

---

## ESTUDIOS / RESEARCH STUDIES

---

### Model for estimating Academic Ranking of World Universities (Shanghai Ranking) scores

A. Pandiella-Dominique\*, L. Moreno-Lorente\*\*, C. García-Zorita\*, E. Sanz-Casado\*

\* Department of Library and Information Science, Carlos III University of Madrid, Spain  
e-mail: [apandiel@bib.uc3m.es](mailto:apandiel@bib.uc3m.es) | ORCID iD: <https://orcid.org/0000-0002-6572-2317>  
e-mail: [czorita@bib.uc3m.es](mailto:czorita@bib.uc3m.es) | ORCID iD: <https://orcid.org/0000-0002-6860-8069>  
[elias@bib.uc3m.es](mailto:elias@bib.uc3m.es) | ORCID iD: <https://orcid.org/0000-0002-0188-7489>

\*\* Research Institute on Higher Education and Science (INAECU), Madrid, Spain.  
Department of Systems and Automation, Carlos III University of Madrid, Madrid, Spain  
e-mail: [moreno@ing.uc3m.es](mailto:moreno@ing.uc3m.es) | ORCID iD: <https://orcid.org/0000-0002-4450-680X>

Received: 13-2-2017; 2<sup>nd</sup> version: 30-08-17; Accepted: 28-09-2017.

**Citation/Cómo citar este artículo:** Pandiella-Dominique, A.; Moreno-Lorente, L.; García-Zorita, C.; Sanz-Casado, E. (2018). Model for estimating Academic Ranking of World Universities (Shanghai Ranking) scores. *Revista Española de Documentación Científica*, 41 (2): e204. <https://doi.org/10.3989/redc.2018.2.1462>

**Abstract:** The interest in global university rankings has grown significantly in the last 10 years. The use of just a handful of indicators, the ease of interpretation of the information they contain and the furtherance of inter-university competition and comparability are some of the factors that have popularised their implementation. However, at the same time their critics have identified some conceptual, technical and methodological issues. This paper addresses three such issues that have prompted intense methodological debate around university rankings: replicability of results, relevance of indicators and data retrieval. It also proposes a tool for estimating the scores for the two indicators of the greatest interest for most universities (Papers published in *Nature* or *Science* and Papers listed in the WoS). It reports on an alternative method developed to calculate any university's score in the two most significant Shanghai ranking indicators. One of the foremost features of the proposed method is that the inputs needed are readily available to policymakers, academic authorities, students and other stakeholders and can be applied directly. Furthermore, with this model, scores can also be estimated for universities not listed among the first 500 in the Shanghai ranking.

**Keywords:** international university rankings; research evaluation; Shanghai ranking.

#### Modelo de estimación de los indicadores del Academic Ranking of World Universities (Shanghai Ranking)

**Resumen:** El interés en los rankings globales de universidades ha crecido significativamente a lo largo de los últimos 10 años. El uso de indicadores simples y sintéticos, la facilidad de interpretación de la información que contienen, el fomento de la competencia entre las universidades, así como la posibilidad de compararlas entre sí, son algunos de los factores que han popularizado su aplicación. Al mismo tiempo, sus críticos han identificado problemas en relación a cuestiones tanto conceptuales como técnicas y metodológicas. Este artículo aborda tres temas que han suscitado un intenso debate metodológico sobre los rankings de universidades: la replicabilidad de los resultados, la relevancia de los indicadores, y la recopilación de datos. También propone una herramienta para estimar las puntuaciones de los dos indicadores de mayor interés para la mayoría de las universidades (Documentos publicados en *Nature* o *Science* y publicaciones en WoS). Se informa sobre un método alternativo desarrollado para calcular la puntuación de cualquier universidad en los dos indicadores de ranking más importantes de Shanghái. Una de las principales características del método propuesto es que los input necesarios son fácilmente accesibles para los gestores de política científica, autoridades académicas, estudiantes, y otros grupos de interés, pudiéndose aplicar directamente. Además, con este modelo también se pueden estimar las puntuaciones para las universidades que no figuran entre las primeras 500 en el ranking de Shanghái.

**Palabras Clave:** Rankings internacionales de universidades; evaluación de la investigación; Ranking de Shanghái.

**Copyright:** © 2018 CSIC. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

## 1. INTRODUCTION

The interest in global university rankings has risen exponentially in recent years (Fowles et al., 2016; Gonzalez-Riano et al., 2014; Safón, 2013). Their popularity is largely due to the fact that they furnish readily understandable, simple and synthetic information, favouring comparisons among intra- or cross-country higher education institutions (Hazelkorn, 2014). This approach has also been criticised, however, for it assumes that academic performance can be assessed with the same simplicity as football teams (van Raan, 2005a). For other authors (Pusser and Marginson, 2013), rankings may be selective, for they ‘...serve as a particularly useful lens for studying power in higher education, as they are used to confer prestige, in the allocation of resources, as a form of agenda setting, as a means of stratifying national higher education systems...’

The appearance of international university rankings may be said to have had beneficial implications for higher education systems, for it has required universities to adapt to a reality fully assumed by other social sectors, namely the need to survive in increasingly complex and competitive global environments. Regarding this, Andersson and Mayer (2017) have pointed out that rankings can provide a “picture of the university concerned combining a variety of pieces of information and, before rankings, there was no measure or semi-quantitative comparison available and there was an opacity in the system”. Nonetheless, the harsh criticism and heated debate to which rankings are subject must not be overlooked. The most prominent beneficial factors include increasing competition among universities and a more attentive focus on measuring higher education institutions’ performance for assessment (Rauhvargers, 2013; Sanz-Casado, 2015; Olcay & Bulu, 2016).

Some of the issues addressed by critics are conceptual, questioning, among others, the capacity of such tools to measure and rank very heterogeneous universities with the same parameters (Rauhvargers, 2011; Andersson & Mayer, 2017), the trend to assess all universities to the same criteria as the top 500 (Rauhvargers, 2011), to rank institutions by size (van Raan, 2005a; Docampo and Cram, 2015; Docampo et al., 2015), favouring those heavy on research (Rauhvargers, 2011) or to prioritise English-speaking establishments (Marginson & van der Wende, 2007; van Raan et al., 2011; Rauhvargers, 2011). In addition, the absence of teaching measures and the lack of recognition to the local or national contribution of the universities have been a shortcoming recognized by international ranking organizations ever since (O’Leary, 2017).

Another conceptual problem is the inequality among different subject areas. Ways of organizing researchers and resources, ways of publication and citation or ways of embedding research teams are not equal in different subject areas, and general (Turner, 2017).

One of the technical factors called into question is the capacity of databases to correctly assign authors’ names and affiliations (van Raan, 2005a; Moed, 2002). The solution to this problem would call for careful data cleansing as well as clear and standardised guidelines. As these technical problems affect 60 % of the Academic Ranking of World Universities (hereafter ARWU) indicators, they introduce considerable uncertainty (van Raan, 2005a). Problems have also arisen around the differences in the values accorded to journals, citations and so on, depending on the database queried (Fiala, 2012; Lascurain et al., 2015).

Regarding methodology, Robinson and Jimenez-Contreras (2017) have pointed out that “the development of proper methodologies for the elaboration of research rankings is an on-going research front in which many variables unexplored and questions still remain unsolved”. In this sense, methodological problems have also been identified in connection with selection criteria (Zornic et al., 2014) and indicator weighting (Stewart, 2014). The scant relevance and discriminatory capacity of factors based on Nobel Prizes or Fields Medals have been challenged, for instance, along with the justification for the choice of indicator time periods or the bias toward research-based criteria (van Raan, 2005a; Billaut et al., 2010). The reliability of the findings (van Raan, 2005a, 2005b; Saisana et al., 2011) and the scant replicability of ranking results (Florian, 2007; Docampo, 2013; van Raan, 2005a; van Raan, 2005b; Jovanovic et al., 2012) have also been criticised.

The perspective adopted here is that solving formulation-related problems holds the key to strengthening the prestige and credibility of rankings among the scholastic community, as well as their ‘practical utility’ as an intra-university assessment tool and a benchmarking instrument for inter-university comparisons.

This paper addresses three controversial matters in connection with the methodology used in the Shanghai ranking. Firstly, the debate revolves around the suitability of the most relevant indicators. Here criticism has focused on the arbitrariness of evaluations based on Nobel Prizes or Fields Medals, a debate that has apparently been internalised, for the ARWU has now launched the possibility of establishing a ‘related’ ranking excluding these

two indicators (Dobrota & Dobrota, 2016). The arguments against Nobel Prize- or Fields Medal-based indicators include the following:

- These indicators are geared to measuring the prestige of only a few universities but are scantily able to evaluate the quality of their student bodies or academic staff. They are barely related to the object of the measurement.
- Moreover, since they add to very few universities' scores, they are assumed to lack the capacity to discriminate among all the institutions assessed. Of the 2000 universities evaluated, approximately 7 % would be assigned points under the 'Award' indicator and only 10 % under 'Alumni'. Moreover, in the high scored group of top 500 universities, only 141 higher education institutions are assigned points for 'Award' and 208 for 'Alumni'.
- Some authors find the time period for assigning points unjustified (Billaut et al., 2010).
- Both indicators may be measuring university characteristics that are better assessed by other variables.
- These indicators fail to quantify universities' efforts to improve their quality, they limit recently founded (under 50 years) universities' visibility and consequently are overly conservative, essentially benefiting universities of consolidated prestige.

In another vein, factors related to the strategies followed to retrieve papers published and their subsequent formulation to compute the scores in the respective indicators are also debated. Problems have been detected in the assignment of points under the indicators based on article counts.

Lastly, inasmuch as some authors have pointed out the difficulty of replicating the findings following the methodology described in the ranking (Florian, 2007; Billaut et al., 2010; Docampo, 2013), this study puts forward a proposal as an alternative to the one recently suggested by Docampo (2013).

This paper puts forward a simpler alternative for estimating the points with which universities are scored under two of the ranking's key indicators: the number of articles in *Nature* and *Science* (N&S) and the number listed in the Web of Science (PUB), based on raw data (non-fractionated counts) obtained directly from the Web of Science (hereafter WoS) using the WoS search field 'Organisation-enhanced'.

It addresses three aspects of the methodological debate on university rankings: the relevance of the indicators, data retrieval and replicability of results. It proposes a tool for estimating the points assigned under these two indicators of highest interest for universities.

## 2. ACADEMIC RANKING OF WORLD UNIVERSITIES (ARWU)

The ARWU ranking appeared in 2003 in the wake of the success obtained with the '985 project' that analysed a number of Chinese institutions to help them attain 'world class university' status. In other words, this first global university ranking was created to determine the worldwide position of Chinese higher education institutions.

According to the methodology described by the founders, the values of each indicator are assigned proportionally, with the highest ranking institution scoring 100 (Liu et al., 2005; Florian, 2007). However, it was unable to replicate the findings when applying that methodology. Docampo (2013) proposed a formula for estimating an institution's final score for each indicator further to the methodology laid down in the ranking. Although that formula is simple and operational, a number of factors hinder the estimation of an institution's score, per indicator and overall: the difficulty of obtaining fractionated counts for articles and determining the search strategy for each institution involved are two such impediments.

### ARWU methodology and indicators

Every year, the ARWU analyses 2 000 universities and ranks 500 in a list published annually. Its methodology focuses on universities' teaching and research performance, assessing six indicators which it weights and classifies under four criteria.

The quality of the education delivered by the institution is based on the number of Nobel Prizes or Fields Medals earned by its alumni. As noted in the introduction, this indicator has been highly criticised.

Academic staff quality is determined by two indicators. The first is the number of professors holding a Nobel Prize or a Fields Medal, weighted depending on the time lapsing since the award. The objections raised to this indicator are likewise discussed in the introduction (Billaut et al., 2010). The second criterion is the number of the institution's researchers heavily cited in the 21 Thomson-Reuters scientific categories (Bornmann, & Bauer, 2015). One of the problems with this

**Table I.** ARWU Ranking for 2015: indicators and weights

Criterion	Indicator	Code	Weight
Quality of education	Institution's alumni awarded Nobel Prizes or Fields Medals	Alumni	10%
Quality of faculty	Institution's staff awarded Nobel Prizes or Fields Medals	Award	20%
	Highly cited researchers under 21 broad subject categories	HiCi	20%
Research output	Papers published in <i>Nature</i> or <i>Science</i>	N&S	20%
	Papers listed in the WoS Science Citation Index-expanded or the Social Sciences Citation Index	PUB	20%
Per capita performance	Institution's per capita academic performance	PCP	10%
Total			100%

indicator identified by some authors is the lack of justification for applying a 10-year timeframe (Billaut et al., 2010).

Institutions' scientific production is also measured with two indicators. The first is the number of articles published in *Nature* and *Science* in the last five years. This indicator weights authors by the order of appearance (corresponding author 100 %, first 50 %, second 50 % and all others 10 %). The lack of justification for this type of count as well as the problems surrounding institutional affiliation are two factors that have been criticised in connection with this indicator (Billaut et al., 2010). The second is the number of papers published by the institution's academic staff in journals listed in the two WoS indices, Science Citation Index-Expanded (SCI-E) and Social Sciences Citation Index (SSCI). Here one of the problems is that the number of papers in the SCI-E doubles the number in the SSCI.

Lastly, the ARWU ranking assesses institution productivity relative to size, calculated from the values of the five indicators described, divided by the number of full-time equivalent academic staff members.

Each institution is ranked in accordance with the findings for the aforementioned indicators and a value of 100 is assigned to the university with the highest score in each indicator. The percentage values of the remaining institutions are then recalculated on that basis.

### 3. RESEARCH OBJECTIVE

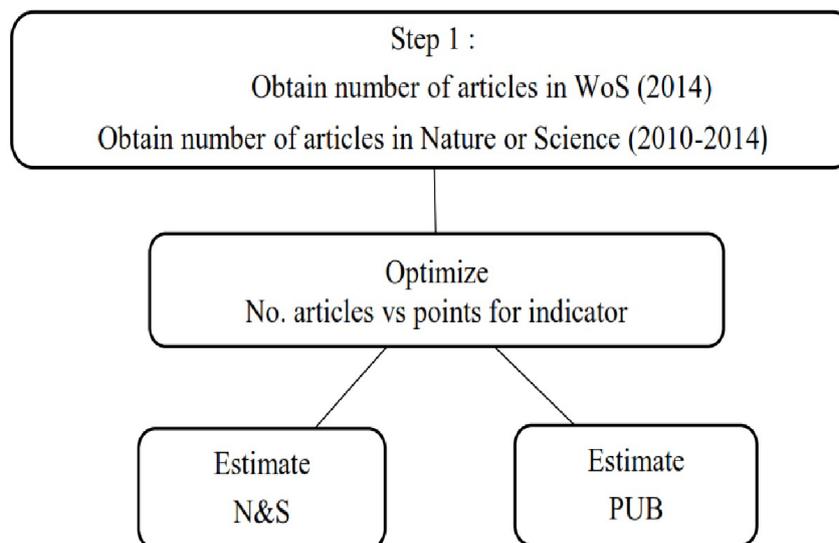
This article aims primarily to establish a reliable method for quickly and easily estimating universities' scores in the Shanghai ranking. The model proposed calculates the scores for research output indicators N&S and PUB based on simple Web of Science queries, with no need to standardise the raw data. It can consequently be used to find the score for any university, whether or not it is included in the Shanghai ranking.

### 4. METHODOLOGY

Firstly, the PUB score was estimated for all the institutions ranked in the 2015 edition of the ARWU based on the number of articles listed in the Web of Science indices in 2014 under each university's standardised name. The same procedure was followed to estimate the N&S indicator, except that the interval used was 5 years (2010-2014), i.e., the same period applied by the Shanghai ranking to calculate this indicator. The flow chart in Figure 1 shows the steps followed. In both cases gross (i.e., non-fractionated) counts were used. The data were obtained on 15 may 2016.

A two-step approach was followed to develop a mathematical model able to minimise the difference between the actual and the predicted points found for PUB and N&S:

- A number of mathematical models were first considered, most based on exponential or logarithmic functions. When, as in this case, the regression curve is non-linear, the problem is broached using the likewise non-linear regression function. Its expression is  $y=f(x,\theta)$ , where  $y$  is the output (the indicator),  $x$  is the data vector (the regressors, here the number of papers published by the institution and the number by the highest scoring institution) and  $\theta$  is the vector for the model parameters ( $\alpha$  and  $\beta$ ). The error is defined as the difference between actual output,  $y_T$ , and model output,  $y$ :  $e = y_T - y$ .
- Secondly, the model parameters ( $\alpha$  and  $\beta$ ) were selected to minimise the total error. Whilst any of a number of error functions can be used, quadratic cost,  $C_T = \sum_i e_i^2$  is the one most commonly applied. An alternative less sensitive to outliers, the absolute value of the error,  $C_T = \sum_i |e_i|$ , was also considered here.
- Several optimisation methods have been developed to determine the parameters that minimise the total error. Typically, the derivative

**Figure 1.** Flow chart for estimating N&S and PUB

of the quadratic cost function is equated to zero and the resulting equation solved, an approach that leads to the well known least squares method. This solution cannot be applied to the absolute value of the error, however, because the derivative is not properly defined at the minimum. As the aim sought in this study was to find an optimisation method compatible with either function, quadratic cost error or the absolute value of the error, the differential evolution optimisation method was adopted. As a sampling-based optimisation method, this procedure does not require the use of the derivative of the total error function and is consequently suitable for determining the minima or maxima of a wide spectrum of non-linear, non-differentiable and multivariate functions.

After testing several non-linear functions, the expression that provided the best fit to the problem posed was as shown in the following paragraph. The same basic form of the mathematical function was used to fit the input data to the output scores considered.

The input data for the N&S non-linear function were the number of articles published by institution (and the university with the largest number) in *Nature* and *Science*, and for PUB the total number of papers listed in the SSCI and SCI-E databases. The output consisted in the scores obtained for the indicator, i.e., the gross values for each institution, including the values for the highest-ranking university under that indicator and the scores obtained by the institution at issue in the

respective year. The resulting function estimated the scores for each institution in the two indicators. After analysing a number of non-linear functions, the one that afforded the best fit was as shown below:

$$y = \frac{\left(\frac{100(1+x)}{c}\right)^{0.625}}{\alpha * \log\left(\frac{c}{1+x}\right) + \beta}$$

where:

$y$  = N&S or PUB score for the institution studied.

$x$  = gross number of publications listed in the WoS under the name of the institution studied.

$c$  = gross number of publications listed in the WoS under the name of the institution with the largest number.

$\alpha$  = intercept of the regression curve ( $\alpha_{\text{N\&S}} = -0.006597$ ;  $\alpha_{\text{PUB}} = -0.020584$ ).

$\beta$  = slope of the regression curve ( $\beta_{\text{N\&S}} = 0.163957$ ;  $\beta_{\text{PUB}} = 0.165899$ ).

The functions were seasonally adjusted to provide a model that could be applied to several years. The term  $(1+x)/c$  represents the relationship between the total articles published (N&S or PUB) by an institution and the institution with largest number of publications (N&S or PUB). This averts any dependence on a specific year's values. The term  $(+1)$ , in turn, excludes the possibility of zero values in the denominator of the expression  $(\log c/(1+x))$ .

Table II shows the summary statistics of the coefficients values obtained after having applied the model to the years 2008-2015.

Figures 2 and 3 show the results: the actual scores published by ARWU 2015 (in blue) and the points estimated after fitting the non-linear function for each indicator (in red).

Whilst the same non-linear function was applied to both indicators to find the regression curve, the coefficients that afforded the best fit logically varied.

The choice of N&S and PUB was justified by the fact that the former was the indicator most closely correlated to the overall ranking (N&S,  $r = 0.935$ ), as

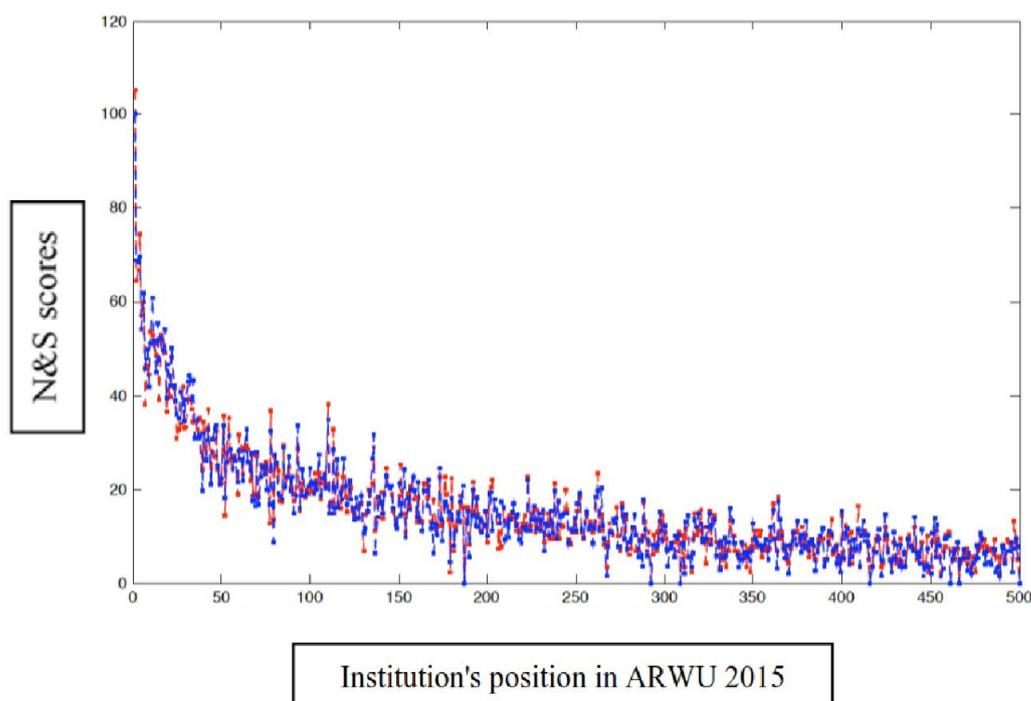
shown in Figure 4, and the latter the sole indicator affecting all universities (see Figure 5). Most of the institutions assessed showed no values for the other three indicators (Alumni, Award and HiCi) or the correlation between the respective scores and the final score was very low. Consequently, only the values for previous years were factored into the equation as a constant to improve the fit.

The goodness of fit for the model proposed was tested with the Kolmogorov-Smirnov procedure. The test findings revealed the similarity between the shapes of the two distributions (the estimated and the actual points), as shown in Table III and Figure 6. The calculations and graphics were performed with 'rgr' data exploration analysis software (Garret, 2015).

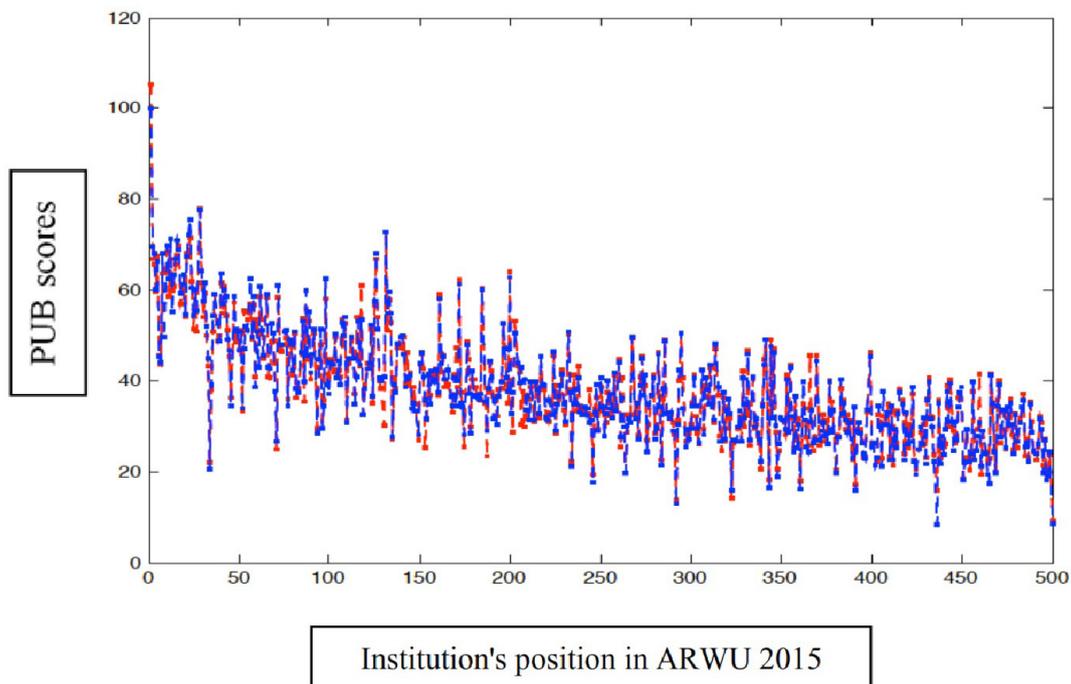
**Table II.** Summary statistics

Coefficients	Coefficients $\alpha$ N&S	Coefficients $\beta$ N&S	Coefficients $\alpha$ PUB	Coefficients $\beta$ PUB
Min	-0.008849	0.159232	-0.020584	0.163831
Max	-0.006314	0.170312	-0.018553	0.165899
Range	0.002535	0.011080	0.002031	0.002068
Mean	-0.007348	0.164230	-0.019465	0.165090
Standard deviation	0.000868	0.003901	0.000749	0.000748

**Figure 2.** Actual and non-linear function-estimated N&S scores



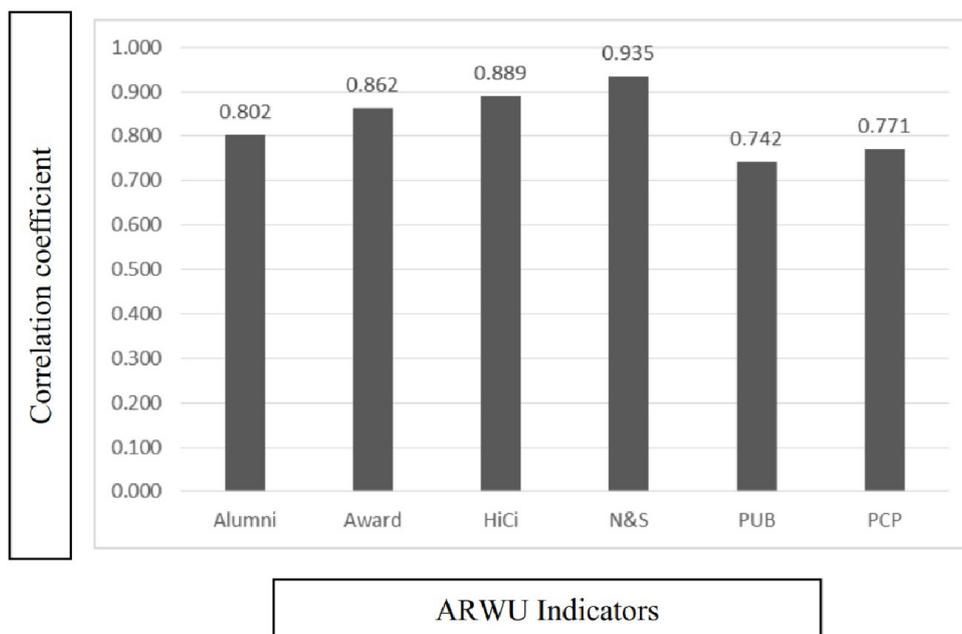
**Figure 3.** Actual and non-linear function-estimated PUB scores



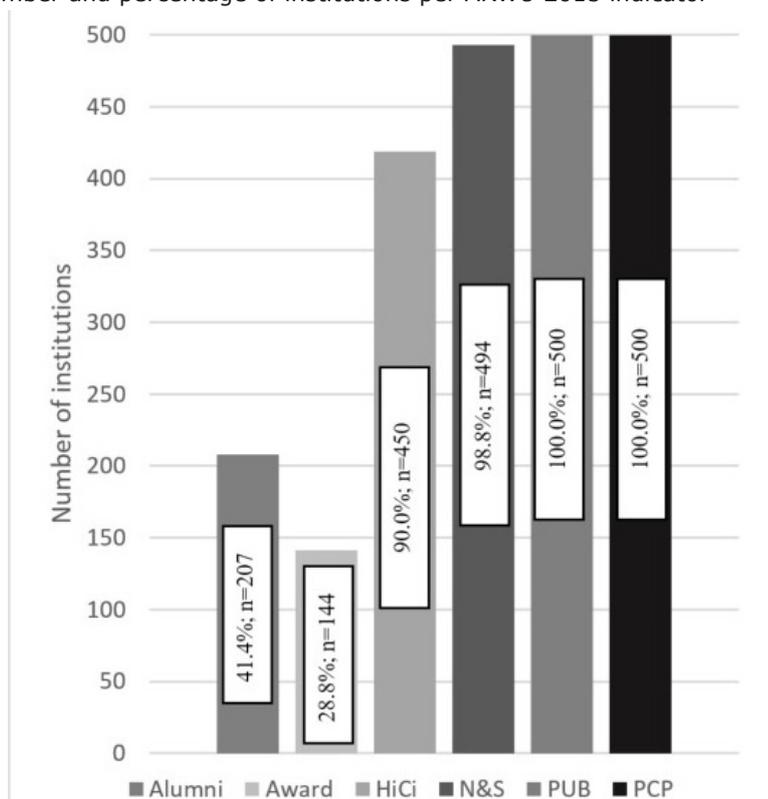
**Table III.** Kolmogorov-Smirnov test

Two-sample Kolmogorov-Smirnov test	N&S	PUB
D	0.053	0.033
p-value	0.498	0.957
Alpha	0.05	0.05

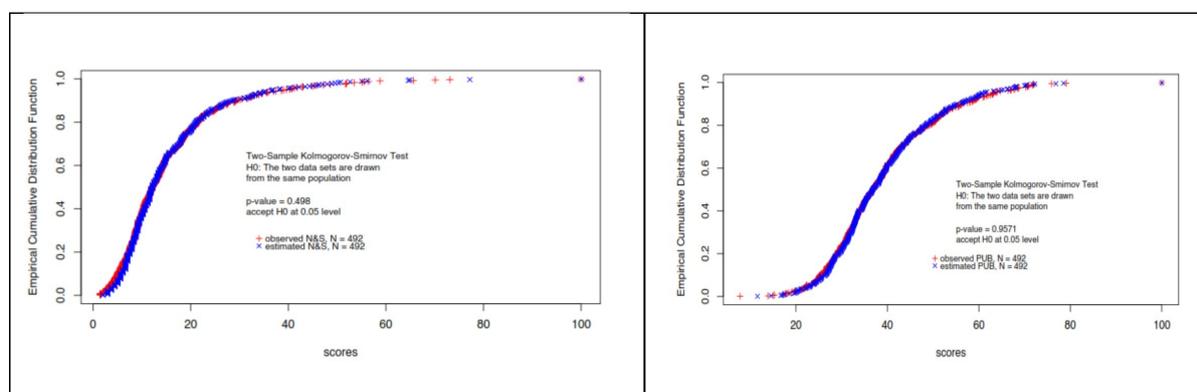
**Figure 4.** Correlation coefficient (r) between ARWU indicators and Shanghai ranking scores (2015). P-values: < 0.0001



**Figure 5.** Number and percentage of institutions per ARWU 2015 indicator



**Figure 6.** Kolmogorov-Smirnov test. N&S left and PUB right



## 5. RESULTS

### 5.1. Nature and Science (N&S) indicator

The determination coefficient ( $R^2$ ) was computed to determine the quality of the model used for the estimate. Table IV lists the determination coefficients for the number of N&S papers in 2010-14 compared to the N&S scores published in the ARWU 2015

and the scores estimated for the indicator with the methodology proposed. The model used here delivered a better fit than the gross values.

By way of example of the fit obtained with the proposed model, Table V shows the estimated N&S scores ( $E_{N\&S}$  ARWU-2015) for 50 universities chosen randomly from the top 500, selecting 10 from each range of 100 institutions.

**Table IV.** Determination coefficients ( $R^2$ ) for gross values, actual and estimated scores for the N&S indicator

Variable	N&S papers (WoS. 2010-14)	N&S score. ARWU-2015	Estimated N&S score. ARWU 2015
N&S papers (WoS. 2010-14)	1	0.835	0.877
N&S score. ARWU-2015	0.835	1	0.953
Estimated N&S score. ARWU-2015	0.877	0.953	1

**Table V.** Papers published in *Nature* and *Science* (N&S), actual and estimated scores

Institution	N&S papers (WoS. 2010-14)	N&S score ARWU-2015	E_N&S ARWU-2015 (2)	Diff. (2) - (1)
California Institute of Technology	319	56.40	56.23	-0.17
The University of Tokyo	212	48.60	44.36	-4.24
University of Minnesota. Twin Cities	125	32.50	32.69	0.19
University of California. Santa Barbara	86	27.50	26.36	-1.14
The University of Manchester	81	26.10	25.47	-0.63
The University of Melbourne	83	25.30	25.83	0.53
The University of Edinburgh	136	30.50	34.31	3.81
Vanderbilt University	72	21.80	23.81	2.01
Uppsala University	84	24.60	26.01	1.41
University of Helsinki	75	22.40	24.37	1.97
Peking University	59	21.50	21.25	-0.25
Indiana University Bloomington	66	21.90	22.65	0.75
University of Frankfurt	46	21.20	18.45	-2.75
University of Amsterdam	66	18.80	22.65	3.85
University of Liverpool	33	18.60	15.29	-3.31
Mayo Medical School	26	12.80	13.38	0.58
University of Leeds	46	18.40	18.45	0.05
Tel Aviv University	41	14.80	17.28	2.48
Queen Mary University of London	24	13.70	12.80	-0.90
University of Science and Technology of China	27	13.90	13.67	-0.23
University of Bologna	18	10.00	10.93	0.93
Delft University of Technology	29	18.10	14.23	-3.87
University of York	34	15.20	15.55	0.35
Yeshiva University	58	21.70	21.04	-0.66
Autonomous University of Madrid	15	9.90	9.89	-0.01
University of South Florida	14	10.40	9.53	-0.87
Charles University in Prague	16	9.00	10.25	1.25
Brandeis University	28	18.80	13.95	-4.85
Complutense University of Madrid	15	8.60	9.89	1.29

City University of Hong Kong	1	1.50	3.00	1.50
University at Albany (State University of New York)	11	8.70	8.38	-0.32
Umea University	26	10.40	13.38	2.98
James Cook University	32	15.40	15.03	-0.37
National Chiao Tung University	7	8.80	6.64	-2.16
Sichuan University	7	4.30	6.64	2.34
Eindhoven University of Technology	10	9.20	7.97	-1.23
University of Oulu	20	7.00	11.58	4.58
University of Duisburg-Essen	15	9.70	9.89	0.19
University of Wyoming	22	14.40	12.20	-2.20
Tianjin University	1	1.50	3.00	1.50
University College Cork	9	7.50	7.55	0.05
Saint Louis University	12	5.00	8.78	3.78
Syracuse University	7	8.00	6.64	-1.36
University of the Basque Country	19	7.10	11.25	4.15
University of Alaska - Fairbanks	13	11.70	9.16	-2.54
University of Jena	20	9.30	7.10	-2.20
University of Palermo	8	6.60	7.10	0.50
Utah State University	14	10.80	9.53	-1.27
University of Essex	1	1.50	3.00	1.50
University of Graz	6	6.20	6.15	-0.05

## 5.2. WoS listing indicator (PUB)

Table VI lists the determination coefficients for the number of papers listed in the WoS in 2014, the PUB scores obtained in the ARWU 2015 and the scores estimated for the indicator in this study (Estimated PUB score, ARWU 2015). The model used here delivered a better fit than the gross values for this indicator also.

Table VII, in turn, gives the estimated PUB scores for the same universities as listed in Table V. The first column shows the number of papers (WoS papers in 2014), the second lists the scores published for each institution in the ranking (PUB score, ARWU 2015) and the third the estimated scores (Estimated PUB score, ARWU 2015).

**Table VI.** Determination coefficients ( $R^2$ ) for gross values, actual and estimated scores for the PUB indicator

Variable	WoS papers in 2014 (PUB)	PUB score. ARWU 2015	Estimated PUB score. ARWU 2013
WoS papers in 2014 (PUB)	1	0.902	0.949
PUB score. ARWU 2015	0.902	1	0.952
Estimated PUB score. ARWU 2015	0.949	0.952	1

**Table VII.** Papers listed in WoS (PUB), actual and estimated scores

Institution	WoS papers in 2014 (PUB)	PUB score. ARWU-2015 (1)	Estimated PUB score. ARWU-2015 (2)	Diff. (2) - (1)
California Institute of Technology	3303	44.00	44.62	0.62
The University of Tokyo	7983	70.80	67.87	-2.93
University of Minnesota. Twin Cities	6474	64.00	61.34	-2.66
University of California. Santa Barbara	2257	37.30	37.45	0.15
The University of Manchester	5153	56.20	55.00	-1.20
The University of Melbourne	6302	66.90	60.55	-6.35
The University of Edinburgh	4555	51.90	51.88	-0.02
Vanderbilt University	4105	51.00	49.39	-1.61
Uppsala University	3919	49.80	48.33	-1.47
University of Helsinki	4330	51.80	50.65	-1.15
Peking University	6908	64.00	63.29	-0.71
Indiana University Bloomington	2345	39.60	38.11	-1.49
University of Frankfurt	2486	37.70	39.14	1.44
University of Amsterdam	4605	53.30	52.14	-1.16
University of Liverpool	2708	41.20	40.70	-0.50
Mayo Medical School	4351	46.50	50.77	4.27
University of Leeds	3027	42.50	42.85	0.35
Tel Aviv University	3784	48.10	47.54	-0.56
Queen Mary University of London	1937	32.10	34.95	2.85
University of Science and Technology of China	4086	50.80	49.29	-1.51
University of Bologna	3781	48.10	47.52	-0.58
Delft University of Technology	2472	38.90	39.04	0.14
University of York	1695	32.50	32.92	0.42
Yeshiva University	1753	31.50	33.42	1.92
Autonomous University of Madrid	2254	39.30	37.43	-1.87
University of South Florida	2410	40.20	38.59	-1.61
Charles University in Prague	3306	43.90	44.64	0.74
Brandeis University	499	18.40	19.64	1.24
Complutense University of Madrid	2897	42.80	41.99	-0.81
City University of Hong Kong	2233	38.50	37.27	-1.23
University at Albany (State University of New York)	948	25.60	25.56	-0.04
Umea University	1731	33.70	33.23	-0.47
James Cook University	1157	27.40	27.84	0.44
National Chiao Tung University	1633	32.10	32.38	0.28
Sichuan University	4701	54.20	52.66	-1.54
Eindhoven University of Technology	1523	31.10	31.40	0.30
University of Oulu	1103	28.50	27.27	-1.23
University of Duisburg-Essen	1765	30.60	33.52	2.92
University of Wyoming	828	23.30	24.14	0.84
Tianjin University	3085	46.80	43.23	-3.57
University College Cork	1131	25.30	27.57	2.27

Saint Louis University	950	24.20	25.58	1.38
Syracuse University	934	25.20	25.40	0.20
University of the Basque Country	2141	36.10	36.57	0.47
University of Alaska - Fairbanks	604	19.30	21.20	1.90
University of Jena	2093	34.00	33.21	-0.79
University of Palermo	1322	28.20	29.50	1.30
Utah State University	906	25.10	25.08	-0.02
University of Essex	580	20.70	20.86	0.16
University of Graz	742	19.90	23.06	3.16

## 6. DISCUSSION AND CONCLUSIONS

Since the Academic Ranking of World Universities (ARWU) was first published in 2003, the number of university rankings has multiplied each using different criteria and indicators to classify higher education institutions. The so-called league tables (ARWU, THE, QS) that have headed this trend are the ones most highly known among the public at large. Hand-in-hand with their growing popularity, considerable research has been conducted to analyse their conceptual grounds, techniques and methodologies. Reproducibility of their findings is one of the areas addressed.

Florian (2007), attempting to calculate universities' scores for the 2005 ARWU, unexpectedly found that the Shanghai 2005 findings could not be reproduced using the methodology outlined by the ranking. A plot of the paper counts and actual PUB scores obtained by listed institutions showed that the relationship between the two variables was not proportional, but rather non-linear. Inasmuch as the function is purported to fit the end scores to the raw paper count, the author reported that non-reproducibility of the results for the ranking could only be justified by the indication in the methodology to the effect that 'the data distribution for each indicator is reviewed for any significant distortion: standard statistical techniques are used to adjust the indicator as necessary' (ARWU, 2015).

Docampo (2013), in turn, addressed the difficulties encountered when attempting to reproduce the ARWU results. His paper included a procedure for estimating an institution's score for each of the ranking indicators with which he confirmed 'that the results of the Shanghai ranking are in fact reproducible'. The discrepancies found by the author between the scores estimated with his procedure and the actual scores published for each university in the ranking were justified by difficulties in assigning published papers to institutions (Docampo, 2013).

The aforementioned approaches used by Florian (2007) to estimate PUB scores and Docampo (2013) to calculate the scores for all the indicators in the ranking entail search (the search strategy applied to identify an institution's output) and paper assignment (fractionated counts and article weighting) processes that appear to be overly complex for non-experts.

The model introduced here is geared to replicating and predicting the ARWU results simply and with readily accessible data. It was successfully applied to estimate scores for the indicator that measures papers listed in WoS databases (PUB) as well as the indicator for the number of articles published in *Nature* and *Science* (N&S), as shown in the results described above.

The prediction model developed delivered a high correlation between the scores predicted and the scores published by the Shanghai ranking. The determination coefficients for the observed and estimated values were  $R^2=0.952$  for N&S in the 2015 edition of the ARWU and a similarly high  $R^2=0.953$  for PUB (tables IV and VI). The Kolmogorov-Smirnov test was run to verify whether the median, mean or shape of the distribution of the two populations (model-estimated and actual points shown in the ranking) differed. The results showed that the two distributions had the same shape for the indicator PUB, as well as for the indicator N&S (table III).

The model proposed supports the reproducibility of the Shanghai ranking for, while the findings could not be reproduced with absolute accuracy, they were estimated very reliably. It may, then, constitute a very useful tool for estimating an institution's PUB and N&S scores with a view to enhancing both its scientific strategy and its benchmarking skills. It may also be applied in the context of a country's research policy to ascertain the role of its universities from an international perspective, for the model can be used to estimate the scores for universities not ranked among the ARWU's top 500.

Lastly, the alternative method proposed in the paper contributes to the theoretical understanding of the methodology used to build rankings and the calculation mechanisms applied, as well as to interpreting the indicators. Practically speaking, the simplicity of the calculations involved makes the model accessible to any user (including policy

makers, students and other stakeholders) seeking to obtain significant information on an institution. It consequently favours the development of strategic plans and benchmarking policies and is particularly useful for assessing universities not listed in the Shanghai ranking.

## 7. REFERENCES

- Andersson, B.; Mayer, T. (2017). University Rankings: Was Academic Life Perfect before Rankings? In: Downing, K. and Ganotice, F. Jr. (eds.), *World University Rankings and the Future of Higher Education*, pp. 70-86. Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-5225-0819-9.ch004>
- ARWU. (2015). *Ranking Methodology of Academic Ranking of World Universities - 2015*. [Online] Available at: <http://www.shanghairanking.com/ARWU-Methodology-2015.html> [Accessed 1 Nov. 2015].
- Billaut, J. C.; Bouyssou, D.; Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, 84 (1), 237-263. <https://doi.org/10.1007/s11192-009-0115-x>
- 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. <https://doi.org/10.1002/asi.23396>
- 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. <https://doi.org/10.1002/asi.23527>
- Docampo, D. (2013). Reproducibility of the Shanghai academic ranking of world universities results. *Scientometrics*, 94 (2), 567-587. <https://doi.org/10.1007/s11192-012-0801-y>
- Docampo, D.; Cram, L. (2015). On the effects of institutional size in university classifications: the case of the Shanghai ranking. *Scientometrics*, 102 (2), 1325-1346. <https://doi.org/10.1007/s11192-014-1488-z>
- Docampo, D.; Egret, D.; Cram, L. (2015). The effect of university mergers on the Shanghai ranking. *Scientometrics*, 104 (1), 175-191. <https://doi.org/10.1007/s11192-015-1587-5>
- Fiala, D. (2012). Bibliometric analysis of CiteSeer data for countries. *Information Processing & Management*, 48 (2), 242-253. <https://doi.org/10.1016/j.ipm.2011.10.001>
- Florian, R. V. (2007). Irreproducibility of the results of the Shanghai academic ranking of world universities. *Scientometrics*, 72 (1), 25-32. <https://doi.org/10.1007/s11192-07-1712-1>
- Fowles, J.; Frederickson, H. G.; Koppell, J. G. S. (2016). University Rankings: Evidence and a Conceptual Framework. *Public Administration Review*, 76 (5), 790-803. <https://doi.org/10.1111/puar.12610>
- Garrett, R.R. (2015). *rgr: Applied Geochemistry EDA. R package version 1.1.11*. <http://CRAN.R-project.org/package=rgr>
- Gonzalez-Riano, M. G.; Repiso, R.; Lopez-Cozar, E. D. (2014). The impact of university rankings in the Spanish press. *Revista Española de Documentación Científica*, 37 (3): e055. <https://doi.org/10.3989/redc.2014.3.1128>
- Hazelkorn, E. (2014). Reflections on a Decade of Global Rankings: what we've learned and outstanding issues. *European Journal of Education*, 49 (1), 12-28. <https://doi.org/10.1111/ejed.12059>
- 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. <https://doi.org/10.1007/s11192-012-0674-0>
- Lascrain, M.L.; Pandiella-Dominique, A.; Sanz-Casado, E. (2015). The Weight of Databases in World University Rankings: Spanish Universities in ARWU. In: Luque Martínez, T. (ed.), *The University of Granada in light of its V Centenary. "Reflections on the Future of the University"*. Granada: Universidad de Granada.
- Liu, N. C.; Cheng, Y.; Liu, L. (2005). Academic ranking of world universities using scientometrics - A comment to the "Fatal Attraction". *Scientometrics*, 64 (1), 101-109. <https://doi.org/10.1007/s11192-005-0241-z>
- Marginson, S.; Van der Wende, M. (2007). Globalisation and Higher Education. *OECD Education Working Papers*, No. 8. OECD Publishing (NJ1). <https://doi.org/10.1787/173831738240>
- Moed, H. F. (2002). The impact-factors debate: the ISI's uses and limits. *Nature*, 415 (6873), 731-732. <https://doi.org/10.1038/415731a>
- Olcay, G. A.; Bulu, M. (2016). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change*, 123, 153-160. <https://doi.org/10.1016/j.techfore.2016.03.029>
- O'Leary, J. (2017). The Origins of International Rankings. In: Downing, K. and Ganotice, F. Jr. (eds.), *World University Rankings and the Future of Higher Education*, pp. 61-69. Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-5225-0819-9.ch003>

- Pusser, B.; Marginson, S. (2013). University Rankings in Critical Perspective. *The Journal of Higher Education* 84 (4), 544-568. <https://doi.org/10.1080/00221546.2013.11777301>
- Rauhvargers, A. (2011). *Global University Rankings and Their Impact* (vol. I). Brussels, Belgium: EUA.
- Rauhvargers, A. (2013). *Global University Rankings and Their Impact II* (vol. II). Brussels, Belgium: EUA.
- Robinson-Garcia, N.; Jiménez-Contreras, E. (2017). Analyzing the Disciplinary Focus of Universities: Can Rankings Be a One-Size-Fits-All? In: Downing, K. and Ganotice, F. Jr. (eds.), *World University Rankings and the Future of Higher Education*, pp. 161-185. Hershey, PA: IGI Global.
- Saisana, M.; d'Hombres, B.; Saltelli, A. (2011). Rickety numbers: Volatility of university rankings and policy implications. *Research Policy*, 40 (1), 165-177. <https://doi.org/10.1016/j.respol.2010.09.003>
- Sanz-Casado, E. (Ed.) (2015). *Guía de buenas prácticas para la participación de las universidades españolas en los rankings internacionales*. Madrid: Ministerio de Educación, Cultura y Deporte.
- Safón, V. (2013). What do global university rankings really measure? The search for the X factor and the X entity. *Scientometrics*, 97 (2), 223-244. <https://doi.org/10.1007/s11192-013-0986-8>
- Stewart, T. J. (2014). Value measurement theory and league tables. *South African Journal of Science*, 110 (3-4), 5-6.
- Turner, D. A. (2017). World Class Universities and the Rest. In: Downing, K. and Ganotice, F. Jr. (eds.), *World University Rankings and the Future of Higher Education* (pp. 109-128). Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-5225-0819-9.ch006>
- van Raan, A. F. J. (2005a). Academic ranking of world universities using scientometrics - A comment to the "Fatal Attraction" - Reply. *Scientometrics*, 64 (1), 111-112. <https://doi.org/10.1007/s11192-005-0242-y>
- van Raan, A. F. J. (2005b). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62 (1), 133-143. <https://doi.org/10.1007/s11192-005-0008-6>
- van Raan, A. F. J., van Leeuwen T.N.; Visser M.S. (2011). Severe language effect in university rankings: particularly Germany and France are wronged in citation-based rankings. *Scientometrics*, 88 (2): 495-498. <https://doi.org/10.1007/s11192-011-0382-1>
- 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. <https://doi.org/10.1002/asi.23207>