universities KTH and DTU, where about 20 per cent of their higher total scores can be attributed better scores on research income (but they only explain 1.3 and 1.2 points respectively of a difference of 6.3 points to UiO).

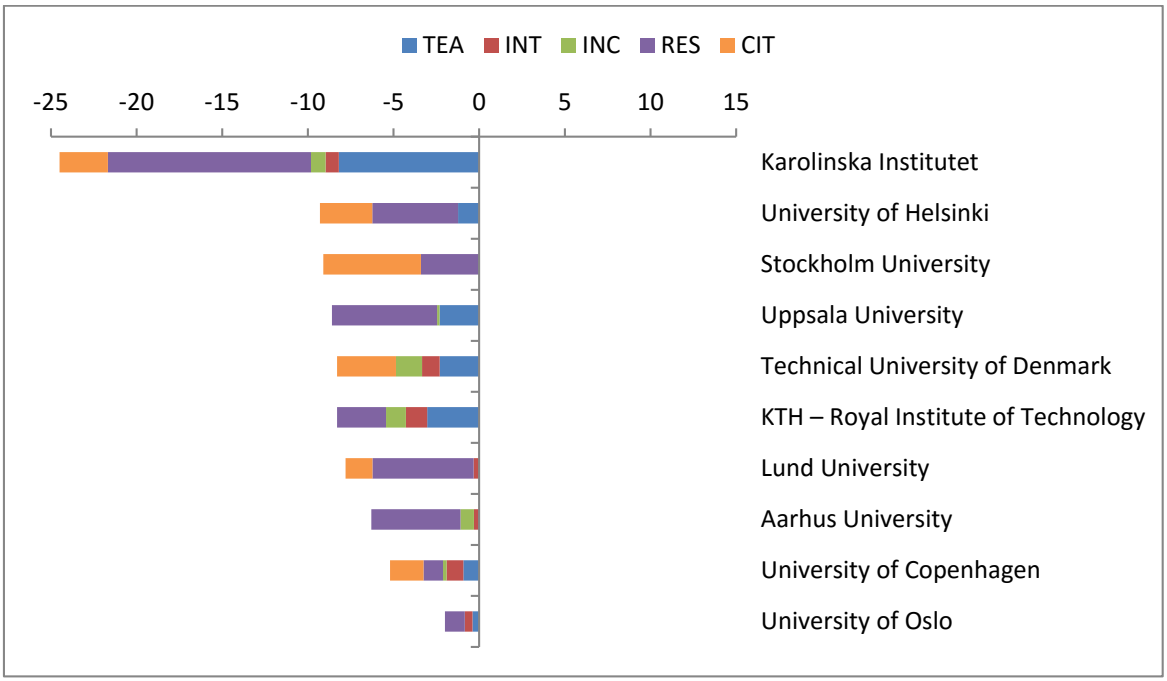
Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

What is unexplained after CIT, INT and INC is attributable to differences in teaching and research indicators. However, these are almost impossible to decompose, as they are the sum of many sub-indicators where the single most important is the reputation survey, whose results are unavailable. UiO scores better at the TEA indicator than the universities in Stockholm, Lund and Aarhus. But TEA is an important explanation for the difference in total scores between UiO and KTH (41 per cent), KI (34), Uppsala University (26), DTU (24), Copenhagen (13) and University of Helsinki (10). DTU is primarily ranked above UiO because of a better student-staff ratio. UiO's high score on the sub-indicator number of doctorates awarded scaled against number of academic staff, was the reason why these universities do not score even more compared to UiO: with 0.15 doctorates per staff member, UiO is only marginally behind KTH (0.16) in Scandinavia. Similarly, much of the difference in total scores stems from the research indicator: Aarhus University (80 per cent), Lund University (70), Uppsala University (67), University of Helsinki (47), KI (46), KTH (28) and Stockholm University (27).

In general, in explaining UiO's 185th place in THE, KI is ranked higher because of better RES and TEA scores. For DTU and the universities in Helsinki, Stockholm and Copenhagen, it is mainly higher CIT scores that separates these from UiO. The universities in Uppsala, Lund and Aarhus are mainly benefitting from higher RES scores, while KTH scores better than UiO on all indicators except CIT.

University of Bergen (UiB) is not on THE's top 200 list, rather in place 201-225 with 43.3 points; close to UiO at 185th place with 45.3 points. UiB scores relatively high at the CIT indicator (ranked 7th among our selected universities), but weakly on INC and RES in a Scandinavian context. The main reason for UiB's low rank is demonstrated in Figure 6. It is foremost RES that separates UiB from other Scandinavian universities. Stockholm University, but also KI, University of Helsinki and DTU are ranked above UiB due to better CIT scores. It is only for KI (and to some degree KTH, DTU and Uppsala University) where the TEA indicator explains much variation to UiB.

Figure 6: University of Bergen's distance to Scandinavian universities in THE, 2013



Two universities (Stockholm University and DTU) are characterized by both being ranked at the very top (in number of points) at THE's CIT indicator and by having their high CIT scores as their by far

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

most important advantage compared to UiB (explaining 63 per cent of the score difference for Stockholm University and 42 per cent for DTU). CIT explains 37 per cent of the variation for Copenhagen, but because the difference is small, this means that only 1.9 points difference in the ranking is explained by CIT. For KI it is the opposite: CIT explains only 12 per cent of the difference; but nevertheless 2.8 points. The INT and INC indicators do not explain much, except DTU and KTH (some 30 per cent of the variation), and UiO and Copenhagen (22 per cent).

The most important indicator – RES – explains between 34 and 83 per cent of the variation between UiB and the universities in Helsinki, Stockholm, Uppsala, Lund and Aarhus, and KI and DTU. RES also explains 22 and 58 per cent respectively of Copenhagen and UiO's higher scores, but total differences here are small, so that this in reality only explains about one point. By comparison, RES explains 11.9 of KI's 24.5 extra points to UiB (biggest score difference) and 5.2 out of 6.3 extra points for Aarhus University (highest percentage explanation: 82.7 per cent). Again we need to ask whether it is the reputation survey that explains this, or if it is researcher productivity or research income that matters. As for UiO, UiB gets a low score for research income, but UiB's productivity of 0.97 publications per staff member is roughly the same as UiO, Copenhagen and Stockholm University's.

All universities with a higher RES score – except UiO – has some of the variation explained by UiB's low score on the purchasing power adjusted income indicator. At the productivity indicator, UiB receives 79 points by THE, and only four universities (KI, KTH, Uppsala and Lund) has a substantially higher researcher productivity than UiB (thus explaining some of the variation to UiB). With exception of KI who scores much higher than most universities on these two indicators (and KTH scoring well on the income indicator), the impact of these two indicators is low, i.e. they explain quite little of the variation. Hence, most of UiB's difference to other universities at the RES indicator must be subscribed the reputation survey. The TEA indicator does not explain as much as RES of why other Scandinavian universities are ranked higher. This indicator consists of more sub-indicators then RES, and the reputation survey means less than in RES (50 vs. 60 per cent). It is not one specific indicator that explains UiB's position. KI is ranked above UiB due to both RES and TEA, while CIT is the main reason why Stockholm University, Copenhagen and DTU are ranked higher.

In Piro et al. (2014) the importance of the reputation survey in explaining differences in the teaching and research indicators in THE for the same universities analyzed here was discussed. When comparing the raw data that is used to provide scores on all sub-indicators for TEA and RES, it was argued that the reputation survey is a very strong explanatory factor, as differences on other sub-indicators (making up 40 per cent of RES and 50 per cent of TEA) are not big enough to account for all variation. For UiO, the reputation survey does not seem to be as important in explaining differences at the teaching indicator. The exceptions are KI, KTH and Uppsala University, where we believe that most differences can be ascribed the reputation survey (for example, KI scores drastically better than UiO (and all other universities) at the student-staff ratio indicator, but considering this indicator's relative low weight, this is not an important explanation compared to UiO). In short, UiO's lower TEA score compared to KI, KTH, Uppsala University, Helsinki University and University of Copenhagen was believed to be a result of a lower score on the reputation survey. In general, in explaining UiO's 185th place in THE, KI is ranked higher because of better RES and TEA scores, where the reputation survey in both cases seem to be the driving factor. The universities in Uppsala, Lund and Aarhus are mainly benefitting from higher RES scores, where the reputation survey explains most. Lastly, KTH's teaching reputation is the most important reason for its higher rank.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Similarly, at the University of Bergen, the reputation survey seemed to explain five universities' higher TEA scores: KI, KTH, Uppsala University, University of Copenhagen and University of Helsinki (the latter two have only marginally better RES score, so the total effect is small). The reputation survey also seemed central in understanding KTH and Uppsala University's higher scores on RES. The universities in Helsinki, Lund, Aarhus, Stockholm and Oslo also got better RES scores primarily because of the reputation survey; since differences in underlying data for the other indicators could not account for strong enough variation between UiB and these universities.

Discussion

What do ranking differences tell us?

The two rankings that we have studied here differ in many ways. ARWU's objective is to compare research excellence, whereas THE has a more comprehensive approach including measures of teaching, internationalization and industrial collaboration as well.

ARWU is an extremely elitist ranking, where most of the results depend upon awards and publications from a handful of researchers at each university. While ARWU in this sense is very oriented towards the individual level, it is also very preoccupied with institutional size, which is seen as a value in itself (regardless of the productivity of the staff). THE has a much broader and institutional focus, with several indicators measuring the composition of staff and students, as well as the universities' relation to the business community etc. Most indicators are weighted by size to some extent. But despite the large number of indicators supposed to cover all of a university's core activities, most of these are given a low weight when the final scores are calculated. Based our findings we argue that the rank differences between the highest ranked Nordic universities are mostly not of a substantial kind (i.e. very small differences in indicator scores), but rather based on either size differences or differences in historical or path-dependent features such as Nobel Prizes or on what has been described in several studies as the 'anchoring effect', i.e. that the rankings themselves guide future assessments of quality in reputation surveys (Bowman & Bastedo, 2011). Safón (2013) found that both ARWU and THE, seem to be based on a common (and strongly correlated) underlying factor (which is a bit more complex for THE), which is the universities' age, scope, activity in hard sciences, whether US-based/English-speaking country, annual income, orientation towards research, and reputation (which seemed to be the most important variable).

In a Scandinavian context, the Norwegian universities are weakly ranked in THE. Two central factors in explaining this are lower citation indexes and low scores at the research indicator (where we believe the reputation survey explains a large share of the differences). Although we believe that the differences must be based on a limited number of votes, it still remains a self-critical question from a Norwegian point of view, why other universities systematically receive (some) more votes. However, only the former finding conform with data available elsewhere: 1) Norwegian universities are generally lower cited than other Scandinavian universities, 2) Norwegian universities produce fewer researchers with the ability to publish in Nature and Science, or to accumulate enough citations to be included in Thomson Reuters' lists of the world's most cited researchers. With these two – and to various degree – important conclusions, we see that the rankings may tell us something about indicators related to scientific excellence, i.e. research. They do not, however, tell us much (or

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

anything at all?) about differences related to teaching quality at the universities. In general, differences in rankings do not give any valuable strategic information before the ranks are decomposed into the actual indicators that they are constructed from.

Changes may be the result of statistical noise or exceptional (and rare) events

In reality, there are two factors that explain variations in THE: a university's citation index and its number of votes in THE's reputation survey. Unfortunately, THE does not make the results from this survey public, i.e. we do not know how many votes a university has received (is it five, ten or two hundred?). Since the universities' scores on this indicator are based on numbers normalized according to the number of votes that Harvard University receives, extremely small differences in number of votes between two Scandinavian universities outside top-100 in the ranking may be crucial to their final rankings (because the normalization procedures may lead to small differences being blown up when the total scores are calculated). Several studies point to changes in universities' rank from one year to another, just being the result of statistical noise (Gnolek, Falciano & Kuncl, 2014; Dichev, 2001). Related to this is changes in scores that may completely be ascribed changes in the formulas of the rankings, e.g. Bookstein et al. (2010), who demonstrates what they call "unacceptably high fluctuations" from year to year, such as when University of Copenhagen's score on the student/staff ratio in THE went up from 51 to 100 from 2007 and 2008, which they claim is practically impossible as this ratio must be stable in any academic institution above a certain size.

Most indicators in ARWU are stable over time. In the period 2008-2013 there have been only minor changes in how Scandinavian universities are ranked. It is a ranking where it is difficult to change position. The exception is if a university should receive its first Nobel prize: when London School of Economics received its first prize in 2010 it went from place 201-302 to 102-150 where it has been since, and when the Norwegian University of Science and Technology (NTNU) won its first prize in 2104, it sent the university up from place 201-300 to 101-150 (but dropped from place 251-275 to 276-300 in THE just a few weeks later). (Similarly, a university with a low number of publications in Nature and Science would boost its score on this indicator in ARWU with just a modest increase in publications in these journals.) We have observed a clear divide between the Scandinavian universities with and the ones without a Nobel prize. It is easy to mock the lack of relevance today that prizes awarded in the 1920s have, but from a Norwegian point of view, it is the three Nobel prizes of UiO's, that keeps it inside the top-100 in ARWU. Similarly, lack of awards is a key factor to UiB's modest position. Interestingly, one of NTNU's two Nobel laureates in 2014, said that the he did not expect to win the prize before he was at least 80 years old (i.e. in some 28 years, probably based on a presumption that most laureates are awarded a long time after their research breakthroughs). If so, NTNU would have been ranked 201-300 in ARWU for yet a very long time, even though this research would have already been taken place.

Unlike ARWU, the universities may change their positions in THE drastically over a relative short time period. Remarkable changes from year to year are easy to spot. A good example is Uppsala University, who climbed 60 positions from 2010 to 2011 (from rank position 147 to 87). The reason was primarily a change in their citation score from 40.7 to 50.3, which is difficult to understand since this indicator is based on a 5-year average. In Norway, UiB has over time climbed down in THE. From 135th place in 2010 to 201-225 in 2013. In this period, UiB's teaching score has dropped from 40 to 30

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

points and the research score from 42 to 28 (UiB experienced a small reduction in their citation index that year). What these indicators have in common is the dependency on the reputation survey.

The reputation survey alone decides 33 per cent of the total score in THE, and the number of votes are kept hidden. The respondents are asked to name up to 15 universities that they consider the best in the world with regards to research and teaching. We know that Harvard University gets the most votes, but not how many (neither do we know how many universities on average the respondents list). In 2013, MIT got the second highest number of votes; 87.6 per cent of Harvard's votes. Purdue University at number 50 in the reputation survey got 6.5 per cent of Harvard's votes, i.e. the number of votes rapidly decrease the farther from Harvard we get. THE only publish results for the 50 best universities in the reputation survey. The argument is that it is not appropriate to provide more information, as differences are small between these universities. We can come up with at least two other good reasons why the data remain hidden. Firstly, that it would appear strange that a ranking would be so dependent on an indicator where most universities in the world score as disproportionally low as they do. Secondly, and related to the low number of votes that most universities receive, that the results are subject to strong annual fluctuations, where a very few extra or lost votes may have a great impact on a university's total score. It would indeed be interesting to know the number of votes that each university actually receives. For the universities from Norway it certainly must be low, given the fact that the scores they received are ridiculously low: UiO got 12 points, UiB 6, Norwegian University of Science and Technology 10 and University of Tromsø 2 (reprinted with permission from the universities).

Size is an important explanation, but not subject to changes in the short term

Another reason why it is difficult to change position in ARWU – at least among the universities that we have studied – is ARWU's size dependency. Docampo & Cram (2015) found that around 30 per cent of the variation in the ARWU indicators are attributed to variation in university size. It is true that ARWU has an indicator (PCP) that adjusts for size, but the impact of this indicator on final scores hardly makes any difference. Next to the Nobel prizes, it is the size-dependent indicator PUB that explains most of the differences between Scandinavian universities. We have observed that even though a university has a lower researcher productivity, it still gets more points on this indicator. This criticism is of course not directed at THE who do not have any indicators that resembles ARWU's PUB indicator, but is rather based on indicators where the output is scaled against staff numbers etc. Since the size of university is not something that it is neither feasible or even desirable for a university to increase within a short period of time, we believe ARWU falls short on an essential element in evaluation theory, namely that the evaluated institutions should have performance ownership to the performance indicators that are used (Carter et al., 1992).

What is the value (and the limitations) of aggregated scores?

While rankings may be used by universities for publicity and marketing issues, and for students in selecting universities (Harvey, 2008), a recent study of Nordic universities found that the rankings have a relatively modest impact on decision-making and strategic actions in these universities (Elken et al., 2016). We believe the latter finding is a natural consequence of the use of aggregated scores in rankings such as THE and ARWU. As management information it seems impossible to use data that has been mathematically transformed and collapsed into aggregated scores the way THE and ARWU

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

does. Without knowledge of the raw data being used, it is impossible to make use of the data for quality improvement. Initiatives have been made into fill such an information gap, e.g. the Leiden ranking – which is in reality not a ranking – but rather an information source for bibliometric indicators, where it is easy for universities to find relevant units for comparison, and relevant measures to use. The U-Multitrank initiative is also based on making relevant comparisons possible, without going through the construction of a league table, where it is unavoidable that focus is on rank positions – and changes in these. A problem with ARWU and THE is that if a university has a set of institutions they would like to compare themselves with – and find some of the indicators in THE and ARWU useful for benchmarking purposes – they get no information whatsoever in these rankings about how they are actually performing relative to these, as the indicator scores are presented as transformed scores which has no intuitive meaning, or because the indicators have been collapsed with other indicators into aggregate scores, where the results on each of these are unknown. In this study, we have demonstrated a method that can be used to analyze the relative importance of the different indicators in THE and ARWU, and used in-depth data from a selected sample of universities to exemplify the reasons for differences in scores on these indicators.

Acknowledgments

This paper is based on a NIFU-report (Piro et al. 2014) commissioned by the Ministry of Education and Research.

References

Bookstein, F.L., Seidler, H., Fieder, M., & Winckler, G. (2010): "Too much noise in the Times Higher Education rankings". *Scientometrics* 85(1): 295-299.

Bowman, N.A. & Bastedo, M.N. (2011): "Anchoring effects in world university rankings: exploring biases in reputation scores". *Higher Education* 61(4): 431-444.

Carter, N., Klein, R., & Day, P. (1992): *How Organizations Measure Success. The Use of Performance Indicators in Government.* London: Routledge.

Dichev, I. (2001): "New or noise? Estimating the noise in the US News university rankings". *Research in Higher Education* 42(3): 237-266.

Docampo, D. & Cram, L. (2015): "On the Effects of Institutional Size in University Classifications: The Case of the Shanghai Ranking". *Scientometrics* 102(2): 1325-1346.

Elken, M., Hovdhaugen, E., & Stensaker, B. (2016): «Global rankings in the Nordic region: challenging the identity of research-intensive universities?». *Higher Education* (publised online 25 January 2016).

Gnolek, S.L., Falciano, V.T. & Kuncel, R.W. (2014): "Modeling Change and Variation in U.S. News & World Report College Rankings: What would it really take to be in the Top 20?". *Research in Higher Education* 55: 761-779.

Harvey, L. (2008): "Rankings of Higher Education Institutions: A Critical Review", *Quality in Higher Education* 14(3): 187-207.

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5

Kehm, B. & Stensaker, B. (eds.) (2009) *University rankings, diversity and the new landscape of higher education*. Rotterdam: Sense publishers.

Liu, N.C., & Cheng, Y. (2005): "The academic ranking of world universities", *Higher Education in Europe* 30(2): 127-136.

Piro, F.N., Hovdhaugen, E., Elken, M., Sivertsen, G., Benner, M., & Stensaker, B. (2014): *Nordiske universiteter og internasjonale universitetsrangeringer. Hva forklarer nordiske plasseringer og hvordan forholder universitetene seg til rangeringene?* NIFU Report 25/2014. Oslo: NIFU.

Safón, V. (2013): "What do global university rankings really measure? The search for the X factor and the X entity". *Scientometrics* 97: 223-244.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., van Eck, N.J., van Leeuwen, T.N., van Raan, A.F.J., Visser, M.S., & Wouters, P. (2012): *The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation.* CWTS Working Paper Series (CWTS-WP-2012-007). Leiden: Centre for Science and Technology Studies (CWTS).

Post-print version
Publisher version: DOI: 10.1007/s11192-016-2056-5