Milica Maričić[1], Nikola Zornić[2], Ivan Pilčević[3], Aleksandra Dačić-Pilević[4]
[1]University of Belgrade, Faculty of Organizational Sciences, Serbia,
[2] South Stream B.V., The Netherlands, 3 British American Tobacco, The Netherlands

ARWU vs. Alternative ARWU Ranking: What are the Consequences for Lower Ranked Universities?

Abstract: The ARWU ranking has been a source of academic debate since its development in 2003, but the same does not account for the Alternative ARWU ranking. Namely, the Alternative ARWU ranking attempts to reduce the influence of the prestigious indicators Alumni and Award which are based on the number of received Nobel Prizes and Fields Medals by alumni or university staff. However, the consequences of the reduction of the two indicators have not been scrutinized in detail. Therefore, we propose a statistical approach to the comparison of the two rankings and an in-depth analysis of the Alternative ARWU groups. The obtained results, which are based on the official data, can provide new insights into the nature of the Alternative ARWU ranking. The presented approach might initiate further research on the Alternative ARWU ranking and on the impact of university ranking's list length.

Keywords: ARWU ranking, Alternative ARWU ranking, Award factor, University ranking
JEL Classification: C10, C38, I23

1. Introduction

Decision makers and government representatives use numbers in the process of policy making in various spheres of life (Porter, 1995). Higher education is just one of thoe that saw the introduction of scores into its assessment methods through university rankings (Daraio & Bonaccorsi, 2017). Since 2003 and the first global university ranking, the university rankings have proliferated. The three world-acknowledged and most often analysed rankings are the Academic Ranking of World Universities (ARWU), the Times Higher Education (THE), and the Quacquarelli Symonds World University Ranking (QS). However, these rankings led to the creation of dozens separate rankings and subrankings: by region, by subject, by field and so on (Hazelkorn & Gibson, 2016). This information goes to show several facts: that there is a need for university rankings, that their scope is changing, and that university rankings have become a serious industry (Hazelkorn & Gibson, 2016). Therefore, the conclusion can be made that university rankings are here to stay (Nature News, 2007).

Throughout the years, institutions that publish university rankings aimed to answer the needs of students, on one hand, and the needs of universities, on the other hand. One such case is the case of the Alternative ARWU ranking. Namely, the ARWU ranking tends to take into account the prestige of the institution by including indicators such as the number of alumni and staff who received the Nobel Prize or the Fields Medal. However, it is tough for an institution to receive any of them. On the other hand, if a university receives one of the two prestigious awards, its rank skyrockets in the next edition of the ranking. For example, the Toulouse School of Economics received its first Nobel prize in 2014 in Economic sciences and entered the ARWU ranking straight into group 201-300 in 2015. Similarly, the London School of Economics received its first Nobel Prize in 2010 and went from group 201–302 to group 102–150. Also, once a university has received the Nobel Prize or Fields Medal its rank position improves and stays constant for a certain period of time (Piro & Sivertsen, 2016). Many universities worldwide pointed such situations out to the Shanghai Ranking Consultancy and suggested an alternative ranking without the Award Factor (*Alumni* and *Award* indicators) (ARWU, 2014). Their valuable observation led to the creation of the Alternative ARWU ranking.

[1]Corresponding author: Milica Maričić, E-mail: milica.maricic@fon.bg.ac.rs

It is important to create a university ranking as reliable and as accurate as possible (Radojicic & Jeremic, 2012). Therefore, an in-depth analysis of ranking methodologies is needed. One of the major research which aimed to tackle the structure of world acknowledged university rankings was done by Aguillo et al. (2010). They attempted to measure the similarity between ARWU, QS-THE, Webometrics ranking, and the Higher Education and Accreditation Council of the Taiwan ranking (HEEACT). The important contribution of their research is that they analysed universities by groups: top 10, top 100, and top 200. Similar research was recently conducted by Shehatta & Mahmood (2016) who performed a detailed correlation analysis of six global rankings by groups top 50, top 100, and top 200. Both of these papers indicate that, besides the overall analysis of university rankings, analysis by groups is also desirable.

Although the methodology of the Alternative ARWU ranking resembles the official ARWU ranking with smaller alterations, since its publication in 2014, there has not been, to our knowledge, an in-depth analysis of its results and the consequences of the removal of the two indicators related to prestigious awards. Herein we provide analysis of the observed differences alongside with the thorough study on the Alternative ARWU ranking by groups. The following section gives a literature review of the methodological issues the rankings encounter. Both ranking methodologies and the research methodology are presented in the following section, while the research results are given in Section 4. Some of the possible future directions of the study are provided in the next section, while the concluding remarks are given in the final chapter.

2. Methodological Issues of the ARWU and Alternative ARWU Rankings

When it first appeared, the ARWU ranking attracted both a positive feedback and rigorous critique. Since 2003, academics specialized in statistics, bibliometrics, and composite indicators have tried to point out some of the methodological flaws this widely recognized university ranking faces.

One of the first papers which aimed to tackle the ARWU methodology was the work by Florian (2007) who pointed out the problem of result irreproducibility. Namely, he attempted to recreate the 2005 ARWU ranking using the official methodology but without success. He encountered discrepancies in the ranking, as well as in the processed data. Although the values of some of the indicators were hard to reproduce, especially the ones obtained directly from the universities, the same should not account for objective bibliometric indicators.

Several years later Dehon et al. (2010) employed the Principal Component Analysis (PCA) to inspect whether the factors measured by the ARWU should be integrated into a single measure. Their analysis showed that there are two different and uncorrelated aspects of research performance: one related to high level research, and the second related to research output. Their result raised the question on whetehr various aspects of university performance should be aggregated at the cost of information loss. The presented research clearly initiated further debate on university rankings and their methodologies which are often presented as "black boxes" (Longden, 2011).

Shortly after Jeremic et al. (2011) suggested an application of a multivariate approach based on the I-distance to analyse the ARWU weighting scheme. Namely, they created an alternative ARWU ranking by proposing new weights. The result of their research provided insights that the university rankings are sensitive to indicator weight alterations. In the following year, Jovanovic et al. (2012) examined the influence of normalization on the results of the ARWU ranking. Namely, they again employed the I-distance method, this time on raw data, and normalized values of indicators. They showed that the normalization significantly alters the rankings meaning that additional research on the type of normalization should be conducted.

Another publication that should be mentioned is the one by Zornic et al. (2014). The main aspect of their research was to scrutinize the indicator *PUB*. Since the indicator is calculated as the simple sum of published papers without taking into account the Impact Factor (IF) of the journal, two universities with the same number of published papers will have the same indicator value regardless of the quality of the journals they published their paper in. They unveiled the example of two Brazilian universities who had the same value of *PUB* that the differences in the IF of the top five journals are staggering. They ranged from 0.301 (Sao Paulo State University, journal *Semina-Ciencias Agrarias*) to 3.730 (Federal University of Rio de Janeiro, journal *PLOS ONE*). One of the suggested directions to overcome the observed issue of inequality is to include the aspects of journal quality into the *PUB* indicator.

The most recent article which at the same time acts as an inspiration for our research is Dobrota & Dobrota (2016). In their interesting paper they conducted the uncertainty and sensitivity analysis of both ARWU and Alternative ARWU rankings. The results clearly showed that the Alternative ARWU ranking is more stable, meaning that the Award Factor distorts the ranking and makes it more volatile. Their result could act as an impetus for further analysis of the Alternative ARWU ranking and the rationale for its creation.

It can be concluded from the selected papers that the contemporary literature on the ARWU ranking is impressive. The ARWU and the Alternative ARWU rankings are composite indicators and, as such, they face some methodological issues (Nardo et al., 2005). So far some drawbacks of the ARWU ranking have been observed and constructive solutions have been recommended. However, the Alternative ARWU ranking has not yet received the attention it deserves. Therefore, we herein propose a statistical approach to scrutinize and explore what the consequences of the exclusion of the *Alumni* and *Award* indicators on the university ranks are, with a special overview of the lower ranked universities and the impact of the university ranking list length.

3. Methodology

3.1 ARWU and Alternative ARWU ranking methodologies

The Academic Ranking of World Universities (ARWU) or the Shanghai ranking is published yearly by the Institute of Higher Education of the Jiao Tong University in Shanghai since 2003. The ARWU aims to rank the world's top 500 universities and to better present its results the universities are devided into groups of 100 by the achieved result. The ranking itself consists of six indicators which aim to rank institutions according to academic and research performance (Liu & Cheng, 2005). Table 1 presents six indicators, their codes and weights which are assigned to them before aggregation.

Table 1: ARWU indicators, their codes and weights

Indicator	Code	Weight
Alumni of an institution winning Nobel Prizes and Field Medals	Alumni	10%
Staff of an institution winning Nobel Prizes and Fields Medals	Award	20%
Highly cited researchers in 21 broad subject categories	HiCi	20%
Papers published in *Nature* and *Science*	N&S	20%
Papers indexed in SCIe and Social SCI	PUB	20%
Per capita academic performance of an institution	PCP	10%

Source: ARWU, 2016

The first indicator, *Alumni,* is related to the number of alumni of an institution who won the Nobel Prize and/or Fields Medal. Alumni are defined as those who obtained Bachelor's, Master's, or Doctor's degrees from the observed institution (Liu & Cheng, 2005). Similarly, the *Award* indicator is related to the number of staff of an institution who won the Nobel Prize and/or Fields Medal. In the case of both indicators different weights are set, according to the periods of obtaining degrees (*Alumni*) or Prizes (*Award*) (Liu & Cheng, 2005). The following indicator, *HiCi,* aims to measure the number of staff who are classified as highly cited researchers by the ISI Web of Knowledge published by Thomson Reuters Corporation. However, an interesting research by Bornmann and Bauer (2015) showed that the ranking of the highly cited researchers depends on the number of institutions named by the authors. They pointed out that the discrepancies are clearly visible when comparing ranking lists of universities based on the firstly named institution and on all the named institutions. Therefore, in the official methodology, it should be clearly stated which counting approach has been used. The next two indicators are bibliometric indicators aiming to measure the research output. The *N&S,* on one hand*,* indicates the number of papers published in the *Nature* and *Science* in the last five years, while the *PUB* indicates the total number of articles indexed by the Science Citation Index-Expanded (SCIe) and the Social Science Citation Index (Social SCI) in the previous year. Both indicators take into account only articles (Liu & Cheng, 2005). However, two things are not clear when it comes to these two indicators. First, why does the counting window differ for *N&S* and *PUB,* and second, why is the number of published papers used instead of citations count? Finally, the *PCP* attempts to measure the academic performance of an institution by the full-time equivalent academic staff. It is calculated as the weighted score of the five indicators divided by the number of full-time employed academic staff (ARWU, 2015a).

To calculate the ARWU scores the raw indicator values are normalized first. Within each indicator, the best performing university is given a score of 100 and becomes the benchmark against which the scores of all other universities are measured. In the next step, the normalized scores are weighted accordingly and aggregated using sum (Dehon et al., 2010).

The methodology of the Alternative ARWU ranking slightly differs from the official ARWU ranking methodology. Firstly, indicators *Alumni* and *Award* are excluded from the framework. Secondly, the *PCP* indicator is recalculated based on the remaining three indicators (ARWU, 2014). In our paper, we will use the *APCP* as the code for the rescaled *PCP* used in the Alternative ARWU ranking methodology. Finally, the weights assigned to the remaining four indicators are 70%; they have not been rescaled to 100%. Besides these three, there are no more methodological differences between the two rankings.

3.2 Research methodology

The research methodology is based on the descriptive statistics, parametric, non-parametric tests, and multivariate analysis, specifically Principal component analysis (PCA) conducted on the ranking indicators, scores, ranks, and rank differences depending on the analysis.

The descriptive statistics, namely means and measures of variability have been used to analyse the Rank and the Absolute Rank Differences between the ARWU and Alternative ARWU rank. Also, cross-tabulation has been performed to observe whether the universities qualified for a higher or a lower ranked group of universities after the Award factor was removed. Such an approach should provide more information on how the rank of universities changed, both on the overall ranking and per group. More detail on these approaches could be found in Holcomb (1997).

In our research, we aimed to explore both rank and score differences. Parametric tests used to analyse the score differences and indicator values were the Pearson's correlation coefficient, Student's t-test, Analysis of variance (ANOVA) and its Post hoc test, Tamhane's test, and Fisher's Z transformation. These tests should unveil whether there is a statistically significant difference between the ARWU and the Alternative ARWU score. On the other hand, non-parametric tests were employed to analyse the rank differences: Spearman's Rho, Kruskal-Wallis test and its Post hoc test, Dunn's test. Similarly, these tests are to unveil whether there is a statistically significant difference between the ARWU and the Alternative ARWU, but this time when it comes to university ranks. More information on the performed analysis could be found in, for example, Tabachnick and Fidell (2001) and Hair et al. (2009).

To obtain an in-depth analysis of the Alternative ARWU structure, we performed the PCA to see whether the indicators should be aggregated in a single metrics. Namely, the PCA is used to extract the important information from the dataset and to represent it as a set of new orthogonal variables, thus reducing the dimensionality of the observed phenomenon (Abdi & Williams, 2010). A detailed overview of the PCA is also given in Tabachnick and Fidell (2001).

4. Results

The dataset on which the analysis was performed contained all six ARWU indicator values and the *Alternative PCP* (*APCP*) for 500 ranked universities for the year 2014. The normalized data are publicly available on the official ARWU website (ARWU, 2014).

The *Alumni* and *Award* indicators are both related to the number of the prestigious Nobel Prize and Fields Medal winners who are currently employed on or are alumni of a university. However, only a few universities can win any of these prizes per year and many universities have never won any of them. Therefore, a significant percent of the observed universities has been assigned the value of 0 for both of these indicators. Namely, 293 universities (58.6%) have the value of 0 for the indicator *Alumni,* 359 (71.8%) have the same value for the indicator *Award,* whereas there are 264 (52.8%) universities that have values of 0 for both indicators*.* Excluding the Award Factor this is one way the Shanghai Ranking Consultancy attempts to provide a more comprehensive and trustworthy ranking.

As explained above, the Alternative ARWU ranking is created when the Award Factor is removed from the ranking framework. The first conducted analysis was the correlation analysis between the two rankings. Correlation based on rank order, measured through Spearman's Rho and correlation based on scores, measured through Pearson's correlation coefficients on the whole sample were r_s=0.955 (p<0.01) and r=0.959 (p<0.01). The high correlation means that the scores and ranks are highly consistent when observing all universities. However, the same does not account if the scores and ranks are analysed by ARWU groups (Table 2). Taking a closer look at the Pearson's correlation coefficient, it can be observed that its values are decreasing going down the ranking. This means that, although the two indicators have been removed and that the *PCP* has been altered, the scores of the top 100 universities have not changed significantly. The reduction of the Pearson's correlation coefficient in the latter groups shows that the new methodology had an effect, especially in the last group in which the correlation was the least, 0.421 (p<0.01). On the other hand, Spearman's Rho shows a different pattern. Again, the correlation is the highest in the first group, 0.827 (p<0.01), meaning that the ranks within the group are highly stable. Correlation is moderate in the other four groups whereas it is the least in the group 401-500, closely followed by group 101-200, meaning their ranks changed the most.

Table 2: Correlation analysis between ARWU and Alternative ARWU per official ARWU group

Group	Pearson's correlation coefficient	Spearman's Rho
1-100	0.918**	0.827**
101-200	0.621**	0.620**
201-300	0.583**	0.651**
301-400	0.485**	0.698**
401-500	0.421**	0.659**

** The correlation is statistically significant at the level 0.01

Shehatta & Mahmood (2016) state that when analysing and comparing university rankings, it is important to take into account the ranking scores. Following their idea of analysing the university ranking scores, an interesting insight can be provided by Table 3 which summarizes the scores of universities ranked 1, 100, 200, 300, 400, and 500. The difference between the ARWU and Alternative ARWU is the highest for the first place. Namely, the two removed indicators mean that the maximum Alternative ARWU score is 70. Nevertheless, going down the ranks, their values are closer. In the case of the values for the 400th their difference is just 0.1. Although the Award Factor has been removed and the *PCP* has been recalculated, the values of lower ranked universities remained very close.

Table 3: Scores of universities ranked 1, 100, 200, 300, 400, and 500 by ARWU and Alternative ARWU

	1	100	200	300	400	500
ARWU	97.53	23.38	16.46	13.07	10.79	7.74
Alternative ARWU	68	21.42	15.86	13.11	10.89	2.88

Table 4 presents five universities whose rank improved the most and five universities whose rank deteriorated the most by the Alternative ARWU ranking. In the case of universities that improved their rank the most, we can see that the difference between the ARWU and the Alternative ARWU values is up to 2 points. Although the two indicators were excluded, these universities improved their score. On the other hand, the universities whose rank significantly declined had the difference between the ARWU and the Alternative ARWU values up to 6.57 points. For example, let's take a closer look at the George Mason University. Namely, it received two Nobel Prizes for Economy; in 1986 and in 2002 (Nobel Prize, 2016). On the other hand, its results of *HiCi* and *N&S* are not that impressive, 5.3 and 9.3 respectably. Therefore, it was quite affected by the exclusion of the Award Factor. A similar situation was with the University of Buenos Aires (4 Nobel Prizes), University of Lorraine (3 Fields Medals), Ecole Normale Superieure – Lyon (Notable Alumni winners of Fields medals), and the London School of Economics and Political Science (16 Nobel Prizes and Notable Alumni).

Table 4: Five universities whose rank improved the most and five universities whose rank declined the most by the Alternative ARWU ranking

University	ARWU	ARWU Rank	Alternative ARWU	Alternative ARWU rank	Score difference	Rank Difference
Highest rank improvement						
Scuola Normale Superiore - Pisa	12.35	335	14.12	255	+1.77	+80
Catholic University of Korea	9.84	471	10.89	400	+1.05	+71
Capital University of Medical Sciences	9.53	487	10.4	438	+0.87	+49
University of Pompeu Fabra	12.82	316	13.77	268	+0.95	+48
University of Wageningen	20.11	148	21.08	102	+0.97	+46
Highest rank decline						
George Mason University	16.65	195	10.08	464	-6.57	-269
University of Buenos Aires	17.47	183	11.13	388	-6.34	-205
University of Lorraine	14.58	254	10.23	450	-4.35	-196
Ecole Normale Superieure - Lyon	14.1	264	10.13	460	-3.97	-196
London School of Economics and Political Science	17.24	150	12.46	331	-4.78	-181

Although it is obvious that there is difference between the ARWU and the Alternative ARWU rankings, the Wilcoxon Signed Ranks test was performed. It showed that the observed difference is statistically significant (Z=-4.184, p<0.01). Namely, 327 universities improved their rank, six remained at the same position, while the remaining 167 universities worsened their position. A descriptive analysis of the absolute rank difference showed that the mean absolute difference is 30.78 ranks, with a standard deviation of 30.48, median 24, and ranges through 269 ranks. The absolute difference is positively skewed (β_1=3.392) and leptokurtic (β_2=16.018).

Another interesting insight is to graphically present and analyse the ARWU rank and the Absolute Rank Difference (Figure 1). As we can see, the first 50 ranked universities are highly stable, without large rank changes. However, the same does not account for the universities up to 250th position whose differences range up to 270 position. Although the rank difference is higher than in the top 50 group, the last 200 show more stability when it comes to an absolute rank change.
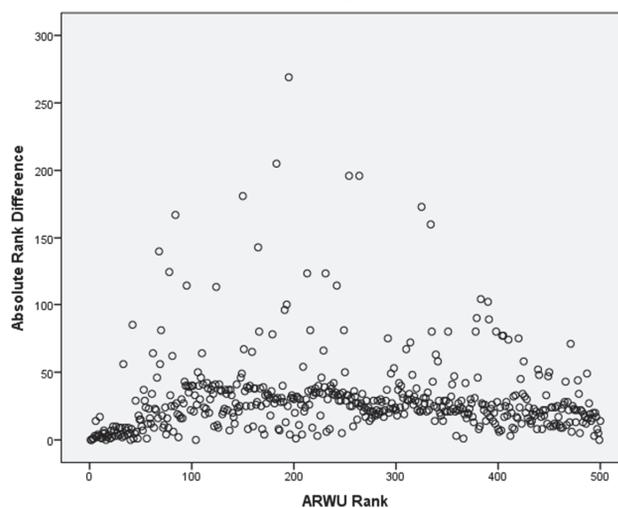
Figure 1: Scatter plot of the ARWU rank and the Absolute Rank Difference

Additionally, the Absolute Rank Difference is observed in each group (Table 5). The mean difference is, as expected, the largest in the 101-200 (39.37) group and the smallest is in the first group (21.45). The least stable group is the 101-200 group whose standard deviation is 39.051. The difference of the following 250 universities proves to be more stable, whereas the ranks in the last group are the most stable according to the standard deviation of 16.541. the median also provides interesting insights. According to the median, the second group is the most volatile. For example, the first and the second groups are wide range groups, but their medians show that the volatility of the first group was not as high as the volatility of the latter group although they both have high standard deviations.

Table 5: Descriptive statistics of the Absolute Rank Difference per each ARWU group

Group	Mean	Median	Std	Min	Max
1-100	21.45	10	29.634	0	167
101-200	39.37	31	39.051	0	269
201-300	35.20	29	31.209	1	196
301-400	34.52	26.5	27.950	1	173
401-500	23.40	20	16.541	0	77

What would also be of value to explore is whether there is a statistically significant difference in the Absolute Rank Differences between the five official ARWU groups. Kruskal-Wallis test confirmed that there are differences ($x^2=75.656$, df=4, p<0.01). As differences were observed, Dunn's test, as a post hoc test was performed with a Bonferroni correction to find out which differences were statistically significant. The test showed that differences between groups 101-200 and 201-300, 101-200 and 301-400, and 201-300 and 301-400 are not statistically significant (p>0.05), which means that those universities oscillated in a similar way. The same accounts for groups 1-100 and 401-500 (p>0.05). This result means that the first and the fifth group oscillated differently than the other three groups, and by looking at the median, we can conclude that they oscillated less.

Finally, the Rank Difference is observed in each group (Table 6). Taking a closer look at the mean Rank difference, we can conclude that the universities from the group 401-500 improved their ranks the most. The group 201-300 has the highest median, meaning 50% of universities improved its position for more than 21.5 places. Standard deviation pointed group 101-200 as the most volatile. The minimum and maximum showed that a university in the group 301-400 improved its place for the most, 80 places, and also that a university in the group 401-500 worsened its rank the least, for 77 places.

Table 6: Descriptive statistics of the Rank Difference per each ARWU group

Group	Mean	Median	Std	Min	Max
1-100	-9.79	-1.0	35.300	-168	43
101-200	-4.59	20.5	55.401	-269	46
201-300	0.10	21.5	47.175	-197	46
301-400	3.62	21.0	44.398	-173	80
401-500	10.66	17.0	26.681	-77	71

After excluding the Award Factor, we were able to calculate the Alternative ARWU ranks. Based on the obtained ranks we created the Alternative ARWU groups. By cross-tabulating the ARWU and Alternative AWRU groups valuable insights can be provided (Table 7). The most sensitive ARWU group of universities is the group from 101st to 200th positon. Namely, 17% of the group improved their ranking groups, while 14% declined, whereas one of them, George Mason University went from 195th to 464th place. Compared to other groups, universities from the group 401-500 improved their ranks the least, 16% of them advanced into the better-ranked group. However, that result should not be undermined. Namely, up to 21% of universities per group advanced into a better-ranked group. This result enforces the fact that the removal of Award Factor and the *Alternative PCP* had a positive impact on the rank of lower ranked universities. Interestingly, no university managed to improve its position for more than one group.

Table 7: Cross-tabulation of the ARWU and Alternative AWRU groups

		Alternative ARWU groups					
		1-100	101-200	201-300	301-400	401-500	Total
ARWU groups	1-100	83	13	4	0	0	100
	101-200	17	69	10	3	1	100
	201-300	0	18	65	15	2	100
	301-400	0	0	21	66	13	100
	401-500	0	0	0	16	84	100
	Total	100	100	100	100	100	500

The universities that dropped from top 100 to the 201-300 group are the Ecole Normale Superieure – Paris (68th – 208th), the Technion-Israel Institute of Technology (78th – 202nd), the Moscow State University (84th – 251st), and the University of Strasbourg (95th – 209th). For example, the Moscow State University had high values of *Award* and *Alumni*, 42.4, 33.0 respectively, but an unexpectedly low *HiCi*, 0. Three universities that significantly worsened their position and entered the 401-500 group are George Mason University (195th – 464th), University of Lorraine (254th – 450th), and the Ecole Normale Superieure - Lyon (264th – 460th).

The next step in our analysis was to explore whether there are statistically significant differences between the indicator values within groups formed by the ARWU and the Alternative ARWU. We aimed to see how and whether the indicator values within the same group rank changed. The suggested analysis could be performed for indicators *HiCi*, *N&S*, and *PUB* as these indicators are used in both methodologies. Namely, 15 Student's t-tests were performed, three per ranking group, and all of them showed that there was no statistically significant difference between the group values ($p < 0.01$ for all conducted tests). This is an important result as it indicates that the values of these indicators do not differ between the two grouping systems. Therefore, attention should be placed on indicators *PCP* and *APCP* per groups, as they are the ones which make the difference. Table 8 gives basic descriptive statistics of the *PCP* and the *APCP* per groups. The first thing that attracts attention is that the mean values of the *APCP* are higher than the *PCP*. Also, the standard deviation of the *APCP* is higher, especially when comparing groups 201-300.

Table 8: Descriptive statistics of indicators PCP and APCP per groups

PCP					APCP				
ARWU group	Mean	Std	min	max	Alternative ARWU group	Mean	Std	Min	Max
1-100	31.662	12.605	17.5	100	1-100	35.953	11.649	20.0	100
101-200	21.909	3.934	12.2	32.2	101-200	27.073	4.822	15.9	41.8
201-300	19.000	3.180	1.5	28.8	201-300	25.276	7.222	16.7	76.6
301-400	18.428	5.813	12	58.9	301-400	22.482	4.839	15.0	45.3
401-500	16.347	4.310	10.3	37.9	401-500	19.306	4.644	5.0	37.8

To additionally explore the alternative groups and their indicator values, we conducted a one-way Analysis of variance (ANOVA) which showed that there is a statistically significant difference between the groups of each of the indicators ($F_{HiCi} = 232.335$, $F_{N\&S} = 199.476$, $F_{PUB} = 218.020$, $F_{APCP} = 77.073$, $p < 0.01$). This result means that the indicator values within groups differentiates. However, another question arises: are there any two groups where the differences are not statistically significant? To answer the question, we performed a Post Hoc analysis, the Tamhane's test. It showed that there is a difference between all groups for all indicators except for the indicator *APCP*. The Post Hoc test revealed that there is no statistically significant difference between groups 101-200 and 201-300 ($p > 0.05$).

Another aspect of the Alternative ARWU ranking should be inspected. Namely, should the four indicators be aggregated into one single measure and are they in a positive correlation by each group? The correlation analysis (Table 9) between the indicators on the entire sample reveals that the correlations between the four

indicators are positive, medium to strong, and statistically significant. The highest correlation is between indicators *N&S* and *HiCi,* 0.869, while the lowest correlation is between *APCP* and *PUB,* 0.482. However, when the correlation analysis is performed by the Alternative ARWU groups, the results attract attention. To compare two correlations coefficients and examine whether the observed difference was statistically significant, we used the test proposed by Cohen and Cohen (1983) which is based on the Fisher's Z-transformation. In the first group, the correlation between the *PUB* and the *HiCi* significantly declined, from 0.649 to 0.444 (Z=2.67, p<0.01) and the correlation between the *APCP* and the *PUB* declined from 0.482 to a statistically insignificant correlation of 0.151 (Z=3.364, p<0.01). Still, all correlations are positive. On the other hand, some of the correlations in the remaining four groups are negative and statistically insignificant. The negative coefficient of correlation means that the two observed measures are not of the same direction, meaning that the increase of one measure will decrease the values of the other. Therefore, before aggregation, all measures should be transformed so that their direction is the same (Nardo et al., 2005). Accordingly, the question arises whether the results of the Alternative ARWU ranking below 100th position are valid.

Table 9: Correlation analysis of the four Alternative ARWU indicators per Alternative ARWU group

	All 500					Group 1-100			
	HiCi	N&S	PUB	APCP		HiCi	N&S	PUB	APCP
HiCi	1				HiCi	1			
N&S	0.869**	1			N&S	0.874**	1		
PUB	0.649**	0.659**	1		PUB	0.444**	0.445**	1	
APCP	0.640**	0.682**	0.482**	1	APCP	0.575**	0.647**	0.151	1
	Group 101-200					Group 201-300			
	HiCi	N&S	PUB	APCP		HiCi	N&S	PUB	APCP
HiCi	1				HiCi	1			
N&S	0.115	1			N&S	-0.03	1		
PUB	-0.431**	-0.426**	1		PUB	-0.527**	-0.368**	1	
APCP	0.056	0.171	-0.107	1	APCP	-0.124	0.017	-0.412**	1
	Group 301-400					Group 401-500			
	HiCi	N&S	PUB	APCP		HiCi	N&S	PUB	APCP
HiCi	1				HiCi	1			
N&S	0.109	1			N&S	-0.104	1		
PUB	-0.674**	-.507**	1		PUB	-0.340**	-0.148	1	
APCP	-0.082	-0.081	-0.238**	1	APCP	-0.180	-0.160*	0.150	1

** Correlation is significant at the level 0.01; * Correlation is significant at the level 0.05

One of the ways to determine whether indicators should be aggregated in a single dimension is to perform the Principal Component Analysis (PCA). The PCA was performed on each five groups and the entire sample. The Kaiser-Meyer-Olkin (KMO) measure varied from 0.801 (entire sample) to 0.257 (group 201-300). The Bartlett's test of sphericity is statistically significant in all cases (p<0.01). The obtained results of the PCA with Varimax rotation are presented in Table 10. Analysing the whole sample, all four indicators should be aggregated in one dimension. Namely, the PCA retained one component which describes 75.212% of variance. The same accounts for group 1-100, whereas its retained component explains 65.962% of variance. However, in the next four groups, the PCA suggested retaining two components. The structure of components for groups 101-200 and 401-500 is the same, while the components of the other two groups differ. The conducted PCA per group indicates that although on the overall sample it is justified to aggregate the indicators in one dimension, the same does not account on the group level.

Table 10: Number of retained components, the % of variance they explain, and indicators which make each component on the entire sample and by Alternative ARWU groups

All			Group 1-100		
Component	% of variance	Indicators	Component	% of variance	Indicators
1	75.212	HiCi, N&S, PUB, APCP	1	65.962	HiCi, N&S, PUB, APCP
Group 101-200			Group 201-300		
Component	% of variance	Indicators	Component	% of variance	Indicators
1	38.781	HiCi, PUB	1	36.184	N&S, PUB
2	29.239	N&S, APCP	2	34.988	HiCi, APCP
Group 301-400			Group 401-500		
Component	% of variance	Indicators	Component	% of variance	Indicators
1	47.413	HiCi, N&S, PUB	1	36.484	HiCi, PUB
2	27.306	APCP	2	31.112	N&S, APCP

4.1 Results of the Universities of the Balkan Peninsula

Out of ten countries of the Balkan Peninsula, three countries had their representatives in the AWRU 2014 ranking: Slovenia, Serbia, and Greece. It would be of interest to explore how their ranks changed in the Alternative ARWU Ranking (Table 11). The only university that advanced into a highly ranked group is the National and Kapodistrian University of Athens; it is in the group 201-300 by the Alternative ARWU ranking. The rest of the universities improved their ranks but remained in the same group. It can be concluded that the Alternative ARWU ranking had a positive effect on the lower ranked universities. However, those universities mostly did not advance in the better-ranked group.

Table 11: Results of ARWU and Alternative ARWU ranks of the universities of the Balkan Peninsula

University	Country	ARWU	ARWU Rank	Alternative ARWU Rank	Difference
National and Kapodistrian University of Athens	Greece	12.76	321	299	22
University of Belgrade	Serbia	11.51	369	353	16
University of Ljubljana	Slovenia	9.68	482	461	21
Aristotle University of Thessaloniki	Greece	9.78	476	463	13

Additionally, to better represent the position of the universities of the Balkan Peninsula we examined the new ARWU 2016 ranking list and compared it to the last year's ARWU 2015. The data are publicly available on the official ARWU website (AWRU, 2016; 2015b). Comparing the results, another Balkan university entered the list in 2016, the University of Zagreb. The University of Zagreb went in and out of the list in the last several years. Its results for 2016 show that in the observed group it has a very good *HiCi* and *N&S*, 10.3 and 5.3, respectively. Table 12 gives the comparison of the 2016 and 2015 results of the four Balkan universities that remained on the list.

Table 12: 2015 and 2016 ARWU groups of the selected universities, the improvement of their indicator values and the score difference

University	2015 group	2016 group	HiCi*	N&S*	PUB*	PCP*	Overall difference
University of Belgrade	301-400	201-300	10.3	1.9	0.4	1.3	2.65
National and Kapodistrian University of Athens	301-400	301-400	3.9	1	-1.5	0.3	0.71
University of Ljubljana	401-500	401-500	0	0.4	0.9	1.1	0.37
Aristotle University of Thessaloniki	401-500	401-500	5.4	0.6	-0.4	0.6	1.18

* The values present the improvement from 2015

The University of Belgrade improved its position the most and entered the 201-300 group. The main reason for such a sharp increase is the *HiCi* indicator. Namely, it improved by 10.3 points. The main reason for such a good result is that mathematics professor Stojan Radenovic has become a highly cited researcher according to Thomson Reuters (Thomson Reuters, 2016). What should also be noted is that the *N&S* indicator increased by 1.9 points, meaning that the number of papers published in the two respectable journals increased. The overall difference in the ARWU score for the University of Belgrade is 2.65. The two Greek universities improved their *HiCi*, but both saw their *PUB* decline. They remained in their respective groups, the same as the University of Ljubljana who minimally improved its score.

5. Further Directions of the Study

One of the possible future directions of study is to integrate more advanced bibliometric indicators such as percentile-based indicators which are based on citation count (Bornmann & Mutz, 2014; Zornic et al., 2015). Recent bibliometric research labelled percentiles as a new method suitable for the normalization of citation counts of publications in terms of subject category (Bornmann, 2013). If percentiles are to be used, the citation window should be at least five years. Currently, the observed bibliometric window by the ARWU is just one year. Adding a more complex bibliometric indicator such as a percentile-based indicator would differentiate universities more and would also take into account the scientific contribution of the published paper measured through citations. Also, differences in citing behaviour between sciences could be taken into account according to Bornmann, de Moya Anegón, and Mutz (2013) who showed that certain subject-specific types of institutions are in an advantageous position when it comes to ranking regarding the citation impact.

The correlation analysis has clearly showed that on the overall level and in the first group the correlations between the four indicators are positive and statistically significant. However, the same does not account for the remaining groups. According to the Joint Research Centre guide for composite indicator development (Saisana, 2012), if "there are negative correlations among indicators, it means that either the desired direction of the indicator is wrong or that there are trade-offs between indicators". More attention should be paid to the observed negative correlation and its impacts on the overall metrics.

As Jovanovic et al. (2012) pointed out, the normalization has a detrimental effect on the rankings. However, they did not examine the effects of various normalization methods. The currently employed normalization method, the percentage of the highest scoring institution, might not be the best solution. Namely, taking a closer look at the normalized data, it can be observed that there is a big difference in indicator values between the best and the second best universities (Table 13). The largest difference is for the indicator *N&S*, 26.9 points, whereas the smallest difference is for the indicator *Awards*, 3.4. Therefore, another type of normalization could be utilized or the dataset should be checked for the presence of outliers.

Table 13: The values of the first and the second best universities of each ARWU and Alternative ARWU indicator

Indicator	Rank	University	Value
Alumni	1	Harvard University	100
	2	University of Cambridge	77.1
Awards	1	Harvard University	100
	2	University of Cambridge	96.6
HiCi	1	Harvard University	100
	2	Stanford University	80.1
N&S	1	Harvard University	100
	2	Massachusetts Institute of Technology (MIT)	73.1
PUB	1	Harvard University	100
	2	University of Toronto	79.1
PCP	1	California Institute of Technology	100
	2	Harvard University	76.6
APCP	1	California Institute of Technology	100
	2	Harvard University	80.0

Discussion and Conclusion

Although statisticians, experts in higher education assessment, and bibliometricians raised their concern regarding the university rankings (for example, Saisana et al., 2011; Marginson, 2014), various stakeholders still believe university rankings are vital for the development of higher education (Kauppi, 2016). Therefore, university ranking creators should place even more effort in creating reliable composite measurements of universities' performance.

One university ranking that has attracted the attention of politicians, students, academics, and universities since its development is the *Academic Ranking of World Universities*, published yearly by the Institute of Higher Education of the Jiao Tong University in Shanghai. Putting aside a vast number of critiques this ranking received (for example Zornic et al., 2014), ARWU ranking should be accounted for being the pioneer in the creation of world university rankings.

The ARWU ranking itself is quite specific. At the same time, it measures the quality of education (*Alumni*), quality of faculty (*Awards*), research output (*HiCi, N&S,* and *PUB*), and per capita performance (*PCP*) (Liu & Cheng, 2005). Therefore, it aims to integrate teaching, research and efficiency of the institution. However, the two indicators related to teaching are highly prestigious and difficult to achieve. Therefore, to create a ranking with a less rigorous methodology without the prestigious indicators the Alternative ARWU ranking was published for the year 2014.

Herein we attempt to provide an in-depth analysis of the Alternative ARWU ranking on two aspects: first, on the consequences of the removal of Award Factor on the scores and ranks of universities, and secondly, on the impact of the list length.

The observed correlation of scores and ranks between the overall ARWU and Alternative ARWU is very high and positive. Nonetheless, the correlation analysis by official ARWU groups showed that the Pearson's correlation coefficient declines to 0.421 in the group 401-500 and that the Spearman's Rho declines to 0.620 in the group 101-200. This means that there are differences which are not visible when the results are analysed on the overall level. To confirm what the correlation analysis showed, the Wilcoxon test was conducted which showed that there was a statistically significant difference between the two ranks. Namely, 65.4% of universities improved their positions in comparison with the ARWU and Alternative ARWU rankings.

The analysis of the Absolute Rank Difference points out the 101-200 group as the most volatile, and the 401-500 group as the most stable. The analysis of the Rank Difference showed that the mean rank difference of the first two groups is negative. In the first group, 36% of universities deteriorated their positions, while in the second group that percent goes to 39%. It can be concluded that the new methodology had an adverse impact on universities ranked from 1-200, while it had a positive impact on the universities ranked from 301-500. As rank differences were observed, it was of interest to find out whether there are differences in the absolute rank differences between the five groups. The Kruskal-Wallis test has confirmed that there is a difference among the groups; the Dunn's post hoc test has showed that groups 1-100 and 401-500 have similar absolute rank differences which differed from the three remaining groups.

The comparison of the ARWU with the Alternative ARWU group showed that the most volatile was the latter group as some of its universities were ranked in the group 401-500. The universities that significantly worsened their position are the universities that recently received or have Nobel Prizes or Fields Medals, but have low research output. After the creation of the Alternative ARWU groups, their results were compared with the ARWU group results for the three same indicators. No statistically significant difference was observed meaning that although the two indicators were removed and the *PCP* was altered, the mean group values do not differ. Therefore, the *PCP* and *APCP* were compared, and the *APCP* shows more volatility per group than the *PCP*.

To explore whether it is reasonable to aggregate the four indicators in one overall score, the correlation analysis of the four Alternative ARWU indicators has been performed alongside with the PCA. Both analyses have justified that on the overall level and on the level of group 1-100, indicators should be aggregated. However, the same does not account for the later four groups: the correlation coefficients are negative and not significant, and the PCA suggested to retain two components.

Although 65.4% of universities improved their rank, the group which most improved its ranks is the group 401-500, 71% of universities. However, this amazing result is not visible. The conclusion can be made that although the Award Factor was removed and although it had a positive impact on the lower ranked universities, their achievement remained the same. This can be seen in the example of universities from the Balkan Peninsula. On the other hand, the universities in the prospective group 101-200 suffered the most after the exclusion of the *Alumni* and the *Award.* It seems that the new reduced methodology did not have the desired impact as just several universities significantly benefited.

The presented paper delivers a thorough analysis of the AWRU and the Alternative ARWU rankings for the year 2014. We believe our research may trigger further academic research on the topic of the Alternative ARWU ranking and the impact of the list length upon the university ranking methodologies.

Acknowledgements

REFERENCES

[1] Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433-459. doi: 10.1002/wics.101

[2] Aguillo, I. F., Bar-Ilan, J., Levene, M., & Ortega, J. L. (2010). Comparing university rankings. *Scientometrics*, *85*(1), 243-256. doi: 10.1007/s11192-010-0190-z

[3] ARWU. (2014). Alternative Ranking 2014 (Excluding Award Factor). Retrieved on July 25, 2016 from http://www.shanghairanking.com/Alternative_Ranking_Excluding_Award_Factor/Excluding_Award_Factor2014.html

[4] ARWU. (2015a). Methodology. Retrieved on July 20, 2016 from http://www.shanghairanking.com/ARWU-Methodology-2015.html

[5] ARWU. (2015b). Academic Ranking of World Universities 2015. Retrieved on August 16, 2016 from http://www.shanghairanking.com/ARWU2015.html

[6] ARWU. (2016). Academic Ranking of World Universities 2016. Retrieved on August 16, 2016 from http://www.shanghairanking.com/ARWU2016.html

[7] Bornmann, L, de Moya Anegón, F., & Mutz, R. (2013). Do universities or research institutions with a specific subject profile have an advantage or a disadvantage in institutional rankings? *Journal of the Association for Information Science and Technology, 64*(11), 2310-2316. doi: 10.1002/asi.22923

[8] Bornmann, L. (2013). How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top cited papers. *Journal of the American Society for Information Science and Technology, 64*(3), 587-595. doi: 10.1002/asi.22792

[9] Bornmann, L., & Bauer, J. (2015). Which of the world's institutions employ the most highly cited researchers? An analysis of the data from highlycited. com. *Journal of the Association for Information Science and Technology, 66*(10), 2146-2148. doi: 10.1002/asi.23396

[10] Bornmann, L., & Mutz, R. (2014). From P100 to P100': A new citation-rank approach. *Journal of the Association for Information Science and Technology, 65*(9), 1939-1943. doi: 10.1002/asi.23152

[11] Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

[12] Daraio, C., & Bonaccorsi, A. (2017). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*. 68, 508-529. doi: 10.1002/asi.23679

[13] Dehon, C., McCathie, A., & Verardi, V. (2010). Uncovering excellence in academic rankings: A closer look at the Shanghai ranking. *Scientometrics*, *83*(2), 515-524. doi: 10.1007/s11192-009-0076-0

[14] Dobrota, M., & Dobrota, M. (2016). ARWU ranking uncertainty and sensitivity: What if the award factor was Excluded?. *Journal of the Association for Information Science and Technology*, *67*(2), 480-482. doi: 10.1002/asi.23527

[15] Florian, R. (2007). Irreproducibility of the results of the Shanghai academic ranking of world universities. *Scientometrics, 72*(1), 25-32. doi: 10.1007/s11192-007-1712-1

[16] Hair, J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate data analysis*. Prentice Hall. ISBN: 0138132631

[17] Hazelkorn, E., & Gibson, A. (2016). Another Year, Another Methodology: Are Rankings Telling Us Anything New?. *International Higher Education*, (84), 3-4.

[18] Holcomb, Z. (1997). *Fundamentals of Descriptive Statistics*. Routledge. ISBN: 1884585051

[19] Jeremic, V., Bulajic, M., Martic, M., & Radojicic, Z. (2011). A fresh approach to evaluating the academic ranking of world universities. *Scientometrics, 87*(3), 587-596. doi: 10.1007/s11192-011-0361-6

[20] Jovanovic, M., Jeremic, V., Savic, G., Bulajic, M., & Martic, M. (2012). How does the normalization of data affect the ARWU ranking?. *Scientometrics, 93*(2), 319-327. doi: 10.1007/s11192-012-0674-0

[21] Kauppi, N. (2016). Ranking and structuration of a transnational field of higher education. In eds. Normand, R. and Derouet, J.L., " A European Politics of Education: Perspectives from Sociology, Policy Studies and Politics.". Routledge. (pp. 92-103)

[22] Liu, N. C., & Cheng, Y. (2005). The academic ranking of world universities. *Higher education in Europe*, *30*(2), 127-136. doi: 10.1080/03797720500260116

[23] Longden, B. (2011). Ranking indicators and weights. In J. C., Shin, R., Toutkoushian & U. Teichler (Ed.), "University Rankings" (pp. 73-104). Springer Netherlands. doi: 10.1007/978-94-007-1116-7_5

[24] Marginson, S. (2014). University rankings and social science. *European Journal of Education*, *49*(1), 45-59. doi: 10.1111/ejed.12061

[25] Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on constructing composite indicators. OECD Publishing

[26] Nature News. (2007). Academics strike back at spurious rankings. *Nature, 447* (May), 514–515. doi: 10.1038/447514b

[27] Nobel Prize. (2016). All Prizes in Economic Sciences. Retrieved August 1, 2016 from https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/

[28] Piro, F. N., & Sivertsen, G. (2016). How can differences in international university rankings be explained?. *Scientometrics*, 209. 2263-2278. doi: 10.1007/s11192-016-2056-5

[29] Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press. ISBN: 9780691029085

[30] Radojicic, Z., & Jeremic, V. (2012). Quantity or quality: What matters more in ranking higher education institutions. *Current science*, *103*(2), 158-162.

[31] Saisana, M. (2012). A do-it-yourself guide in Excel for composite indicator development, European Commission, Joint Research Centre, Italy, http://composite-indicators.jrc.ec.europa.eu

[32] Saisana, M., d'Hombres, B., & Saltelli, A. (2011). Rickety numbers: Volatility of university rankings and policy implications. *Research policy*, *40*(1), 165-177.

[33] Shehatta, I., & Mahmood, K. (2016) Correlation among top 100 universities in the major six global rankings: policy implications. *Scientometrics*, 109 (2), 1231-1254. doi: 10.1007/s11192-016-2065-4

[34] Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate analysis*. California State University Northridge: Harper Collins College Publishers. ISBN: 0321056779

[35] Thomson Reuters. (2016). Highly cited researchers. Retrieved on August 17, 2016 from http://hcr.state-ofinnovation.thomsonreuters.com/

[36] Zornic, N., Bornmann, L., Maricic, M., Markovic, A., Martic, M., & Jeremic, V. (2015). Ranking institutions within a university based on their scientific performance: a percentile-based approach. *El profesional de la información*, *24*(5). doi: 10.3145/epi.2015.sep.05

[37] Zornic, N., Markovic, A., & Jeremic, V. (2014). How the top 500 ARWU can provide a misleading rank. *Journal of the Association for Information Science and Technology*, *65*(6), 1303-1304. doi: 10.1002/asi.23207

/// **About the Author**

Milica Maričić
milica.maricic@fon.bg.ac.rs

Milica Maričić is a teaching associate at the Department of Operations Research and Statistics at the Faculty of Organizational Sciences, University of Belgrade (UB). After graduation in 2014 at the Faculty of Organizational Sciences, she got her MSc at the same faculty, where she specialized in business analytics and statistics. Currently, she is on her PhD studies at the Faculty of Organizational Sciences, where she is studying Quantitative management with special interest in computational and applied statistics.

Nikola Zornić
nikola.zornic@fon.bg.ac.rs

Nikola Zornić is a teaching associate at the Department of Management at the Faculty of Organizational Sciences, University of Belgrade. After graduation in 2013 at the Faculty of Organizational Sciences, he got his MSc at the same Faculty, where he specialized in business intelligence and decision making. Currently, he is on his PhD studies at the Faculty of Organizational Sciences, where he is studying Quantitative management with special interest in simulation models.

Ivan Pilčević
ivan.pilcevic@gmail.com

Ivan Pilčević works in South Stream B.V. company for gas pipeline design and construction. Ivan is IT professional with more than 10 years of experience in managing ERP (SAP), CRM, WMS implementation, TM&D systems integration and IT services. His current professional interests are related to IT services and business processes and design and IT systems implementation.

Aleksandra Dačić-Pilčević
aleksandra.dacic@gmail.com

Aleksandra Dačić-Pilčević is senior IT Manager with 11 years of Information Technology management experience in British American Tobacco - international world's second-biggest tobacco company. Her current professional interests are: electronic business, IT systems implementation, IT project management.