



HAL
open science

An anatomy of the academic ranking of world universities (Shanghai ranking)

D. Docampo, D. Egret, L. Cram

► **To cite this version:**

D. Docampo, D. Egret, L. Cram. An anatomy of the academic ranking of world universities (Shanghai ranking). *SN Social Sciences*, 2022, 2 (8), pp.146. 10.1007/s43545-022-00443-3 . hal-03744350

HAL Id: hal-03744350

<https://hal.science/hal-03744350>




Submitted on 2 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



An anatomy of the academic ranking of world universities (Shanghai ranking)

D. Docampo¹  · D. Egret²  · L. Cram^{3,4} 

Received: 16 November 2020 / Accepted: 28 June 2022
© The Author(s) 2022

Abstract

International academic rankings of research universities are widely applied and heavily criticised. Amongst the many international rankings, the Shanghai ranking has been particularly influential. Although this ranking's primary data are generally accessible and its methods are published in outline format, it does not follow that its outputs are predictable or straightforward. In practice, the annual and time series Shanghai rankings rely on data and rules that are complex, variable, and not fully revealed. Patterns and changes in the ranking may be misinterpreted as intrinsic properties of institutions or systems when they are actually beyond the influence of any university or nation. This article dissects the rules that connect raw institutional data to the published ranking, using the 2020 edition as a reference. Analysing an ARWU review of ranking changes over 2004–2016, we show how exogenous or methodological changes have often driven changes in ranking. Stakeholders can be misled if they believe that changes are intrinsic to institutions' performance. We hope to inform and warn the media, governments, and institutions about the merits and risks of using the Shanghai ranking to evaluate relative institutional performance and its evolution.

Keywords University ranking · World-class universities · Reproducibility · Academic ranking of world universities · Shanghai ranking

✉ L. Cram
Lawrence.Cram@anu.edu.au

¹ AtlanTTic Research Centre for Information and Communication Technologies, Universidad de Vigo, Campus Universitario, 36310 Vigo, Spain

² Observatoire de Paris, Université PSL, 92190 Meudon, France

³ Research School of Physics and Engineering, Australian National University, Acton, ACT 0200, Australia

⁴ College of Engineering, IT and Environment, Charles Darwin University, Darwin, NT 0815, Australia

Introduction

Around 2500 universities worldwide conduct research on a scale that supports institution-level comparative evaluation.¹ Stakeholders use these comparative evaluations to help inform judgements and choices about universities and university systems. One subclass of comparative evaluation—the global academic ranking of research universities—has been both an instrument for transformation and the target of substantial criticism (e.g. Marginson and Van der Wende 2007; Salmi 2009; Rauhvargers 2011; 2013; Hazelkorn 2017; Vernon et al. 2018).

The International Observatory on Academic Ranking and Excellence² (IREG) lists approximately two-dozen “global university rankings.” The earliest of these (Siwinski et al. 2021) and the subject of this paper is the Academic Ranking of Universities (ARWU, also known as the Shanghai Ranking, Shanghai Ranking or Shanghai Jiao Tong Ranking, Liu and Cheng 2005). Published annually since 2003, ARWU gathers its raw data from the public domain or paywalled sources and applies incompletely documented procedures to produce indicator scores and the overall ranking (ARWU 2021). From its inception, ARWU has attracted worldwide attention from media, governments, academia, alumni, prospective students, and other stakeholders, and it remains widely referenced both inside and outside academia.

We have searched for scholarly literature about ARWU (using Elsevier’s Scopus database for convenience), finding over 400 articles with titles, abstracts, or keywords similar to “ARWU” or “Shanghai Ranking” or which list amongst their references a citation to its initial description (Liu and Cheng 2005; Liu et al. 2005) or the early critique by Florian (2007). Guided by the use of Latent Dirichlet Allocation on article abstracts (Blei et al. 2003), co-citations (Small 1978) and bibliographic coupling (Yan and Ding 2012), we identify six to eight overlapping themes that fall into three clusters. The largest cluster of themes and articles relates to different aspects of the *application of ARWU* in studies of disciplines, regions, nations, institutions, researchers, and journals, or for administrative purposes such as planning, strategy, budgeting, faculty promotion or grant allocation. Examples include studies of university corporate governance based on the websites of public universities selected from the ARWU top 100 (Garde Sánchez et al. 2020) and how some institutions’ strategies have influenced and responded to ARWU over 15 years (Dowsett 2020). We have ourselves released applications as (unreviewed) articles that apply ARWU methods to regional university systems at ResearchGate.³

A second cluster discusses the social, economic, geographical, cultural and political *consequences of ARWU* and other university rankings. These works often cite Marginson and Van der Wende (2007) or Marginson (2014). Examples include a

¹ The University Ranking by Academic Performance (URAP) project aims for comprehensive coverage of academic (research-related) performance, publishing data on 2500 universities; Times Higher Education World University Rankings include 1500 universities; QS Top Universities and Shanghai ARWU list 1000; CWTS Leiden Ranking lists around 1200.

² <https://ireg-observatory.org/en/initiatives/fireg-inventory-of-international-rankings/> (accessed 2020/11/10).

³ <https://www.researchgate.net/project/Shanghai-Ranking-beyond-ARWU-500>.

monograph about rankings as engines of anxiety (Espeland et al. 2016), a report on elite-making higher education policies of China since the 1990s (Allen 2017) and observations about the place of ARWU in Chinese diplomacy and the projection of soft power (Lo 2011; Charroin, 2015). The third cluster of works relates to the *methodology of ARWU*, including discussions and criticisms of its reproducibility (Florian 2007; Docampo 2013), measurement error (Ioannidis et al. 2007), believability (Billaut et al. 2010), robustness (Saisana et al. 2011), consistency (Jeremic et al. 2011), validity (Selten et al. 2020), and discriminatory power (Claassen 2015) as well as the non-linear transformation of raw scores (Alaşehir et al. 2014; Docampo and Cram 2014). Most of these works cite or are cited by the reviews by Moed (2017) or Fernández-Cano et al. (2018).

In practice, the widespread recognition and influence of ARWU arise not from such academic studies but rather from its treatment by traditional and social media. Barats and her co-workers (Barats and Leblanc 2013; Barats 2017, 2019, 2020) provide an illuminating account of the importance of media actions for ARWU by analysing French media “occurrences” about ARWU taken from articles, dispatches, interviews and forums between 2003 and 2014, supplemented by interviews with reporters and media users. The analysis shows how French media amplified and consolidated the ARWU ranking due to (i) the simplicity, regularity, familiarity, independence and plasticity of ARWU’s ranking reports and (ii) the “électrochoc” of putative low rankings of French institutions. Barats notes how the media and other commentators often use an *argumentative concession* that acknowledges problems with the ARWU ranking whilst also using it without examination to report or advocate a position or proposition. One scholarly example is the extensive and critical review of rankings by Fernández-Cano et al. (2018), which summarises many shortcomings and potential improvements of university rankings, whilst observing (perhaps paradoxically) that due to its black-box approach and consideration of only a few indicators, the ARWU ranking “receives widespread critical acceptance”.

As noted by Fernández-Cano et al. (2018), “ARWU does not give any contextual consideration, only cold classifications”. Under this policy, it is unsurprising that the authors of ARWU have themselves used the *argumentative concession*. For example, Wu and Liu (2017) claim that “ARWU has been providing trustworthy performance information on universities in different countries for 14 years” yet caution in the same article that “people should pay more attention to the methodologies and indicators of rankings to understand and apply their results more accurately”. By dissecting ARWU’s methodology we show how the ranking of institutions described as improving or declining by Wu and Liu (2017) may change due to exogenous or methodological factors that are not inherent to the institution, although their article is indifferent to this distinction.

ARWU and other global university rankings now span over 15 years, and systematic time series studies have become feasible. However, whilst ARWU does not display the volatility of some other rankings (Holmes 2013; Hou and Jacob 2017; Sayed 2019), reliable time series of any ranking requires either stability or well-explained variations to the methodology. ARWU methodology does vary (cf. ARWU12 2012; ARWU20 2020), but the consequences are not well understood. In practice, stakeholders often accept a ranking as an indicator of relative institutional

performance and a ranking change as an indicator of improvement or decline, ignoring the warnings of the *argumentative concession*. Our paper aims to assist ARWU's users appreciate its limitations and extend its results through an independent assessment of the ARWU methodology.

The paper dissects the Shanghai ranking to reveal undocumented rules and choices that affect the results. Because ARWU results cannot be re-generated *ab initio*, we rely on the 2020 edition (hereinafter ARWU20) as the reference point; we cannot know whether further changes will occur. We first present the computational procedures that connect raw scores with the final ranking and then detail the steps from data sets to raw scores, including uncertainties and data cleaning activities. We then explain how ARWU outcomes may be due to drivers that are exogenous to both rankers and institutions, or methodological, or inherent to institutions. We conclude by discussing specific issues an institution may encounter when using ARWU to benchmark research performance.

ARWU data and methods

The Academic Ranking of World Universities (ARWU or Shanghai ranking) has been published annually since 2003, first by the Center for World-Class Universities (CWCU), Graduate School of Education (formerly the Institute of Higher Education) of Shanghai Jiao Tong University and after 2009 by the independent higher education intelligence organisation ShanghaiRanking Consultancy. Initially constructed as a public policy instrument to track the relative performance of Chinese universities as the Chinese university system grew and modernised (Liu and Cheng 2005; Liu et al. 2005; Liu 2009), the ARWU attracts worldwide attention from university stakeholders.

ARWU data cleaning

The passage from raw to cleaned ARWU data is subject to many alternatives and choices, many having been described in Liu and Cheng (2005), Van Raan (2005), and Liu et al. (2005). Liu et al. (2005) estimate that uncertainties due to attribution or human error are around 1–2% in the raw indicators. We emphasise here some additional considerations that may be undocumented or overlooked, particularly in the context of ARWU20.

Nobel Prizes and Fields Medals

Alumni is the count of graduates with a bachelor, master, or doctoral degree (or multiple degrees) from an institution who have won a Nobel Prize or a Fields Medal (ARWU20 includes awards in 2019 and earlier years). The indicator is discounted according to the year of graduation. In ARWU20, 100% of the score goes to recipients who graduated

after 2010, 90% to graduates between 2001 and 2010, and so on, down to 10% to graduates between 1921 and 1930.

Award is the fractional count of staff employed at an institution when they were awarded a Nobel Prize in Chemistry, Economic Sciences, Physiology, and Medicine or Physics or a Fields Medal in Mathematics (ARWU20 includes awards in 2019 and earlier years). The indicator is discounted according to the year of award: in ARWU20, 100% of the score goes to recipients of the prize after 2010, 90% to laureates between 2001 and 2010, and so on, down to 10% to laureates between 1921 and 1930.

Nobel Prize Laureate information is usually available at <https://nobelprize.org/> and Fields Medalist information is usually available at <http://www.mathunion.org/>. The 2020 Alumni and Award scores span awards made between 1920 and 2019. Table 1 sets out (for 2018, the most recent year with Fields Medal awards) the data cleaning applied to raw scores for university affiliations.

For Alumni, data cleaning includes identification of the contemporary ARWU name of alma maters and verification of the year of graduation so that the decadal ageing algorithm can be applied. Some institutions have been restructured over the century spanned by this indicator and Alumni points may be allocated to the relevant parts if they can be identified (e.g. University of Vienna and Medical University of Vienna) or descendent institutions may be excluded from ARWU (e.g. Berlin University). For Award data, cleaning includes identifying valid affiliation(s) at the time of the award, apportioning affiliations for shared and joint awards, and managing the year of an award to apply the decadal ageing algorithm. For convenience in treating shared awards, we have allocated three (3) points for a full Nobel Prize or a Fields Medal in Table 1.

Highly cited researchers

HiCi is the number of affiliated Highly Cited Researchers (HCR) in the HCR list released by Clarivate Analytics (ARWU20 uses the December 2019 release). Clarivate's methods for identifying HCRs and recording their affiliations have changed in material ways since the inception of ARWU rankings. Docampo and Cram (2019) trace the changes, including replacing lifetime awards with annual awards based on the previous decade of work, recalibration of the number of awards available in different research fields, and in the 2018 HCR list the introduction of a new cross-field category.

ARWU has also changed its HiCi methodology in material ways since 2003. For the HiCi indicator, previously complicated weightings now entail straightforward mapping of primary affiliated institutional names to ARWU names, noting that some HCRs are listed in more than one research category and counted multiple times.

Nature and Science

N&S is the weighted count of affiliations listed in articles published in the journals *Nature* and *Science* over the previous 5 years (for ARWU20, the years are 2015–2019). *Nature and Science* bibliographic data are publicly available on the

Table 1 Details of nobel prize and fields medal affiliations (universities only) for 2018 awards

Prize winner	Prize	Score (award)	Affiliation (award)	Year of graduation	Score (alumni)	Affiliation (alumni)
Frances H. Arnold	Chem	1.5	California Institute of Technology	1979	0.6	Princeton University
George P. Smith	Chem	0.75	University of Missouri	1985	0.7	University of California, Berkeley
Gregory P. Winter	Chem	0.375	University of Cambridge	1963	0.5	Haverford College
William D. Nordhaus	EcSci	1.5	Yale University	1970	0.5	Harvard University
Paul M. Romer	EcSci	1.5	New York University	1973	0.6	University of Cambridge
Olga Tokarczuk	Lit			1973	0.6	Yale University
Denis Mukwege	Peace			1967	0.5	Massachusetts Institute of Technology
Denis Mukwege	Peace			1983	0.7	University of Chicago
Arthur Ashkin	Phys			1985	0.7	University of Warsaw**
Gerard Mourou	Phys	0.375	University of Michigan	1983	0.7	University of Burundi
Gerard Mourou	Phys	0.375	Ecole Polytechnique	2015	1.0	Université Libre Bruxelles**
Donna Strickland	Phys	0.75	University of Waterloo	1947	0.3	Columbia University
James P. Allison	Med	0.75	University of Texas	1947	0.3	Columbia University
Tasaku Honjo	Med	1.5	Kyoto University	1952	0.4	Cornell University
Caucher Birkar	Fields	3	University of Cambridge	1967	0.5	Université Grenoble Alpes
Alessio Figalli	Fields	3	ETH Zurich	1973	0.6	Sorbonne Université
Peter Scholze	Fields	3	University of Bonn	1975	0.6	McMaster University
				1981	0.7	University of Rochester
				1989	0.7	University of Texas
				2000	0.8	Kyoto University
				2004	0.9	University of Tehran**
				2006	0.9	University of Nottingham
				2007	0.9	Scuola Normale Superiore—Pisa
				2007	0.9	University of Pisa
				2012	1.0	Ecole Normale Supérieure—Lyon
						University of Bonn

Table 1 (continued)

Prize winner	Prize	Score (award)	Affiliation (award)	Year of graduation	Score (alumni)	Affiliation (alumni)
Akshay Venkatesh	Fields	3	Stanford University**	1998	0.8	University of Western Australia**
				2002	0.9	Princeton University

**Denote affiliations that are not counted in ARWU20

journal websites or by subscription to Clarivate Analytics' *Web of Science* (WoS). ARWU N&S raw data are the sequence of authors and their affiliations in the categories of corresponding authors (CA) and other authors (OA) for each N&S article. Each publication yields a list of unique universities⁴ built up from all the affiliations of the first CA (and then any subsequent CAs taken in order) and then for all OAs in taken order. Each listed university's raw points are accumulated according to their position in each publication in this ordered list, reflecting author status and order.

Articles may have more than one corresponding author, and corresponding authors may have more than one university affiliation. There are also articles with dozens of authors and articles where the authors are listed under a consortium name rather than as individuals. As with the Publication indicator, institutional names must be disambiguated, and the validity of associated entities such as medical research institutions must be established.

Publications

PUB is the weighted count of affiliated articles indexed by Science Citation Index-Expanded (SCI-E) and Social Science Citation Index (SSCI) in the previous year (for ARWU20, the year is 2019). Bibliographic data for the *PUB* indicator can be accessed by subscription to WoS. An article is weighted more highly if indexed in SSCI, with a weight of 2 if indexed only in the SSCI and 1.5 if indexed in both the SSCI and the SCI-E.

Data cleaning for the *PUB* indicator focuses on identifying the valid institutional components and name variants selected by ARWU for each institution listed in ARWU20. The WoS organisation-enhanced (OG) field tag proves to be a useful starting point. However, some of the variants assembled into the OG concordance are components that ARWU may not include (e.g. associated hospitals). On the other hand, a WoS single organisation (OO) field tag will often miss name variants included in ARWU's unpublished list of name variants and institutional components.

PCP and FTE

ARWU intends that PCP is a per capita indicator derived by normalising the weighted total of the five (raw) indicators to the number of full-time equivalents (FTE) academic staff (Liu and Cheng 2005). Where ARWU identifies a reliable source, they take FTE values from national data. In other cases, they use a "weighted score" with a nominal or dummy value of FTE derived from the median value of the reliable FTE values in the corresponding year.

For countries where ARWU takes FTEs from national lists, the validity of FTE requires completeness, reliable concordance between institutions in the national list and ARWU's list, and amongst faculty levels, tenure status, and employment

⁴ ARWU does not allocate N&S points to non-university institutions.

contract classifications across and between different higher education systems. Anomalies are common. Senior faculties with untenured or grant-funded status are sometimes but not always counted (cf. ANU and Caltech). Distinctive institutions may sometimes be omitted from national lists, distorting their ranking and within-country comparisons (e.g. Mines ParisTech in France).

ARWU sometimes discovers or revises national FTE lists so that most institutions in an entire country undergo undocumented changes in PCP and total score. For example, ARWU included institutional FTE values for Polish institutions in ARWU20: FTE for Jagiellonian University and the University of Warsaw increased from the nominal value (950) in 2019 to 3375 and 2960, respectively, in 2020 (approximate FTE values). Jagiellonian moved from the 301–400 band to the 401–500 band, whilst Warsaw moved in the opposite direction. Warsaw's rise was due to ARWU's recognition of the delayed 2018 Nobel Prize of alumna Olga Tokarczuk slightly outweighing the significant decline in PCP.

Significant and sometimes irregular changes in FTE/PCP have occurred in Australia (2006, 2010), Austria (2010), Brazil (2010 concerning only the University of Sao Paulo), Canada (2009), Switzerland (2010), the USA (2010), China (irregular), the Czech Republic (2017), Spain (2010), France (2008, 2017), Great Britain (2005, 2010), Sweden (2010), Italy (2012), Japan (2011), Korea (2005), Norway (2010), New Zealand (2005, 2017), and the USA (2010). Germany, Israel, Finland, Denmark, Ireland, and South Africa are countries where ARWU uses the dummy value. In ARWU20, the dummy FTE value changed substantially, from around 950 to 1500, reflecting the difference in the median of the known FTE values. PCP decreased accordingly for institutions where ARWU uses the dummy FTE.

Computation of the indicator scores

The raw indicators Alumni, Award, HiCi, N&S, and PUB have disparate scales and distribution functions that are adjusted before combining in the total score. Raw indicator values are *normalised* by dividing by the corresponding raw value of the best performer (to date, always Harvard University) and *compressed* by taking the square root. Let H be the raw score of Harvard on any of the five indicators and X the raw score of another institution on that same indicator. The ARWU-published indicator score S is computed by Eq. (1) rounded to one significant digit:

$$S = 100 \sqrt{\frac{X}{H}}. \quad (1)$$

ARWU publishes S rather than raw scores X or H . Provided many values of X and S are known, the H value used in ARWU20 can be verified or estimated by fitting Eq. (1). Our estimations of Harvard's raw indicator scores for ARWU20 are exhibited in Table 2.

The PCP indicator is derived from the five fundamental indicators, adjusted to include the effect of institutional size on the amount of output. A convenient

Table 2 Raw indicator scores for Harvard in ARWU20

Indicator	Alumni	Award	HiCi	N&S	PUB
Harvard value	33.91	36.29	204	477.85	19,041

way to compute PCP is first to define WSS as the weighted sum of the squares of the five indicator scores (in effect, reversing Eq. 1), denoted by self-evident names:

$$\text{WSS} = 0.1 \text{ Alumni}^2 + 0.2 [\text{Award}^2 + \text{HiCi}^2 + \text{N\&S}^2 + \text{PUB}^2]. \quad (2)$$

If WSS_C is the weighted sum of squares for Caltech (the university with the highest score in the indicator PCP) and FTE_C is the FTE faculty of Caltech, the PCP indicator of any university N with a weighted sum of squares WSS_N and FTE faculty FTE_N is then

$$\text{PCP} = 100 \sqrt{\frac{\text{WSS}_N \text{ FTE}_C}{\text{FTE}_N \text{ WSS}_C}}. \quad (3)$$

By fitting raw scores, we estimate that in ARWU20, $\text{FTE}_C = 294.5$ and $\text{WSS}_C = 2617.9$ so that

$$\sqrt{\frac{\text{FTE}_C}{\text{WSS}_C}} = 0.3409 \quad \text{and} \quad \text{PCP} = 34.09 \sqrt{\frac{\text{WSS}_X}{\text{FTE}_X}}. \quad (4)$$

ARWU ranks universities individually or into bands based on a total score S_T computed as the weighted sum of individual indicator scores computed and assembled as described below:

$$S_T = K(0.1[\text{Alumni} + \text{PCP}] + 0.2[\text{Award} + \text{HiCi} + \text{N\&S} + \text{PUB}]). \quad (5)$$

The factor K sets $S_T = 100$ for the top performer (Harvard) since the value of PCP_H is not 100:

$$K = [0.9 + 0.1\text{PCP}_H/100]^{-1}. \quad (6)$$

ARWU modifies (1) and (5) for a small number of institutions whose social science and humanities mission means that *Nature* or *Science* are used only infrequently by their researchers. The N&S indicator is null for these institutions, with other indicators appropriately re-weighted. For ARWU20, the institutions are London School of Economics and Political Science, Université Toulouse 1 Capitole, Stockholm School of Economics, Bocconi University, Tilburg University, Renmin University of China, Copenhagen Business School, London Business School, School for Advanced Studies in the Social Sciences, and the WU Vienna University of Economics and Business.

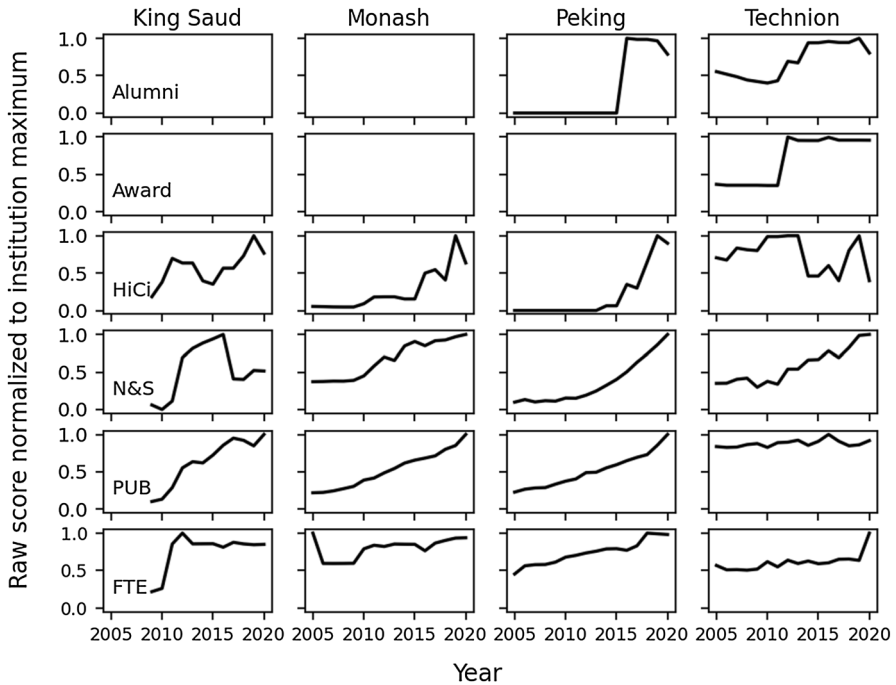


Fig. 1 Time series of raw scores scaled to Harvard or Caltech for institutions selected for their rapid improvement in the ranking. Three enter the top 100 and one (King Saud) enters the ranking. The range is normalised to the maximum value in each panel

Application

In 2017, two ARWU authors published a review of trends in ARWU between 2004 and 2016 (Wu and Liu 2017). The review mentioned twenty-seven institutions showing “great improvements” and around forty institutions that exhibited “major declines” in ARWU ranking. We have applied the procedures described above to link ARWU’s reported ranking movements (excluding 2004 due to an abundance of irregularities compared with later years and updating the series to include 2017–2020) to the underlying changes in raw scores of four institutions in each category. In the “improvement” category, we selected the earliest Saudi university that entered the top 500 (King Saud) and three of the four universities that entered the top 100 (Peking, Monash, and Technion, omitting the second Chinese institution Tsinghua). In the “decline” category, we select the first institutions that moved out of the top 100 in four different countries (Rochester, Tohoku, Vienna, and Sheffield).

Figures 1 and 2 display the time series of raw scores for the chosen institutions, scaling raw scores to the reference institution (Harvard or Caltech) at each time step. For each indicator, the time series plot is normalised to the maximum value of the scaled raw score over the period 2005–2020. The time series thus reveals changes relative to the changes of the reference institution.

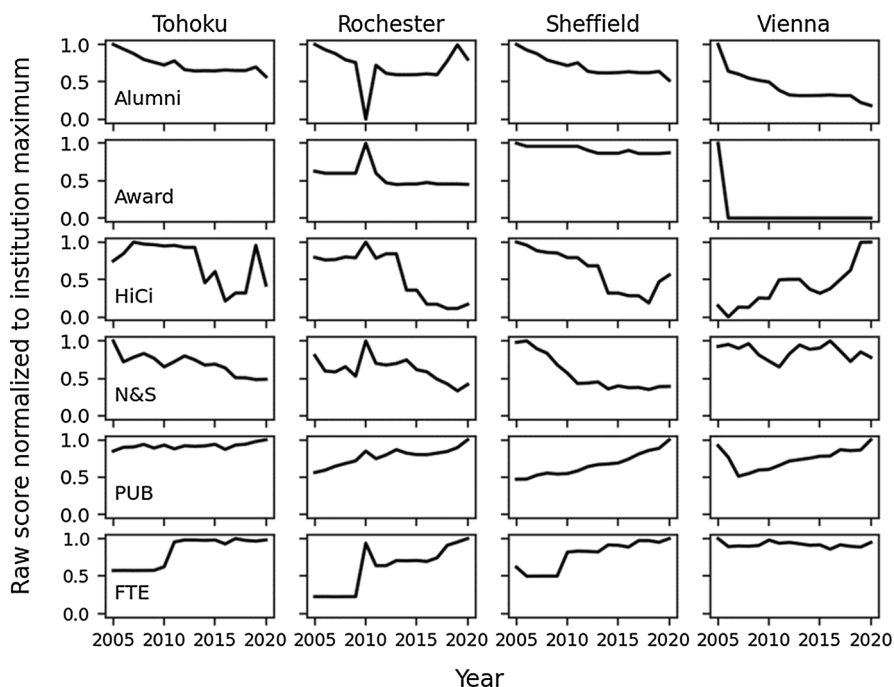


Fig. 2 Time series of raw scores scaled to Harvard or Caltech for institutions selected for their rapid decline out of the top 100. The range is normalised to the maximum value in each panel

Figures 1 and 2 exhibit considerable temporal variability. Table 3 describes the drivers of the variability using three categories guided by the “precautionary principles” (Cheng and Liu 2008) of the Berlin Principles on Ranking of Higher Education Institutions (IREG 2021). “Exogenous” change arises when factors other than the routine application of existing ARWU ranking rules are involved in a change. “Methodological” changes arise when ARWU adjusts its existing data sets or ranking rules. “Inherent” changes occur when changes in an institution’s raw score are not due to either of the other two categories. The categories are not mutually exclusive and are often entangled. Note also that an institution’s ranking (but not indicator scores) is exogenously responsive to the insertion or removal of other institutions lying above it between one annual ranking and the next.

Discussion and conclusion

By successfully reproducing ARWU20, we confirm that ARWU relies on non-subjective raw data (cf. Vernon et al. 2018). It is, however, not possible to reproduce ARWU results *ab initio*. There are ambiguities and undocumented choices made in producing indicator scores that must be understood and resolved by post-facto iteration to match published scores. It is a significant challenge to discover, clean, and manipulate the entire data set to arrive at the published ranking. It is an even

Table 3 Annotation of changes exhibited in Figs. 1 and 2

	Exogenous	Methodological	Inherent
All institutions (HiCi)	Clarivate changed the definition of HCR in 2014/2015 Clarivate introduced the cross-field category in 2017/2018	ARWU changed rules for HCR with multiple affiliations (2014/2015/2016/2017)	
All institutions (FTE)	National FTE data omit certain institutions	ARWU adopts a new national FTE data set ARWU extends the list of institutions, revising the dummy FTE value	
Declining institutions (prizes)		Decadal ageing of Alumni and Award raw scores	Harvard gains new Alumni and Awards
King Saud		Uplift of N&S articles in 2010/11 passes through a 5-year window	High growth rate (Al-Khedhairi 2012)
Monash			High growth rate
Peking			High growth rate One Alumni (1955/2015)
Technion			Three Alumni (1966/2013; 1972/2011; 1981/2004) Two-part Awards (2004) and one full Award (2011)
Tohoku			N&S and PUB decline with steady FTE
Rochester		Anomalous results in 2011	N&S and PUB decline, whilst FTE increases
Sheffield			N&S decline, whilst FTE increase
Vienna	Medical school separated in the 1940s	Revision to initial allocation of Alumni, Award, PUB, and FTE	

more complex challenge to infer what a presently un-ranked institution might need to do to enter the ranking because this requires forecasting the scores and ranks of an unknown number of other institutions. Our paper assists the process by exhibiting the steps that lead to final scores and ranks. Siniksaran and Satman (2020) published simulation software for performing ranking studies, which could be valuable once the connections between published ARWU data and the underlying raw scores are addressed.

Each ARWU indicator poses significant hurdles for users. Although they affect only a small number of institutions, the affiliations arising from new Award and Alumni announcements may take a few years to be resolved (cf. Table 1). Recognising the contrast between the interests of older and rising universities concerning the long-term anchoring effect of the Alumni and Award indicators, in 2014 and 2015 ARWU published an alternative ranking⁵ that omitted the Award and Alumni indicators. The above methods will allow institutions to produce alternative versions of the ARWU ranking that may be more appropriate to their mission and strategy.

The category HCR that underlies the HiCi indicator has been adjusted several times since the inception of ARWU (Docampo and Cram 2019). Moreover, HiCi is the sole indicator that continues to be subject to year-to-year variations due to the inward or outward redirection of affiliations by researchers. The redirection of an HCR affiliation is particularly potent when there are only a few HCRs in an institution, potentially leading to a change of more than 100 places in the rank order.

The N&S and PUB indicators both raise issues of ambiguity in affiliation names. Institutions can reduce the leakage of affiliations by monitoring the format of submitted articles and, perhaps, avoiding journals and group publication practices that lead to the omission of affiliation information in the bibliometric indexes. The N&S indicator is further subject to considerable complexity owing to the ARWU rules for weighting affiliations according to authorship order. Whether or not an institution attracts the raw score premium for corresponding, first or second author depends on quite complex rules that might not be evident to researchers when they agree on the layout of the article affiliation by-line.

ARWU developed the PCP indicator (named “size” in early editions of ARWU) to respond to commentary about the trend for other indicators to be automatically larger in larger institutions, other things being equal. If this is a concern, a substantial additional weighting for PCP could be contemplated (Liu and Cheng 2005). The price of including PCP is the difficulty of finding national data on FTE and harmonising how reported FTE are related to academic employment conditions across national systems. Size-independent research evaluation metrics are available in other rankings (e.g. the CWTS Leiden Ranking: Waltman et al. 2011, 2012) and can be compared with the size-dependent ARWU total score and rank and the size-adjusted ARWU PCP score and rank.

Our study of ARWU’s report on examples of long-term “improvement” and “decline” (Wu and Liu 2017) reveals complex developments that entail exogenous,

⁵ http://www.shanghairanking.com/Alternative_Ranking_Excluding_Award_Factor/Excluding_Award_Factor2015.html.

methodological, and inherent changes, perhaps entangled. Wu and Liu's report is *indifferent* to explaining the underlying drivers of the actual ranking outcomes (see also Holmes 2013; Lim 2018 regarding indifference of other global rankers), a position that might be inconsistent with some of the Berlin Principles (Cheng and Liu 2008). Perhaps an *annotation schema* building on Table 3 could be adopted by ARWU as a contribution to the improvement and evaluation of its ranking, as envisaged by the Berlin Principles.

Many of the complications described in this paper are challenges typical of bibliometric practice and occur alongside mistakes, errors, and uncertainties in the data that underpins the ARWU (or any other) ranking. For this reason, as well as concerns about the usefulness of quantitative, low-dimensional evaluation (e.g. Moed 2017), like many others, we strongly encourage readers to (1) understand the changes reported by ARWU in terms of exogenous, methodological, and inherent drivers and (2) explore and use institutional evaluation methods beyond the ARWU.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Not applicable.

Data availability The article includes URLs and descriptions of proprietary data.

Code availability Not applicable.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alaşehir O, Çakır MP, Acartürk C, Baykal N, Akbulut U (2014) URAP-TR: a national ranking for Turkish universities based on academic performance. *Scientometrics* 101(1):159–178. <https://doi.org/10.1007/s11192-014-1333-4>
- Al-Khedhairy AA (2012) Saudi University policy: King Saud response. *Science* 335(6072):1040–1040. <https://doi.org/10.1126/science.335.6072.1040-b>
- Allen RM (2017) A comparison of China's "Ivy League" to other peer groupings through global university rankings. *J Stud Int Educ* 21(5):395–411. <https://doi.org/10.1177/1028315317697539>
- ARWU12 (2012) Methodology for the computation of the HiCi indicator. <http://www.shanghairanking.com/ARWU-Methodology-2012.html>. Accessed 19 Oct 2015
- ARWU20 (2021) Academic ranking of world universities—methodology. <http://www.shanghairanking.com/ARWU-Methodology-2020.html>. Accessed 27 Apr 2021

- Barats C (2017) Diffusion d'un classement académique en France: analyse des logiques sociales et des discours de presse sur le classement dit «de Shanghai». Habilitation à Diriger des Recherches en Sciences de l'information et de la communication, soutenue le 6 juin 2017 à l'université Paris Est Créteil. *Education et Sociétés Plurilingues* 43:97–106. <https://doi.org/10.4000/esp.1729>
- Barats C (2019) À propos de la médiatisation du classement de Shanghai: l'apport de l'analyse du discours. *Revue Française de Linguistique Appliquée* 24(1):21–35. <https://doi.org/10.3917/rfla.241.0021>
- Barats C (2020) Dissemination of international rankings: characteristics of the media coverage of the Shanghai Ranking in the French press. *Palgrave Commun* 6(1):1–11. <https://doi.org/10.1057/s41599-020-0440-5>
- Barats C, Leblanc J-M (2013) Généalogie de la co-construction médiatique du « classement de Shanghai » en France. *Corpus de presse et usages sociodiscursifs. Mots. Les langages du politique*, N 102/2013. <https://doi.org/10.4000/mots.21313>
- Billaud JC, Bouyssou D, Vincke P (2010) Should you believe in the Shanghai ranking: an MCDM view. *Scientometrics* 84(1):237–263. <https://doi.org/10.1007/s11192-009-0115-x>
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3(1):993–1022
- Charroin J (2015) The Shanghai ranking: a diplomatic leverage tool? *Revue Internationale et Stratégique* 97(1):48–60
- Cheng Y, Liu NC (2008) Examining major rankings according to the Berlin principles. *High Educ Eur* 33(2–3):201–208. <https://doi.org/10.1080/03797720802253686>
- Classen C (2015) Measuring university quality. *Scientometrics* 104(3):793–807. <https://doi.org/10.1007/s11192-015-1584-8>
- 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 (2014) On the internal dynamics of the Shanghai ranking. *Scientometrics* 98(2):1347–1366. <https://doi.org/10.1007/s11192-013-1143-0>
- Docampo D, Cram L (2019) Highly cited researchers: a moving target. *Scientometrics* 118(3):1011–1025. <https://doi.org/10.1007/s11192-018-2993-2>
- Dowsett L (2020) Global university rankings and strategic planning: a case study of Australian institutional performance. *J High Educ Policy Manag.* <https://doi.org/10.1080/1360080X.2019.1701853>
- Espeland WN, Sauder M, Espeland W (2016) *Engines of anxiety: academic rankings, reputation, and accountability.* Russell Sage Foundation, New York
- Fernández-Cano A, Curiel-Marin E, Torralbo-Rodríguez M, Vallejo-Ruiz M (2018) Questioning the Shanghai ranking methodology as a tool for the evaluation of universities: an integrative review. *Scientometrics* 116(3):2069–2083. <https://doi.org/10.1007/s11192-018-2814-7>
- Florian RV (2007) Irreproducibility of the results of the Shanghai academic ranking of world universities. *Scientometrics* 72(1):25–32. <https://doi.org/10.1007/s11192-007-1712-1>
- Garde Sánchez R, Flórez-Parra JM, López-Pérez MV, López-Hernández AM (2020) Corporate governance and disclosure of information on corporate social responsibility: an analysis of the top 200 universities in the Shanghai ranking. *Sustainability* 12(4):1549. <https://doi.org/10.3390/su12041549>
- Hazelkorn E (2017) *Rankings and higher education: reframing relationships within and between states.* Centre for Global Higher Education, working paper no. 19. <https://www.researchcghc.org/perch/resources/publications/wp19.pdf>
- Holmes R (2013) *Watching the rankings.* Amazon Digital Editions.
- Hou Y-W, Jacob WJ (2017) What contributes more to the ranking of higher education institutions? A comparison of three world university rankings. *Int Educ J Comp Perspect* 16(4):29–46
- Ioannidis JP, Patsopoulos NA, Kavvoura FK, Tatsioni A, Evangelou E, Kouri I et al (2007) International ranking systems for universities and institutions: a critical appraisal. *BMC Med* 5(1):30. <https://doi.org/10.1186/1741-7015-5-30>
- IREG (2021) Berlin principles on ranking of higher education institutions. http://ireg-observatory.org/en_old/berlin-principles. Accessed 19 Apr 2021
- Jeremic V, Bulajic M, Martic M, Radojicic Z (2011) A fresh approach to evaluating the academic ranking of world universities. *Scientometrics* 87(3):587–596. <https://doi.org/10.1007/s11192-011-0361-6>
- Lim MA (2018) The building of weak expertise: the work of global university rankers. *High Educ* 75(3):415–430. <https://doi.org/10.1007/s10734-017-0147-8>
- Liu NC (2009) The story of academic ranking of world universities. *Int High Educ.* <https://doi.org/10.6017/ihe.2009.54.8409>

- Liu NC, Cheng Y (2005) Academic ranking of world universities: methodologies and problems. *High Educ Eur* 30(2):127–136. <https://doi.org/10.1080/03797720500260116>
- Liu NC, 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>
- Lo WYW (2011) Soft power, university rankings and knowledge production: distinctions between hegemony and self-determination in higher education. *Comp Educ* 47(2):209–222. <https://doi.org/10.1080/03050068.2011.554092>
- Marginson S (2014) University rankings and social science. *Eur J Educ* 49(1):45–59. <https://doi.org/10.1111/ejed.12061>
- Marginson S, Van der Wende M (2007) To rank or to be ranked: the impact of global rankings in higher education. *J Stud Int Educ* 11(3–4):306–329. <https://doi.org/10.1177/1028315307303544>
- Moed HF (2017) A critical comparative analysis of five world university rankings. *Scientometrics* 110(2):967–990. <https://doi.org/10.1007/s11192-016-2212-y>
- Rauhvargers A (2011) EUA report on global rankings and their impact—report I. European University Association, Brussels
- Rauhvargers A (2013) Global university rankings and their impact: report II. European University Association, Brussels
- Saisana M, d’Hombres B, Saltelli A (2011) Rickety numbers: volatility of university rankings and policy implications. *Res Policy* 40(1):165–177. <https://doi.org/10.1016/j.respol.2010.09.003>
- Salmi J (2009) The challenge of establishing world-class universities. World Bank Publications, Washington, DC
- Sayed OH (2019) Critical treatise on university ranking systems. *Open J Soc Sci* 07(12):39–51. <https://doi.org/10.4236/jss.2019.712004>
- Selten F, Neylon C, Huang C-K, Groth P (2020) A longitudinal analysis of university rankings. *Quant Sci Stud* 1(3):1109–1135. https://doi.org/10.1162/qss_a_00052
- Siniksaran E, Satman MH (2020) WURS: a simulation software for university rankings—software review. *Scientometrics* 122(1):701–717. <https://doi.org/10.1007/s11192-019-03269-8>
- Siwinski W, Holmes R, Kopanska J (eds) (2021) IREG inventory of international university rankings 2021. Perspektywy Education Foundation, Warsaw
- Small HG (1978) Cited documents as concept symbols. *Soc Stud Sci* 8(3):327–340. <https://doi.org/10.1177/030631277800800305>
- Van Raan AFJ (2005) 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>
- Vernon MM, Balas EA, Momani S (2018) Are university rankings useful to improve research? A systematic review. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0193762>
- Waltman L, van Eck NJ, van Leeuwen TN, Visser MS, van Raan AFJ (2011) Towards a new crown indicator: some theoretical considerations. *J Informetr* 5(1):37–47. <https://doi.org/10.1016/j.joi.2010.08.001>
- Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, van Eck NJ et al (2012) The Leiden ranking 2011/2012: data collection, indicators, and interpretation. *J Am Soc Inf Sci Technol* 63(12):2419–2432. <https://doi.org/10.1002/asi.22708>
- Wu Y, Liu NC (2017) Academic ranking of world universities (ARWU): methodologies and trends. In: Cantu-Ortiz FJ (ed) *Research analytics: boosting university productivity and competitiveness through scientometrics*. CRC Press, Boca Raton, p 95
- Yan E, Ding Y (2012) Scholarly network similarities: how bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and crowd networks relate to each other. *J Am Soc Inform Sci Technol* 63(7):1313–1326. <https://doi.org/10.1002/asi.22680>