Is it possible to rank universities using fewer indicators?
A study on five international university rankings

Güleda Doğan and Umut Al

Department of Information Management, Hacettepe University, Ankara, Turkey

Abstract

Purpose - This study aims to analyse the similarity of intra-indicators used in research-focused international university rankings (ARWU, NTU, URAP, QS, and RUR) over years, and show the effect of similar indicators on overall rankings for 2015. The research questions addressed in this study in accordance with these purposes are as follows: 1) At what level are the intra-indicators used in international university rankings similar? 2) Is it possible to group intra-indicators according to their similarities? 3) What is the effect of similar intra-indicators on overall rankings?

Design/methodology/approach - Indicator-based scores of all universities in five research-focused international university rankings for all years they ranked, forms the dataset of this study for the first and second research questions. We used multidimensional scaling and cosine similarity measure to analyse similarity of indicators and to answer these two research questions. Indicator-based scores and overall ranking scores for 2015 are used as data and Spearman correlation test is applied to answer the third research question.

Findings - Results of the analyses show that the intra-indicators used in ARWU, NTU, and URAP are highly similar and that they can be grouped according to their similarities. We also examined the effect of similar indicators on 2015 overall ranking lists for these three rankings. Using one of the indicators in the similar indicators group each time, we created new overall rankings and correlated these with the existing overall rankings for 2015. It was found that the new overall rankings created were highly correlated with the existing ones. NTU and URAP are affected least from the omitted similar indicators, which means it is possible for these two rankings to create very similar overall ranking lists to the existing overall ranking using fewer indicators.

Research limitations/implications - This study covers the five research-focused international university rankings (ARWU, NTU, URAP, QS, and RUR) that create overall rankings annually and presents indicator-based scores and overall ranking scores. CWTS, Mapping Scientific Excellence, Nature Index, and SIR (until 2015) are not included in the scope of this article, since they do not create overall ranking lists. Likewise, THE, CWUR, and US are not included because of not presenting indicator-based scores. Some difficulties were meet while obtaining the desired dataset from QS. Ranking lists (so the indicator-based scores) of 2010 and 2011 were not accessible for QS. Moreover, although QS ranks more than 700 universities, it reveals the scores of some indicators for only first 400 or 500 universities in the 2012-2015 rankings. Therefore, only first 400 universities in 2012-2015 rankings were analyzed for QS. Although, QS's (as far as possible) and RUR's data analysed in this study, it was statistically not possible to reach any conclusion from the results of multidimensional scaling conducted for QS and RUR because of the S-Stress values calculated over 0.200 that refers to poor concordance.

Practical implications - The results of this study may be considered mainly by ranking bodies, policy- and decision-makers. The ranking bodies may use the results to review the indicators they use, to decide on which indicators to use in their rankings, and to question if it is necessary to continue overall rankings. Policy- and decision-makers may also benefit from the results of this study by thinking of giving up using overall ranking results as an important input in their decisions and policies.

Originality/value - This study is the first to use multidimensional scaling and cosine similarity measure for revealing the similarity of indicators and similar indicator groups. Ranking data is skewed that requires conducting nonparametric statistical analysis, that is the reason for using multidimensional scaling (as a nonparametric multivariate statistical method). The study covers all ranking years and all universities in the ranking lists, and is different from the similar studies in the literature that analyse data for shorter time intervals and top-ranked universities in the ranking lists. It can be said that the similarity of intra-indicators for URAP, NTU, and RUR is analysed for the first in this study, based on our literature review.
Keywords International university rankings, Ranking indicators, Similarity of intra-indicators, Research-focused indicators, Multidimensional scaling, Redundant indicators

Paper type Research paper

Introduction

The early history of ranking universities has started in the US and going back to 1870s. Ranking universities became widespread in 1980s with the popular press in US, particularly US News & World Reports' *Best Colleges* (Stuart, 1995). Although numerous national rankings were developed following *Best Colleges* (see http://ireg-observatory.org/en/ranking-profile), the first international ranking of universities was done by Shanghai Jiao Tong University in 2003, under the name *Academic Ranking of World Universities* (ARWU). While the main aim of ARWU was to determine the position of Chinese universities among other world universities and to identify their strengths and weaknesses (ARWU, 2015a), ARWU attracted intense interest along with the discussions on the subject (Liu *et al.*, 2005; Van Raan, 2005a; Van Raan, 2005b). However, these discussions did not prevent the rapid increase in the number of rankings. *Ranking Web of Universities* (http://webometrics.info/) was developed by Cybermetrics Lab, an initiative of the Spanish National Research Council, in 2004, immediately after ARWU's first international university ranking list in 2003. Another ranking that emerged in 2004 out of a partnership between Times Higher Education (THE) and Quacquarelli Symonds Ltd. (QS) was *Times Higher Education Supplement* (THES), followed by *THE-QS World University Rankings*, which then continued as two different rankings by 2010: *QS World University Rankings* and *THE World University Rankings* (Holmes, 2010, p. 91). The increase in the number of international university rankings continued with *uniRank*, formerly named *4 International Colleges & Universities* (4icu.org), which ranks 12,358 colleges and universities in 200 countries according to their web popularity (uniRank, 2017). Another international university ranking is the *NTU Ranking: Performance Ranking of Scientific Papers for World Universities* initiated by the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT) in 2007 and transferred to Taiwan National University in 2012 (NTU, 2017a). *CWTS Leiden Ranking* (http://www.leidenranking.com/) was introduced by the Leiden University Center for Science and Technology Studies (CWTS) in 2008 with a different approach to rankings. *CWTS Leiden Ranking* provides indicator-based rankings, but not an overall ranking weighting the indicators. *SCImago Institutions Rankings* (SIR) has been ranking institutions in different sectors including universities by research performance, innovation and web visibility since 2009 (SIR, 2017a; SIR, 2017b). In 2010, three international rankings emerged: *University Ranking by Academic Performance* (URAP), *Round University Ranking* (RUR), and *Universitas Indonesia* (UI) *GreenMetric University Ranking*. *UI GreenMetric University Ranking* focuses on green campuses and sustainability (UI GreenMetric, 2015). RUR ranks universities according to 20 different indicators under the categories of research, education, international diversity, and financial sustainability (RUR, 2017). URAP (http://www.urapcenter.org), developed by the Informatics Institute of the Middle East Technical University in Turkey, ranks universities by academic performance and lists 2000 universities. Starting in 2012, *Youth Incorporated Global University Rankings* has used indicators focusing on the opportunities that universities provide to students (Youth Incorporated, 2015). The *Nature Index*, which emerged in 2013, ranks by the number of articles calculated in three different ways: the number of articles, the fractional number of articles, and the weighted fractional number of articles. They update their ranking lists on a monthly basis (A Guide to the Nature Index, 2017; Nature Index, 2017). *Mapping Scientific Excellence* (http://www.excellencemapping.net) emerged in 2013 and has a very different approach than the abovementioned rankings. It is a web application that lists universities and research-focused institutions in specific areas according to scientific performance with visual representation. *Mapping Scientific Excellence* does not create ranking lists on an annual basis. They create a new version by updating current rankings according to five-year periods (2005–2009, 2006–2010, 2007–2011, 2008–2012) (Bornmann *et al.*, 2014, p. 28; Whitcroft, 2013). U-Multirank, which emerged in 2014 and is funded by the European Union, has introduced a multidimensional and user-focused approach to rank universities internationally by categorizing them into one of five different groups from "very good" to "weak" according to selected indicators (Butler, 2010; U-Multirank, 2017a; U-Multirank, 2017b; Van Vught and Ziegele, 2012). US News & World Report has begun to rank universities internationally in 2014 (*Best Global Universities Rankings*) (Morse *et al.*, 2016). Reuters, which ranked first in 2015 (*Reuters Top 100: The World's Most Innovative Universities*), lists universities by innovation performance and totally bases its rankings on surveys, unlike other rankings (Ewalt, 2016; Reuters, 2017).

International university rankings have been an important topic for higher education studies in recent years (Lo, 2014). However, we have seen that the ranking bodies indicate the difficulty in measuring quality and scientific performance and the impossibility of creating excellent and totally objective rankings. Therefore, it is important to examine the methodology of the rankings carefully and understand what they measure (Baty, 2017; Billaut *et al.*, 2010; Degener, 2014; Liu and Cheng, 2005; Van Noorden, 2010). The main problem with existing university rankings is the attribution of results without regarding the methodological details or what they measure (Loughran, 2016). Policy- and decision-makers' non-expert bibliometric use and "fast and dirty" bibliometric analyses are the most significant problems for university rankings (Van Raan, 2005a). Top-ranked universities are perceived as "best universities" by decision-making senior managers (King, 2009); ranking results are used by universities as marketing tools via web pages, student orientations, international meetings, and various media organs. It is known that some countries have changed their funding policies (e.g., China, see Li, 2012) with a focus on creating "world universities" (King, 2009). For example, individual and institutional performance measures in Taiwan's higher education system are constructed according to the indicators of rankings (Lo, 2014, p. 99). These indicators used in rankings are counted as one of the crucial factors for the position of universities (Doğan, 2017, p. 213; Shin and Toutkoushian, 2011, p. 10). In a study on ARWU (Markovic *et al.*, 2014), it was understood that many universities were going to increase their scores on the number of publications (indicator) which they would change more easily than other indicators to obtain a place on the ARWU list, one of the most popular international rankings. It is stated in the same study that there are universities (e.g., University of Belgrade) included in the ranking lists which use this method and those university administrations have increased the number of publications required for professorship for this purpose. These examples show the dilemmas of "universities that are good at using different ranking systems and applying different indicators" and "good universities" (Shehatta and Mahmood, 2016, p. 1232).

There are two groups of studies on international university rankings that one of these groups focus on the indicators and methodologies of ranking by examining the existing indicators (Bookstein *et al.*, 2010; Florian, 2006; Holmes, 2006; Liu and Cheng, 2005; Soh, 2011; Van Raan, 2005a; Van Raan, 2005c; Vincke, 2009) and criticizing the methodology of rankings and suggest methodological improvement (Billaut *et al.*, 2010; Daraio *et al.*, 2015; Dobrota *et al.*, 2016; Jajo and Harrison, 2014; Soh, 2013a, Soh, 2013b, Soh, 2017, Tofallis, 2012), while the other group of studies compare ranking results in terms of the position of universities (Aguillo *et al.*, 2010; Bornmann and Mutz, 2015; Bornmann *et al.*, 2013; Chen and Liao, 2012; Claassen, 2015; Jöns and Hoyler, 2013; Marginson, 2007; Moed, 2017; Piro and Sivertsen, 2016; Shehatta and Mahmood, 2016; Williams, 2008).

Most of the international university rankings are research-focused and they use a wide range of bibliometric indicators to measure research performance of universities (Rauhvargers, 2014, p. 40; Shehatta and Mahmood, 2016, p. 1232). In this study, the aim is to analyse the similarities of intra-indicators used in overall rankings of research-focused international university rankings over years. Besides the similarity of intra-indicators, we aim to show the effect of similarity on overall rankings for 2015. Having highly correlated indicators that measure the same thing in a metric is statistically wrong and undesirable, that is known as collinearity/multicollinearity problem in statistics (Enders, 2013). This indicates that the measurement is not accurate and valid. Likewise, using similar indicators in ranking systems may cause some problems, which means ranking can be easily manipulated. For example, by focusing on/targeting a particular indicator, scores on other indicators that are highly related to that indicator can also be increased.

CWTS, Mapping Scientific Excellence, Nature Index, and SIR (until 2015) are out of the scope of this study, since they do not create overall ranking lists. Ranking Web of Universities, Reuters, uniRank, Green Metrics, U-Multirank, and Youth Incorporated are also not in the scope of the study because of not being research-focused. Although being research-focused and creating overall rankings, THE, CWUR, and US are not included in the study, since they do not present indicator-based scores. Thus, our study considers ARWU, QS, URAP, NTU, and RUR that are both research-focused and presents indicator-based scores and overall ranking scores annually.

In the light of the above, the research questions addressed in this study are as below:

1. At what level are the intra-indicators used in international university rankings similar?
2. Is it possible to group intra-indicators according to their similarities?
3. What is the effect of similar intra-indicators on overall rankings?

There are similar studies in the literature and one of these studies (Shehatta and Mahmood, 2016) examined the inter-correlation of ARWU's indicators based on scores and their correlation with the overall ranking scores. In another study (Dehon *et al.*, 2010), robust principal components analysis was performed for the first 150 universities in ARWU. The number of publications and the Nobel Prize winners were the two factors found to have a very strong effect. Another study (Docampo and Cram, 2015) attempted to determine the factors causing the overall change in ARWU, using principal components analysis. The analysis was based on ARWU's five indicators, and the number of full-time faculty was considered. The study revealed that two factors explain 85% of the variance in overall score. In a similar study (Soh, 2011), inter-correlations of indicators and contribution of indicators to the overall score were investigated. Correlation tests and multiple regression analysis were conducted for the top 100 universities in ARWU's 2010 ranking, top 200 in QS's 2008 ranking, and top 200 in THE's 2010 ranking. Results revealed that indicators of rankings are inter-correlated and there are redundant indicators for each of the three rankings. It was found that two indicators of ARWU (number of articles published in Nature and Science and number of international awarded alumni) and two categories of THE (teaching and citations) explain 94% of the variance in overall score, whereas three indicators of QS (academic peer review, number of students per faculty, and number of citations per faculty) explain 91% of the variance in overall score, which means both the sufficiency of two–three indicators for these rankings and the redundancy of other indicators used in these three rankings (Soh, 2011). Van Raan's (2005a) research, found very low correlations in THE-QS 2004 between bibliometric indicators and indicators based on expert opinions (R^2=0.5%) and pointed out that indicators based on expert opinions are not reliable. ARWU is criticized with that prize-winning alumni and faculty are thought to represent the quality of education; however, these prizes cannot accurately measure a current situation (Billaut *et al.*, 2010). ARWU presented an alternative ranking list in 2014 that excludes two disputable award indicators: internationally awarded alumni and internationally awarded faculty numbers (see ARWU, 2014). A study compared the 2014 ARWU ranking with the alternative one excludes these two award indicators based on an uncertainty and stability analysis (Dobrota and Dobrota, 2016). Less uncertainty and higher stability were found in the alternative ranking. Abramo states that the presence of highly cited researchers has a significant effect on university performance and claims that a change in methodology will not easily or genuinely affect the result (Abramo *et al.*, 2013).

The abovementioned studies in the literature that focus on the indicators of international university rankings study on particular rankings (ARWU, THE, or QS), mostly with one-year ranking data that are limited to top-ranked universities (e.g., top 100, top 150, and top 200), take ranking scores as data, and use methods that require normality assumption for the data (e.g., principal components analysis and Pearson correlation test). ARWU is the first international ranking and, THE and QS are the popular commercial ones but international university rankings are not limited to these three rankings. Considering this situation and to be able to obtain more comparable results, this study focuses on research-focused international university rankings. Most of the similar studies mentioned above consider the top-ranked universities in those lists and one-year ranking data but it is not possible this way to get accurate and valid results on the similarity or correlation of indicators. In this context, this study aim to analyse the entire ranking lists instead of top-ranked universities and considers all ranking years, insofar as possible.

Almost all studies mentioned above use parametric statistical analysis (such as principal component analysis and Pearson correlation test) that require normally distributed score data. The ranking scores have a skewed distribution because of their structure and require conducting nonparametric analysis. Therefore, this study used multidimensional scaling (MDS) that is a nonparametric multivariate statistical method and is first to use MDS analysis (that doesn't require normality assumption) on university ranking indicators.

The Indicators Used in Five International University Rankings

International university rankings that offer an annual overall ranking list select a number of indicators, weight these indicators, and get a score summing up these weighted indicators ($Y = W_1 \text{ st}(X_1) + W_2 \text{ st}(X_2) + \ldots + W_n \text{ st}(X_n)$) (Tofallis, 2012; Usher and Savino, 2006, p. 3). We can see that the indicators used in a significant portion of international university rankings are grouped by certain categories and these groupings differ according to ranking systems. NTU categorizes indicators as research productivity, research impact, and research excellence (NTU, 2017b). ARWU has categories for educational quality, quality of faculty, research output, and per capita performance (ARWU, 2017). RUR groups 20 indicators under four titles: teaching, research, international diversity, and financial sustainability (RUR, 2018). URAP and QS, conversely, do not categorize the indicators (QS, 2017). From Table 1, we can observe that the most intensively used indicators are the ones based on the number of publications and citations.

Table 1. The indicators used in the rankings within the scope of this study

Indicators	RUR	QS	URAP	NTU	ARWU
Number of total publications			•		
Number of articles in the last 11 years				•	
Number of articles in the current year			•	•	•
Number of articles published in Nature and Science					•
Number of publications per faculty	•				
Number of articles normalized for subject			•		
Number of citations in the last 2 years				•	
Number of citations in the last 3 years			•		
Number of citations in the last 11 years				•	
Number of citations normalized for subject	•		•		
Number of citations per faculty	•	•			
Number of citations per publication				•	
Number of publications in the top 1% by citation				•	
Number of articles in the top 5% journals by impact factor				•	
h-index				•	
Number of highly cited researchers					•
Performance per faculty					•
International co-authored publication rate	•				
International co-authored publication number			•		
International student rate	•	•			
International faculty rate	•	•			
Average score of internationality indicators	•				
Number of students per faculty	•	•			
Number of undergraduate degrees per faculty	•				
Doctorate-to-bachelor ratio	•				
Number of doctorates awarded per faculty	•				
Doctoral degrees per admitted PhD	•				
Number of internationally awarded alumni					•
Number of internationally awarded faculty					•
Teaching reputation survey	•	•			
Academic reputation survey	•	•			
International teaching reputation	•				
Institutional income per faculty	•				
Institutional income per student	•				
Research income per faculty	•				
Research income per institutional income	•				
Papers per research income	•				

Rankings, except QS, use indicators based on number of publications, but their approaches differ. RUR uses number of publications per faculty, considers "article" and "note" as publications and divide publication number to the number of academic staff and researchers the year before (RUR, 2017). URAP has three indicators based on the number of publications, and for NTU and ARWU, this number is two. Number of articles published in the current year is a common indicator for URAP, NTU, and ARWU, but data sources differ. Although ARWU and NTU use *SCIE* and *SSCI* as data sources, URAP uses *InCites*. Conversely, ARWU counts only articles as publication type, where URAP considers reviews and notes in publications types in addition to articles. Moreover, ARWU multiplies the number of *SSCI* articles by two. Number of articles in the last 11 years is another indicator for NTU based on publication number, and NTU uses *Essential Science Indicators (ESI)* for this indicator. URAP's other indicator that is based on publication number is the number of total publications, considering conference papers, reviews, letters, discussions, and scripts in addition to journal articles indexed in *WoS* during the previous five years. ARWU and RUR differ from other rankings in terms of their use of an indicator for the number of articles published in "highly influential journals". These journals defined as *Nature* and *Science* for ARWU and *Nature, Science,* and *PNAS (Proceedings of the National Academy of Sciences)* for RUR. URAP normalizes the number of articles for 23 different subjects with another indicator, called "number of articles normalized for subject" (URAP, 2015a; URAP, 2015b).

It is seen in Table 1 that citation-based indicators appear to be widely used in rankings. URAP and NTU use the total number of citations as an indicator. Although URAP considers the last three years for citations, NTU considers the last two years and the last 11 years for two different indicators based on the number of citations (NTU, 2015; URAP, 2015c). NTU has two other indicators related to citations, which are the number of citations per publication (independent from university size), and the number of publications in the top 1% by citations (NTU, 2015). URAP, conversely, uses a second indicator based on citations called number of citations normalized for subject, which is also used as an indicator in RUR. QS also normalizes the number of citations in terms of faculty, which is similar to RUR but different than other rankings (QS, 2015; RUR, 2017). ARWU uses the number of highly cited researchers as an indicator based on Thomson Reuters' highly cited researchers' lists. ARWU created a new indicator in 2004 weighting five other indicators and dividing those by the number of faculty. ARWU categorized this new indicator under a new title, called performance per faculty (ARWU, 2015b).

NTU has two other indicators related to both number of publications and number of citations: h-index and the number of publications in the top 5% of journals by impact factor. NTU calculates the h-index for the last two years using *SCIE* and *SCI* data (NTU, 2015). *Journal Citation Reports* (*JCR*) data is used to determine the 5% of journals for the indicator mentioned above as the number of publications in the top 5% of journals by impact factor (NTU, 2015).

Indicators related to international collaboration in academic studies have only been used in RUR and URAP. While RUR uses international, co-authored publication rates, URAP measures international collaboration using international, co-authored publication numbers (RUR, 2017; URAP, 2015d).

Indicators measuring academic performance are widely used in international university rankings, and there are also indicators related to educational environment, reputation indicators based on surveys, financial indicators related to income information, and indicators related to innovativeness and web performance. RUR and QS differ from the other three rankings with regard to the reputation surveys (academic reputation and teaching reputation). They use these survey results as indicators to determine university reputation. Besides, these two rankings also use indicators related to the educational environment (QS, 2015; RUR, 2017). Being related to the educational environment, international student rate, international faculty rate, and number of students per faculty are the common indicators for RUR and QS. RUR also includes the ratio of doctoral students to undergraduate students and the number of under graduates and doctoral graduates per faculty, doctoral degrees per admitted PhD as indicators (RUR, 2017). ARWU claims to measure education quality using the number of international awarded alumni as an indicator. Likewise, ARWU claims to measure the quality of faculty using the number of internationally awarded faculty as an indicator (ARWU, 2015b). Only RUR uses indicators for income, which are institutional income per faculty and per student, research income per faculty and per institutional income, and papers per research income.

Data and Method

Data

The indicator-based scores of universities in five international university rankings (ARWU, NTU, URAP, QS, and RUR) from the year of their introduction up until 2015 and overall ranking scores for 2015 form the dataset for this study.[1] The scores are used as data because they are more reliable than ranks. Because, 0.3-0.4 changes in score can easily cause about 200-300 changes in rank at the end of the list, while it is not possible to change only 1 rank with these scores for the top of the ranking lists.

Considered the most basic and reliable source for accessing the ranking lists, websites for the rankings were primarily used. Ranking lists were obtained through their websites, except for QS. The QS ranking lists for 2010 and 2011 were not accessible from its own website (http://www.topuniversities.com/). We tried to obtain the missing rankings for these two years through http://www.university-list.net, which complies the university ranking lists, and *Internet Archive: WayBack Machine* (https://web.archive.org/web/), which provides access to the previous versions of this website (http://www.university-list.net), but indicator-based rankings were not accessible this way either. After a certain rank, some universities in the QS 2012-2015 ranking lists do not have indicator-based scores. These universities were excluded because of the impossibility of performing a valid analysis. QS ranks about 700–800 universities in 2012-2015 ranking lists, but academic and teaching reputation scores are only available for the first 400, and scores for the other indicators are available for the top 500 universities. Therefore, analysis results for QS reflect the results for the first 400 universities in the 2012-2015 ranking lists.

ARWU (http://www.shanghairanking.com/) and NTU (http://nturanking.lis.ntu.edu.tw/) have ranked 500 universities and URAP has ranked 2000 universities every year. All ARWU, NTU and URAP ranking lists are accessible through their websites (http://www.urapcenter.org/). All data of RUR was also accessible from its website (http://roundranking.com/) that ranked 567 universities in 2010, 564 in 2011, 635 in 2012, 672 in 2013, 687 in 2014, and 750 in 2015. After all, indicator-based scores and overall ranking scores of universities in URAP's lists for 2010–2015, in ARWU's lists for 2003–2015, in NTU's lists for 2007–2015, in QS's lists for 2012–2015, and in RUR's lists for 2010–2015 are used as data for this study.

Method

The five international university rankings in the scope of this study that provide indicator-based scores and overall ranking scores were examined. MDS, one of the multivariate statistical methods, was performed to determine the grouping of indicators via biplots. Cosine similarity coefficients were calculated to show the similarity of the indicators that are close on biplots. Spearman correlation test is used to show the effect of similar indicators on overall rankings.

MDS is a statistical method used to analyse a group of variables according to their similarities. It provides a graphical representation via biplots, which shows the positions of variables in two- or three-dimensional space in terms of their similarities (distances). It is possible to reveal any kind of distances through MDS, that is popular in marketing and psychological research. MDS is used in marketing research to determine the similarities of the same product with different brands, whereas it is used widely for people's perceptions of traits in psychological studies. MDS is preferred in this study, as the ranking data have a skewed distribution because of their structure, and MDS is a multivariate statistical method that does not require normal distribution of data (Multidimensional Scaling, 2017). By this way, it was also possible to show the inter-similarity of ranking visually. The results of MDS are interpreted according to the stress value obtained from the analysis. Stress value is a measure of the

[1] As mentioned earlier, THE, CWUR and US data is not included within the scope of this study, since they do not present indicator-based scores.

degree of how concordant the distances of the points on the biplot are with their real distances and is calculated according to the Shepard diagram, which shows the concordance between scattering of distances in MDS space and scattering of real distances. Stress values range from 0 to 1, and a stress value of 0 indicates excellent concordance. Although it is not a highly recommended approach to determine intervals for stress values, a stress value over 0.20 indicates poor concordance. The R^2 value found by subtracting the square of the stress value from 1 provides information about how much of the variance/variability of distances determined by the number of dimensions used in graphical representation is explained (Borg and Groenen, 2005, p. 3, 42, 47; Borg *et al.*, 2013, pp. 1–6, 21–26; Cox and Cox, 2001). The analyses were conducted with IBM SPSS 21.0 ALSCAL and S-Stress values are calculated as stress values. Two dimensions were used because of the ease of interpretation, and explained variances were evaluated according to the R^2 values presented with the S-Stress values. Cosine similarity, which measures the angle between two vectors, is often used to solve various information access and machine learning problems (Kryszkiewicz, 2014, p. 2498; Li and Han, 2013, p. 611). By using cosine similarity coefficients in this study, distances of indicators on biplots that represents the similarity of intra-indicators could be interpreted more accurately.

After determining the similar indicators using MDS, we analysed the effect of similar indicators for 2015. We created new overall rankings using only one of the indicators from the similar indicators group each time. The weightings of the omitted indicators were distributed to the other indicators proportionally with their existing weightings and new overall scores calculated via the indicators' scores and new weightings of indicators. We correlated new overall ranking scores with the existing overall ranking scores in terms of ranking scores.

Findings

In this section, the similarity of intra-indicators for ARWU, NTU, URAP, QS, and RUR were examined, using the results of MDS analysis and cosine similarity matrices. After determining the similar intra-indicator groups, the effect of similar indicators on overall rankings were analysed and reported.

S-Stress values for NTU rankings are between 0.036-0.124, and R^2 values are between 0.973 and 0.998 for 2007-2015. The results of MDS of six years in Figure 1 show that the indicators form four different groups according to their positions (similarities). Two indicators on the number of articles (number of articles in the last 11 years and number of articles in last 2 years) are similar and form a group. Four indicators that are relevant with citations (number of citations in the last 2 and last 11 years, number of articles in the top 5% journals, and number of publications in the top 1% by citation) form the second group. According to cosine similarity matrices created for each year between 2007–2015, similarity of two indicators in the first group on the number of articles are calculated to be between 0.984 and 0.999. The cosine similarity coefficients for the four indicators relevant to citations in second group changes between 0.953 and 0.999 for 2007–2015. Although the other two indicators, h-index and number of citations per publications, are separate from each other and from the other two groups of indicators on biplot (Figure 1), considerable cosine similarity coefficients are calculated between h-index and the indicators in the first and second groups for the years 2007–2015 (0.788–0.997): between 0.920 and 0.995 for number of articles in the last 11 years; between 0.849 and 0.997 for number of citations in the last 11 years; between 0.923 and 0.994 for number of articles in the last 2 years; between 0.874 and 0.998 for number of citations in the last 2 years; between 0.858 and 0.997 for number of articles in the top 5% journals; and between 0.788 and 0.997 for number of publications in the top 1% by citation. According to the cosine similarity measure, similarity measures between number of citations per publication and the four indicators in the second group relevant to citations for 2007, 2009, and 2011 were the lowest values (0.704–0.792). All similarity measures for 2013–2015 fell between 0.972 and 0.999.

New overall scores were calculated considering only one of the indicators from each of the similar indicator groups (by giving weights of the subtracted indicators to the other indicators). Eight new overall ranking formed for eight possible cases and Spearman correlation coefficients were found by comparing the overall scores for eight new ranking lists with the overall scores of current 2015 ranking. We obtained very high correlations (0.976–0.995) that indicate overlap of new ranking lists with the existing one between 95%-99%. It is understood from this result that a very similar ranking to that of NTU with eight indicators can be generated with four indicators.

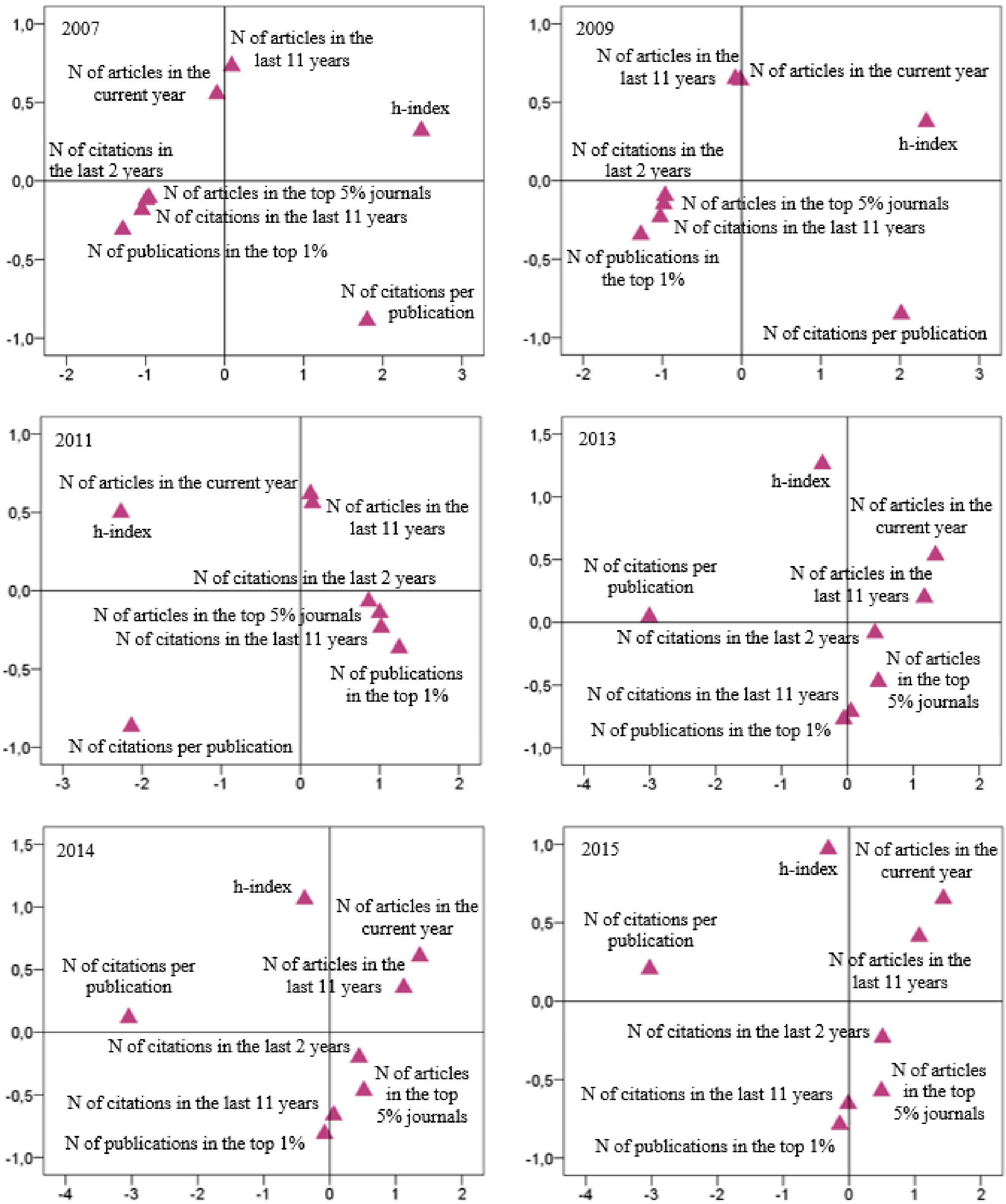

Figure 1. Results of MDS analysis for indicators of NTU ranking (2007, 2009, 2011, 2013–2015)

Figure 2 shows the results of MDS according to the indicators of six years selected from 13 years (2003–2015) of ranking data from ARWU. S-Stress values change between 0.085 and 0.109, and R^2 values change between 0.966 and 0.978.

When the graphical representations of MDS results in Figure 2 are examined, the closest indicators with respect to the first dimension are number of articles in Nature and Science and number of highly cited researchers. The cosine similarity values calculated for these two indicators are between 0.930 and 0.946, except for 2003, when it was 0.909. Other indicators that are most similar to each other in terms of cosine similarity coefficients are performance per faculty and number of articles published in Nature and Science (0.893–0.930), and performance per faculty and number of articles in the current year (0.930–0.950, except for 2004 when it was 0.897). Figure 2 shows that the indicators performance per faculty and number of articles published in Nature and Science are close according to the first dimension and that the indicators performance per faculty and number of articles in the current year are close according to the second dimension.

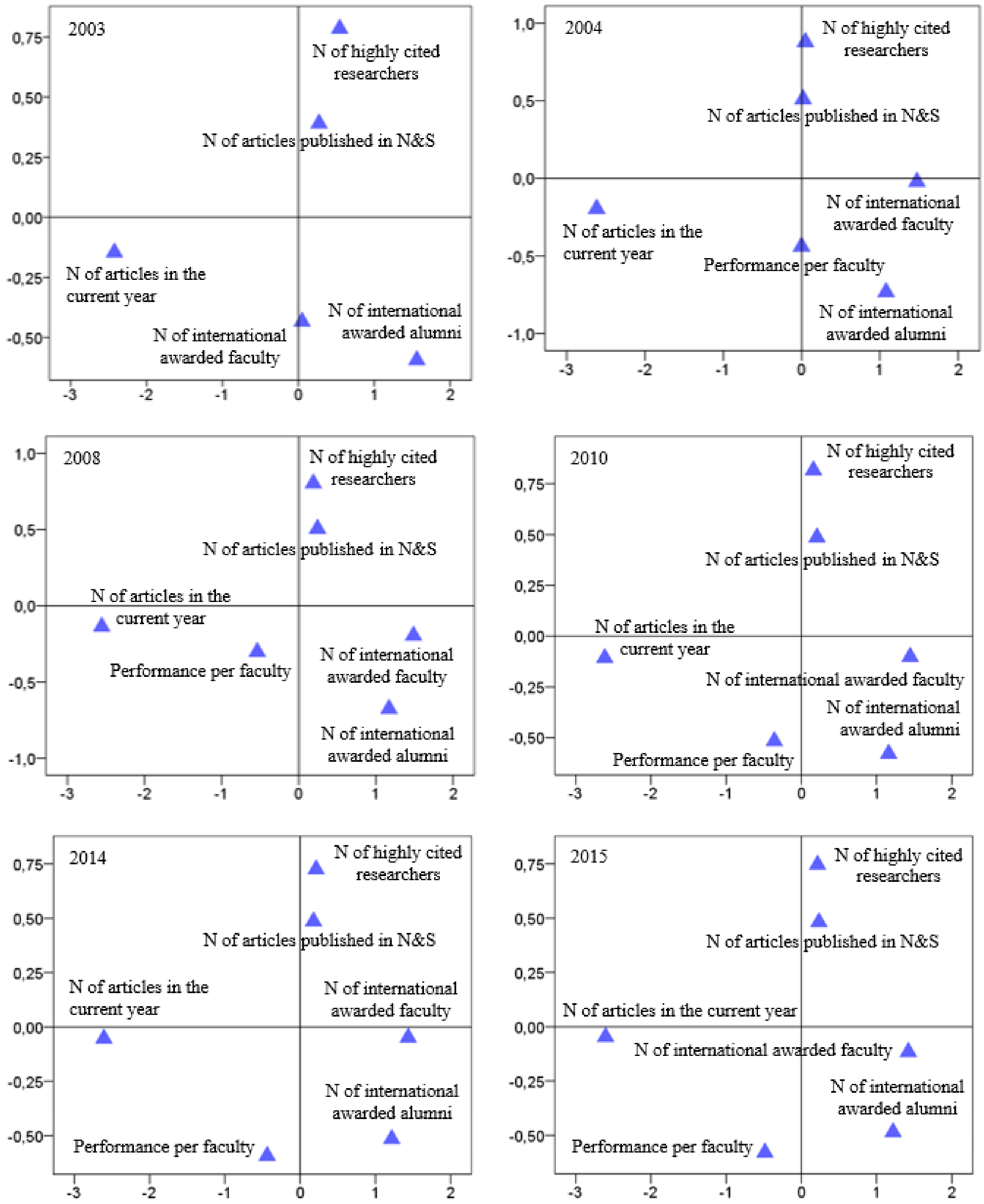

Figure 2. Results of MDS for indicators of ARWU ranking (2003, 2004, 2008, 2010, 2014, and 2015)

The first indicator group that has constituents close to each other according to both dimensions and therefore can be grouped together consist of number of articles published in Nature and Science and number of highly cited researchers. The performance per faculty indicator added to the ARWU ranking in 2004 is close to this group but has moved away year by year. Number of internationally awarded faculty and number of internationally awarded alumni (both indicators) can be thought of as the second group in that they are close in Figure 2. The cosine similarity coefficients for these two indicators are between 0.801 and 0.816 for 2004–2015 and 0.701 for 2003. It is also apparent from Figure 2 that number of articles in the current year (indicator) is located far from all the other indicators according to the first dimension and maintains this location year by year. The lowest similarity coefficients for almost each year of ARWU is calculated between this indicator and the two award indicators, which change between 0.528 and 0.544 for number of internationally awarded faculty and between 0.622 and 0.656 for number of internationally awarded alumni (except 2003 when it was 0.903).

By taking only one of the similar indicators into account, four new overall rankings were generated for 2015, and each was compared with the current ranking list for 2015 in terms of overall scores. They were founded as highly correlated with the existing ranking for 2015 (0.929–0.956). The overlap of new lists with the existing list changes between 86% and 91% which means that a ranking very similar to the ARWU's 2015 overall ranking can be generated with four indicators instead of six indicators.

Figure 3 shows the results of MDS for 2010–2015 rankings of URAP, which draws attention to frequent changes in indicators. The range of S-Stress values found with MDS for the 6-year period indicates a good fit (0.044–0.057). The R^2 values for the same period (0.983–0.994) show that almost all of the variance can be explained in two dimensions.

It is possible to examine the 2010–2015 URAP rankings in three parts (2010, 2011–2013, and 2014–2015) in terms of the indicators used in each year. According to the MDS for 2010, there seems to be no clear grouping of indicators. It was found that the most similar indicators group for 2010 includes h-index, impact, and number of citations in the last 3 years (0.986–0.996). In 2011–2013, number of citations in the last 3 years, journal citation efficiency total, and total journal efficiency multiplier indicators are close to each other. The cosine similarity coefficients calculated for these indicators are the highest similarity coefficients found for these years (2011: 0.994–0.998, 2012: 0.995–0.998, 2013: 0.994–0.997). Similar to 2011–2013, the indicator group including number of citations in the last 3 years, journal citation efficiency total, and total journal efficiency multipliers draw attention with their close positions to each other in 2014–2015. Cosine similarity coefficients were obtained for these three indicators between 0.991 and 0.997 for 2014 and between 0.913 and 0.963 for 2015. For 2015, the similarity coefficient between the number of article in the current year and the number of total publications was also found to be high (0.971). It is worthwhile to note in the evaluation of the results that the similarity coefficients for all the indicators in 2010–2014 are between 0.935 and 0.997, whereas they are between 0.768 and 0.971 for 2015.

Considering only one of the three similar indicators in URAP (in other words, considering four indicators instead of six), a ranking list that has an overlap about 99% with the existing ranking can be obtained. In order to test this, three new ranking lists, each including one of these three similar indicators, have been created, and the correlation of three new rankings with the URAP's 2015 overall ranking list in terms of overall scores showed a high degree of correlation (0.988–0.996).

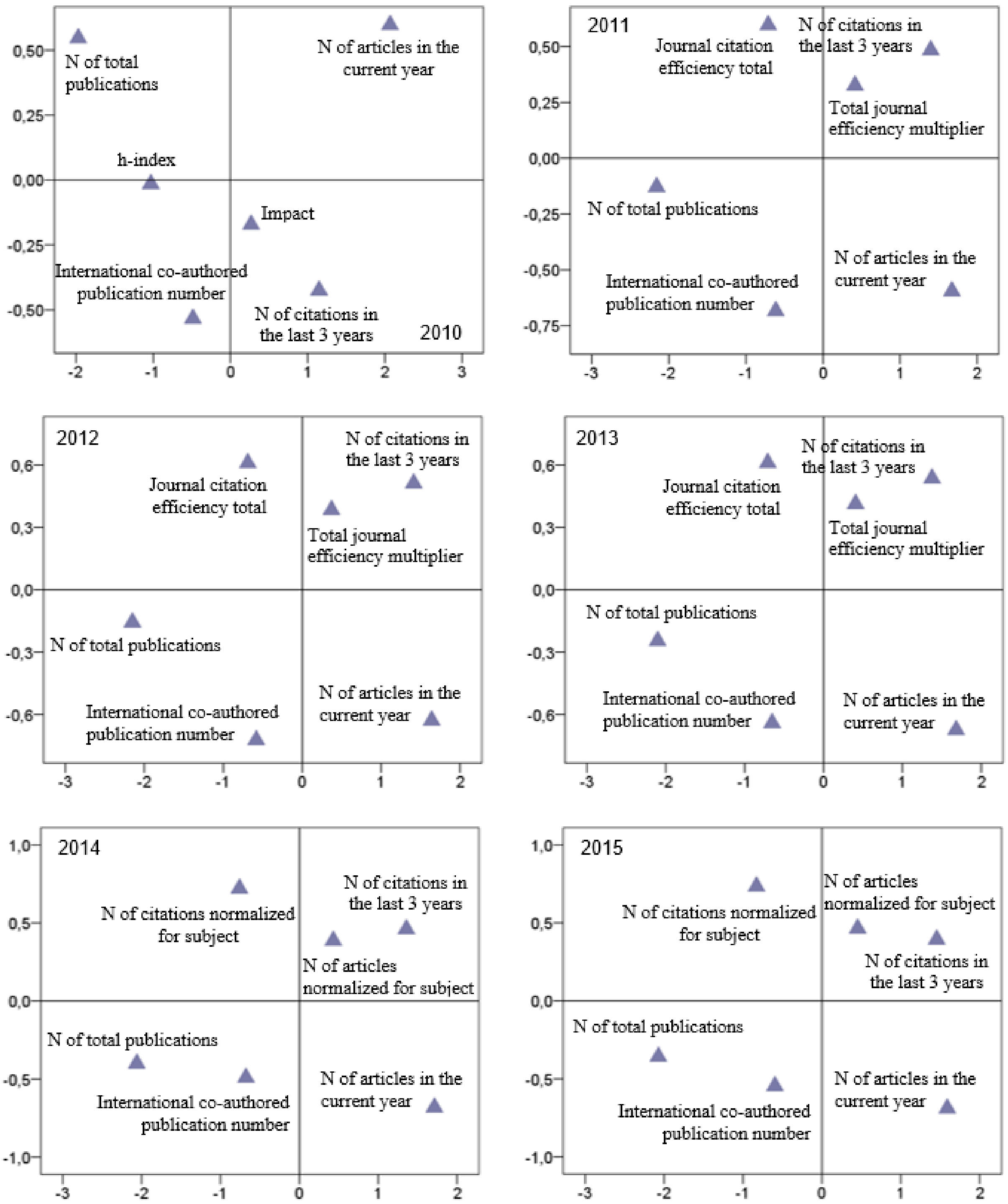

Figure 3. Results of MDS analysis for indicators of URAP rankings (2010–2015)

We calculated S-Stress values higher than 0.200 for QS (2012-2015) and RUR (2010-2015) that indicate low concordance (between 0.216-0.283 for QS and 0.313-0.326 for RUR). R^2 values also indicate relatively low variance explained compared to other rankings included in this research (between 0.567–0.886 for QS and 0.632-0.667 for RUR). Although it isn't statistically possible to draw any conclusion on the QS and RUR in terms of the S-Stress and R^2 values calculated, it appears that the indicators form groups according to their positions. The most obvious result for QS is that international student rates and international faculty rates are close to each other in both dimensions. These two indicators were found to be among the most similar indicators, with both considering the cosine similarity matrices generated for each year (0.912–0.917). It can be seen that the indicators of academic reputation survey and teaching reputation survey, which are the most similar indicators in 2013–2015 (0.952–0.956), are very close, particularly in 2015. In addition, according to the similarity matrices, academic reputation surveys and number of citations per faculty are among the most similar indicators (0.889–0.890). On the other hand, some indicators of RUR are highly similar and very close on biplots for almost each year analysed. Research reputation survey indicator is very similar with two other indicators: teaching reputation survey (0.985-1.000) and international teaching reputation (0.959-0.985). International teaching reputation has also a high similarity with teaching reputation survey with cosine similarity values between 0.961 and 0.987. The other two indicators that are very similar are the number of citations per faculty and the number of publications per faculty (0.978-0.984). The number of students per faculty and the number of undergraduate degrees per faculty are highly similar two indicators with the cosine similarity values between 0.947 and 0.950. Two indicators under the category of international diversity were founded very similar, that are international student rate and international bachelors rate (between 0.961-0.968 except for 2011 that is 0.857 and for 2014 that is 0.842). Institutional income per students and per faculty are two financial indicators that are very similar with the cosine similarity values between 0.949 and 0.956.

Discussion and Conclusion

In this study, we investigated the similarity of indicators used in research-focused international university rankings (ARWU, NTU, URAP, QS, and RUR) over years. In this context, similar indicator groups were determined and the effect of these similar indicators on overall rankings were presented. Because we calculated above 0.200 S-Stress values for QS and RUR, it is statistically not possible to reach any conclusion from the analysis of these two rankings.

One of the findings of this study is that ARWU, URAP and NTU use similar intra-indicators, so a redundancy problem exists for these three rankings. For ARWU, this is a common finding of the studies analysing the correlation of ranking indicators (Liu and Cheng, 2005, Saisana *et al.*, 2011, Shehatta and Mahmood, 2016, Soh, 2011). In a study investigating the correlation between the indicators of the ARWU 2004 ranking (Liu and Cheng, 2005, pp. 132-133), there were higher than 0.80 correlation values between the total number of publications and all other indicators, which interpreted as “compliance of the indicators.” However, high correlation and similarity refer to redundant indicators and require reduction of indicators in that the existing indicators measure the same thing (Van Noorden, 2010). Furthermore, correlation between the indicators that are used for a composite measure is a statistical methodological problem, and various statistical methods have been developed for dimension reduction (e.g., principal components analysis, factor analysis, and structural equation modelling). Additionally, it also points to a waste of resources for the ranking bodies. In another study (Shehatta and Mahmood, 2016, p. 1240) that examined the correlation between the indicators of the ARWU 2015 ranking by the top 100 universities, the most relevant indicators were also the same as those found in this study: the number of highly cited researchers and the number of articles in Nature and Science. Similar to this study, Liu and Cheng (2005, p. 132) examined the correlation between the indicators of the ARWU 2014 ranking and found the lowest correlations (0.50 and 0.55) between number of articles and the two indicators related to awards, a finding that is consistent with that of this study. The same result was obtained by another study (Saisana *et al.*, 2011) that examined the correlations between the indicators of the ARWU 2008 ranking (0.48 and 0.52). This finding was also supported by the findings of another study (Soh, 2011) that examined the inter-correlations of three existing international university rankings (ARWU, THE, and QS) using one-year ranking data and top-ranked universities (ARWU 2010 top 100, QS 2008 top 200, THE 2010 top 200). The number of highly cited researchers and the number of articles in Nature and Science were found to be the most relevant for ARWU (0.87). The most relevant indicators for QS with a moderate correlation coefficient of 0.64 were international faculty rate and international student rate. According to this study two indicators of ARWU explained 94% of variance in overall score, two categories of THE explained 94% of variance in overall score, and three indicators for QS explained 91% of variance in overall score and were “sufficient”. In other words, the indicators outside of the indicated ones are redundant for each of these rankings (Soh, 2011).

With this study, we reinforced the findings of the previous studies that intra-indicators of ARWU was similar. This study differs from the priors conducting investigating the similarity of intra-indicators of URAP, NTU and RUR for the first time. Although the results for QS and THE were reported in previous studies, we found that it was statistically not correct to conclude from the analysis conducted for QS, and it was not possible to analyse THE data (as indicator-based scores were not accessible). Considering more university rankings, using data of all ranking years and all universities in ranking lists, and using a nonparametric multivariate statistical technique

(MDS) for the analysis make this study differ from the priors. We found that it was possible to form indicator groups in terms of their similarities for ARWU, NTU, and URAP. Using only one of the indicators from the similar indicator group each time, we created new overall rankings for each of these three university rankings and we found very high correlations between the existing overall rankings and these new overall rankings we created.

It is not possible to measure the “quality of a university”, which is supposed to be measured by international ranking systems, by using the relevant data, methodologies and indicators. The fact that the international ranking systems, in particular the indicators they claim to measure the research activities, are too similar to each another, point to resource waste as well as a statistical methodological problem of ranking systems.

While scores generate very little position change at the top of the ranking lists, it appears that they lead to very large differences as the list goes down. As universities tend to use the results of the ranking lists according to their aims, they naturally give their rankings prominence instead of their scores, and they are not willing to face the fact that 400-rank difference in their positions is actually caused by 0.3 or 0.4 points. For example, in the overall ranking of CWUR 2017, there is only a 0.3 points difference between the 1000th and 711th universities.

Although there are some recent approaches that indicate “university rankings can be effective in the decisions of the people with the information they provide, but the ultimate decision-maker is the person him/herself” (Degener, 2014) and some applications such as U-Multirank and Mapping Scientific Excellence, which compare universities but do not aim to sort them, these new perspectives are less likely to attract interest than overall ranking lists or rankings offered by commercial firms, indicating that unwanted results tend to persist in comparison among universities or in revealing the “success” of a university. In response to a sudden jump in the rankings of several Canadian universities, Usher (2017) made a comment that points out the changes in the ranking methodology and stated, “the methodologists can take back what they had given”. Moreover, it is noteworthy that the ranking systems that are expected to make the evaluation of the university performances have also been the matter of evaluation, in recent years (Butler, 2010). Thus, ranking systems should not be taken as a basis in the formation of national and institutional policies (Holmes, 2006). Another important problem is that, universities cannot fulfil their other important duties as they try to increase their performance for certain indicators because they are too focused on the rankings and take their places in these ranking systems too serious (Rauhvargers, 2013). On the other hand, the important thing for a university should be to focus on what they want to accomplish nationally and internationally rather than to place higher in the ranking systems, and to develop its own criteria for success rather than adopting the standards developed by others (Degener, 2014). It is frequently mentioned that rankings do not reflect the research quality and many different measures must be taken into consideration in evaluations. For this purpose, the importance of qualitative evaluations (peer evaluation) to be done by experts in a certain field is emphasized (Frey and Rost, 2010; Shin, 2011, p. 32; Van Raan, 2005a).

The results of this study indicate that caution should be given when developing any policy, based on rankings. For example, if they want to make decisions based on ranking results, policymakers can focus on indicator-based rankings instead of overall rankings. This will allow them to make more accurate interpretations. It should also be noted here that using the ranking results, as a stand-alone evaluation tool will lead to incorrect results. Prior to using ranking systems as an evaluation tool, their methodology and criteria should be thoroughly examined and it should be accurately determined what exactly is measured and how. On the other hand, for example, rectors should be more precise when evaluating or using the results of these rankings. The objectives of the university administrations should not be to raise the universities’ ranks. Thus, the problem of targeting the indicators for which scores can be easily raised, and hence raising the scores of others associated with those indicators through manipulation, will cease to exist. Before looking at the position of universities in ranking results, university administrations must understand the indicators and methodological details used in the ranking system. The results of the study also show that developers of rankings can make more efficient rankings with fewer indicators, hence with fewer amounts of time and money. Developers of these systems should understand that it is a more correct approach to produce indicator-based ranking by giving up general ranking that is made by weighting the indicators.

The results of this study can be used by ranking bodies to decrease the number of indicators of their rankings, thus make use of their resources more effectively. Policy- and decision-makers can utilize the findings of this study focusing on indicator-based rankings instead of focusing on overall rankings and could question decision-making based on overall rankings of universities, which is a composite measure calculated with highly correlated and similar indicators. Correspondingly, a similar suggestion can be made for all potential audiences of rankings, namely, students, academics, researchers etc. They can use these findings to scrutinize the rankings with a different approach and think twice before equalizing the rank with quality.

References

A guide to the Nature Index. (2017, 23 March). *Nature*, Vol. 543 No. 33. doi:10.1038/543S33a

Abramo, G., Cicero, T. and D'Angelo, C. A. (2013). "The impact of non-productive and top scientists on overall university research performance", *Journal of Informetrics*, Vol. 7 No. 1, pp. 166-175.

Aguillo, I. F., Bar-Ilan, J., Levene, M. and Ortega, J. L. (2010). "Comparing university rankings", *Scientometrics*, Vol. 85 No. 1, pp. 243-256.

ARWU. (2017). "Methodology", available at: http://www.shanghairanking.com/ARWU-Methodology 2016.html#2 (accessed December 24, 2017).

ARWU. (2015a). "About academic ranking of world universities", available at: http://www.shanghairanking.com/aboutarwu.html (accessed December 24, 2017).

ARWU. (2015b). "Methodology", available at: http://www.shanghairanking.com/ARWUMethodology-2014.html (accessed December 24, 2017).

ARWU. (2014). "Alternative ranking 2014 (Excluding award factor)", available at: http://www.shanghairanking.com/Alternative_Ranking_Excluding_Award_Factor/Excluding_Award_Factor2014.html (accessed December 24, 2017).

Baty, P. (2017, February 8). "This is why we publish the World University Rankings" [Blog post], available at: https://www.timeshighereducation.com/blog/why-we-publishworld-university-rankings (accessed December 24, 2017).

Billaut, J. C., Bouyssou, D. and Vincke, P. (2010). "Should you believe in the Shanghai ranking? An MCDM view", *Scientometrics*, Vol. 84 No. 1, pp. 237-263.

Bookstein, F. L., Seidler, H., Fieder, M. and Winckler, G. (2010). "Too much noise in the Times Higher Education rankings", *Scientometrics*, Vol. 85 No. 1, pp. 295-299.

Borg, I. and Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). United States of America: Springer.

Borg, I., Groenen, P. J. F. and Mair, P. (2013). *Applied multidimensional scaling*. Heidelberg: Springer.

Bornmann, L. and Mutz, R. (2015). "How well does a university perform in comparison with its peers? The use of odds, and odds ratios, for the comparison of institutional citation impact using the Leiden rankings?" [Brief communication], *Journal of the Association for Information Science and Technology*, Vol. 66 No. 12, pp. 2711-2713.

Bornmann, L., Mutz, R. and Daniel, H-D. (2013). "Multilevel-statistical reformulation of citation-based university rankings: The Leiden Ranking 2011/2012", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 8, pp. 1649-1658.

Bornmann, L., Stefaner, M., Moya Anegón, F. and Mutz, R. (2014). "The scientific excellence mapping tool", [Correspondence]. *European Science Editing*, Vol. 40 p. 28, available at: http://europeanscienceediting.eu/wpcontent/uploads/2016/05/corresp_1.pdf html (accessed December 24, 2017).

Butler, D. (2010, 3 March). "University rankings smarten up" [Special report], *Nature*, Vol. 464 No. 7285, pp. 16-17.

Chen, K. and Liao, P. (2012). "A comparative study on world university rankings: A bibliometric survey", *Scientometrics*, Vol. 92 No. 1, pp. 89-103.

Claassen, C. (2015). "Measuring university quality", *Scientometrics*, Vol. 104 No. 3, pp. 793-807.

Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional scaling* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Daraio, C., Bonaccorsi, A. and Simar, L. (2015). "Rankings and university performance: A conditional multidimensional approach", *European Journal of Operational Research*, Vol. 244 No. 3, pp. 918-930.

Degener, J. (2014, June 4). "Controversial and powerful: University rankings under scrutinity", *Euro Scientist*, available at: http://www.euroscientist.com/controversialpowerful-university-rankings-under-scrutinity/ (accessed December 24, 2017).

Dehon, C., McCathie, A. and Verardi, V. (2010). "Uncovering excellence in academic rankings: A closer at the Shanghai ranking", *Scientometrics*, Vol. 83 No. 2, pp. 515-524.

Dobrota, M., Bulajic, M., Bornmann, L. and Jeremic, V. (2016). "A new approach to the QS university ranking using the composite I-distance indicator: Uncertainty and sensitivity analyses", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 1, pp. 200-211.

Dobrota, M. and Dobrota, M. (2016). "ARWU ranking uncertainty and sensitivity: What if the award factor was excluded?", [Brief communication]. *Journal of the Association for Information Science and Technology*, Vol. 67 No. 2, pp. 480-482.

Docampo, D. and Cram, L. (2015). "On the effects of institutional size in university classifications: The case of the Shanghai ranking", *Scientometrics*, Vol. 102 No. 2, pp. 1325-1346.

Doğan, G. (2017). *Akademik performans odaklı uluslararası üniversite sıralama sistemlerinin genel sıralamalarına ve ölçütlerine göre değerlendirilmesi* [*Evaluation of academic performance oriented international university ranking systems based on general rankings and criteria*]. Unpublished Ph.D. Dissertation, Hacettepe University.

Enders, F. B. (2013, December 24). “Collinearity”, *Encyclopedia Britannica*, available at: https://www.britannica.com/topic/collinearity-statistics (accessed August 18, 2018).

Ewalt, D. (2016, September 28). “Reuters top 100: The world’s most innovative universities – 2016”, available at: http://www.reuters.com/article/amers-reuters-rankinginnovative-univers-idUSL2N1C406D (accessed December 24, 2017).

Florian, R. V. (2006). “Irreproducibility of the results of the Shanghai academic ranking of world universities”, *Scientometrics*, Vol. 72 No. 1, pp. 25-32.

Frey, B. S. and Rost, K. (2010). Do rankings reflect research quality?. *Journal of Applied Economics*, Vol. 13 pp. 1-38. doi:10.1016/S1514-0326(10)60002-5

Holmes, R. (2010). “The THE-QS world university rankings, 2004-2009”, *Asian Journal of University Education*, Vol. 6 No. 1, pp. 91-113, available at: http://education.uitm.edu.my/v1/images/stories/formtodownload/AJUE/ajue61.pdf (accessed December 24, 2017).

Holmes, R. (2006). “The THES university rankings: Are they really world class”, *Asian Journal of University Education: Faculty of Education*, Vol. 2 No. 1, pp. 1-14, available at: http://education.uitm.edu.my/v1/images/stories/formtodownload/AJUE/ajue21.pdf (accessed December 24, 2017).

Jajo, N. K. and Harrison, J. (2014). “World university ranking systems: an alternative approach using partial least squares path modelling”, *Journal of Higher Education Policy and Management*, Vol. 36 No. 5, pp. 471-482. doi:10.1080/1360080X.2014.936090

Jöns, H. and Hoyler, M. (2013). “Global geographies of higher education: The perspective of world university rankings”, *Geoforum*, Vol. 46, pp. 45-59. doi:10.1016/j.geoforum.2012.12.014

King, R. (2009). *Governing universities globally: Organizations, regulation and rankings*. Edward Elgar Publishing Limited: United Kingdom.

Kryszkiewicz, M. (2014). “The cosine similarity in terms of the euclidean distance”, *The Encyclopedia of Business Analytics and Optimization* (pp. 2498-2508). doi:10.4018/978-1-4666-5202-6.ch223

Li, B. and Han, L. (2013) “Distance weighted cosine similarity measure for text classification”, H. Yin et al. (Ed.) *Intelligent Data Engineering and Automated Learning - IDEAL 2013. Lecture Notes in Computer Science, 8206*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-41278-3_74

Liu, N. C. and Cheng, Y. (2005). “The academic ranking of world universities”, *Higher Education in Europe*, Vol. 30 No. 2, pp. 127-136.

Liu, N. C., Cheng, Y. and Liu, L. (2005). “Academic ranking of world universities using scientometrics - a comment to the “fatal attraction””, *Scientometrics*, Vol. 64 No. 1, pp. 101-109.

Lo, W. Y. W. (2014). *University rankings: Implications for higher education in Taiwan*. Springer: Singapore.

Loughran, G. (2016, September 19). “Why university rankings may be harming higher education”, *The Irish Times*, available at: http://www.irishtimes.com/news/education/why-university-rankings-may-be-harminghigher- education-1.2793532 (accessed December 24, 2017).

Marginson, S. (2007). “Global university rankings: where to from here?”, *Asia-Pacific Association for International Education, National University of Singapore, 7-9 March 2007*, available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.662.8013&rep=rep1&type=pdf (accessed May 20, 2018).

Markovic, A., Jeremic, V. and Zornic, N. (2014). “How the top 500 ARWU can provide a misleading rank” [Letter to editor], *Journal of the Association for Information Science and Technology*, Vol. 65 No. 6, pp. 1303-1304.

Moed, H. F. (2017). “A critical comparative analysis of five world university rankings”, *Scientometrics*, Vol. 110, No. 2 pp. 967-990.

Morse, R., Krivian, A. and Jackwin, A. (2016, October 24). “How U.S. News calculated the best global universities rankings”, available at: https://www.usnews.com/education/best-global-universities/articles/methodology (accessed May 20, 2018).

Multidimensional Scaling. (2017). “MDC and factor analysis”, available at: http://www.statsoft.com/Textbook/Multidimensional-Scaling#index (accessed May 20, 2018).

Nature Index. (2017). “Institution outputs”, available at: http://www.natureindex.com/institutionoutputs/generate/All/global/All/weighted_score (accessed December 24, 2017).

NTU. (2017a). “Performance Ranking of Scientific Papers for World Universities 2017 – Background”, available at: http://nturanking.lis.ntu.edu.tw/BackgroundMethodology/Background-enus.aspx (accessed December 24, 2017).

NTU. (2017b). “Performance Ranking of Scientific Papers for World Universities 2017 – Methodology”, available at: http://nturanking.lis.ntu.edu.tw/BackgroundMethodology/Methodology-enus.aspx#2 (accessed December 24, 2017).

NTU. (2015). “Performance ranking of scientific papers for world universities 2014 – Methodology”, available at: http://nturanking.lis.ntu.edu.tw/BackgroundMethodology/Methodology-enus.aspx (accessed December 24, 2017).

Piro, F. N. and Sivertsen, G. (2016). “How can differences in international university rankings be explained?”, *Scientometrics*, Vol. 109 No. 3, pp. 2263-2278.

QS. (2017). “Methodology”, available at: https://www.topuniversities.com/qs-world-university-rankings/methodology (accessed December 24, 2017).

QS. (2015). “QS world university rankings: Methodology”, available at: http://www.topuniversities.com/university-rankings-articles/world-universityrankings/qs-world-university-rankings-methodology (accessed December 24, 2017).

Rauhvargers, A. (2014). Where are the global rankings leading us? An analysis of recent methodological changes and new developments, *European Journal of Education*, Vol. 49 No. 1, pp. 29-44.

Rauhvargers, A. (2013). “Global university rankings and their impact (Report II)”, available at: http://www.eua.be/Libraries/publications-homepage-list/EUA_Global_University_Rankings_and_Their_Impact_-_Report_II.pdf?sfvrsn=2 (accessed August 2, 2018).

Reuters. (2017). “Methodology: Top 100 innovative universities 2016”, available at: http://www.reuters.com/most-innovative-universities-2016/methodology (accessed December 24, 2017).

RUR. (2018). “Round University Ranking”, available at: http://roundranking.com (accessed August 1, 2018).

RUR. (2017). “Round University Ranking Methodology”, available at: http://roundranking.com/library/methodology.html (accessed December 24, 2017).

Saisana, M., d’Hombres, B. and Saltelli, A. (2011). “Rickety numbers: Volatility of university rankings and policy implications”, *Research Policy*, Vol. 40 No. 1, pp. 165-177.

Shehatta, I. and Mahmood, K. (2016). “Correlation among top 100 universities in the major six global rankings: policy implications”, *Scientometrics*, Vol. 109 No. 2, pp. 1231-1254.

Shin, J. C. (2011). “Organizational effectiveness and university rankings”, J. C. Shin, R. K. Toutkoushian and U. Teichler (Ed.), *University rankings: Theoretical basis, methodology and impacts on global higher education* (pp. 19-34). The changing academy - The changing academic profession in international comparative perspective 3. Springer: Dordrecht

Shin, J. C. and Toutkoushian, R. K. (2011). “The past, present, and future of university rankings”, J. C. Shin, R. K. Toutkoushian and U. Teichler (Ed.), *University rankings: Theoretical basis, methodology and impacts on global higher education* (pp. 1-16). The changing academy - The changing academic profession in international comparative perspective 3. Springer: Dordrecht

SIR. (2017a). “Scimago institutions rankings”, available at: http://www.scimagoir.com (accessed December 24, 2017).

SIR. (2017b). “SIR methodology”, available at: http://www.scimagoir.com/methodology.php (accessed December 24, 2017).

Soh, K. V. (2011). “Don’t read university rankings like reading football league tables: Taking a close look at the indicators”, *Higher Education Review*, Vol. 44 No. 1, pp. 15-29. https://www.researchgate.net (accessed December 24, 2017).

Soh, K. (2013a). “Rectifying an honest error in world university rankings: A solution to the problem of indicator weight discrepancies”, *Journal of Higher Education Policy and Management*, Vol. 35 No. 6, pp. 574-585.

Soh, K. (2013b). “Misleading university rankings: cause and cure for discrepancies between nominal and attained weights”, *Journal of Higher Education Policy and Management*, Vol. 35 No. 2, pp. 206-214.

Soh, K. (2017) “The seven deadly sins of world university ranking: a summary from several papers”, *Journal of Higher Education Policy and Management*, Vol. 39 No. 1, pp. 104-115.

Stuart, D. L. (1995). Reputational rankings: Background and development. New Directions For Institutional Research, Vol. 1995, No. 88, pp. 13-20.

Tofallis, C. (2012). “A different approach to university rankings”, *Higher Education*, Vol. 63, No. 1, 1-18.

U-Multirank. (2017a). “The project”, available at: http://www.umultirank.org/#!/about/project?trackType=about&sightMode=undefined (accessed December 24, 2017).

U-Multirank. (2017b). “Our approach to ranking”, available at: http://www.umultirank.org/#!/about/methodology/approach-toranking?trackType=about&sightMode=undefined (accessed December 24, 2017).

UI GreenMetric. (2015). “Welcome to UI GreenMetric”, available at: http://greenmetric.ui.ac.id/what-is-greenmetric/ (accessed December 24, 2017).

uniRank. (2017). “About us”, available at: http://www.4icu.org/about/ (accessed December 24, 2017).

URAP. (2015a). “Citation impact total (CIT)”, available at: http://www.urapcenter.org/2014/indicator.php?q=5 (accessed December 24, 2017).

URAP. (2015b). “World ranking indicators”, available at: http://www.urapcenter.org/2014/methodology.php?q=3 (accessed December 24, 2017).

URAP. (2015c). “Citation”, available at: http://www.urapcenter.org/2014/indicator.php?q=2 (accessed December 24, 2017).

URAP. (2015d). “International collaboration”, available at: http://www.urapcenter.org/2014/indicator.php?q=6 (accessed December 24, 2017).

Usher, A. (2017, 20 February). “Canada’s rankings run-up. Higher Education Strategy Associates”, available at: http://higheredstrategy.com/canadas-rankings-run-up/ (accessed August 1, 2018).

Usher, A. and Savino, M. (2006). “World of difference: A global survey of university league tables (Report)”, Canadian Education Report Series, Toronto: Educational Policy Institute, available at: http://www.educationalpolicy.org/pdf/World-of-Difference-200602162.pdf (accessed December 24, 2017).

Van Noorden, R. (2010). “Metrics: A profusion of measures”, *Nature*, Vol. 465, pp. 864-866. doi:10.1038/465864a

Van Raan, A. F. J. (2005a). “Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods”, *Scientometrics*, Vol. 62 No. 1, pp. 133-143.

Van Raan, A. F. J. (2005b). “Reply to the comments of Liu et al” [Correspondence], *Scientometrics*, Vol. 64 No. 1, 111-112.

Van Raan, A. F. J. (2005c). “Challenges in ranking of universities”, *Invited paper for the First International Conference on World Class Universities, Shanghai Jaio Tong University, Shanghai, June 16-18, 2005*, available at: http://www.shanghairanking.com/wcu/wcu1/van%20Raan.pdf (accessed May 20, 2018).

Van Vught, F. A. and Ziegele, F. (Ed.) (2012). *Multidimensional ranking: The design and development of u-multirank*. Springer: Dordrecht.

Vincke, P. (2009). “University rankings”, D. Jacobs and C. Vermandele (Ed.) *Ranking universities, Brussels, 2009* (pp. 11-26). Edition de l’Universit´e de Bruxelles, available at: https://dipot.ulb.ac.be/dspace/bitstream/2013/101994/1/14%20Vincke%20Ranking%20universities.pdf (accessed May 20, 2018).

Whitcroft, A. (2013, August 8). “Mapping scientific excellence – ranking the world’s best scientific organisations” [Blog post], available at: http://sciblogs.co.nz/miscience/ 2013/08/08/mapping-scientific-excellence-ranking-the-worlds-best-scientificorganisations/ (accessed December 24, 2017).

Williams, R. (2008). “Methodology, meaning and usefulness of rankings”, *Australian Universities’ Review*, Vol. 50 No. 2, pp. 51-58, available at: https://issuu.com/nteu/docs/aur_50-02 (accessed December 24, 2017).

Youth Incorporated. (2015). “Global University Rankings 2015”, available at: http://www.youthincmag.com/wp-content/uploads/ranking/55e75554ab887.pdf (accessed December 24, 2017).