Check for
updates

Inter-ranking reputational effects: an analysis of the Academic Ranking of World Universities (ARWU) and the Times Higher Education World University Rankings (THE) reputational relationship

Vicente Safón[1]

Abstract

Global university rankings have become a critical factor in the higher education sector, engendering increasing interest and exerting a notable influence over a wide variety of stakeholders. They are presented to different audiences as tools that evaluate and rank universities according to quality. However, some authors are of the opinion that rankings express reputational factors to a large extent. This article presents a model of the intra- and inter-ranking relationships from the perspective of reputation along with an empirical study on two of the most influential rankings: the Academic Ranking of World Universities and the Times Higher Education World University Rankings. Data from these two rankings between 2010 and 2018, and the application of ordinal regressions, provide evidence that both rankings are mutually influential, generating intra and reciprocal reputational effects over time.

Introduction

In a little under 15 years, global university rankings have become a critical factor in the higher education sector, engendering increasing interest and exerting a notable influence over a wide variety of stakeholders (Hazelkorn 2016; Van Raan 2005). They are presented to different audiences as tools that evaluate and rank universities according to quality. However, despite the diversity of methods used to create them, some authors are of the

✉ Vicente Safón
  safon@uv.es

[1] Department of Management, Faculty of Economics, University of Valencia, Avenida de los Naranjos, s/n, 46022 Valencia, Spain

opinion that rankings express reputational factors to a large extent, particularly for institutions at the top of the lists (Bastedo and Bowman 2009; Locke et al. 2008; Safón 2013; Shehatta and Mahmood 2016; Shin 2011; Stake 2006; Stella and Woodhouse 2006).

A university's reputation, i.e. the overall and final addition to the perceptions of all stakeholders over time (Fombrun 1996), is the origin, and also the consequence, of academic institution rankings. The dominant perspective of academic reputations (Safón 2012), grounded in signaling theory (Spence 1973), is based on the belief that reputation is an outcome of the achievements of an academic institution (Keith 2001), which act as signals of organizational reputation (Rindova et al. 2005). It is assumed that research published by institutions represents the foundation of their prestige (Armstrong and Sperry 1994; Cole and Cole 1967; Smyth 1999). This prestige was reflected in the first ranking that appeared in 1925 in the US. This ranking and those that have followed have acted as signals of reputation, creating or modifying the perceptions of individuals (Rindova et al. 2005; Safón 2007, 2009; Sweitzer and Volkwein 2009); perceptions that have subsequently been incorporated into rankings that use reputational surveys, generating a powerful time-based interaction between rankings and perceptions (Bastedo and Bowman 2009; Bowman and Bastedo 2011; Safón 2009; Stake 2006).

Various global rankings do not use opinions as part of their methodology, and hence are unable to activate the 'ranking-perceptions-ranking' mechanism. However, the elements that make them up, normally research performance indicators, are also subtly affected by reputation signals derived from rankings, and are thus influenced by them.

Previous literature has extensively analyzed global rankings and the relations between them (e.g. Aguillo et al. 2010; Çakır et al. 2015; Moed 2017; Olcay and Bulu 2017; Rauhvargers 2011; Shehatta and Mahmood 2016; Waltman et al. 2012). Until now, no study has analyzed the reciprocal influence between rankings of different types in terms of the creation of reputational signals; nor have they produced a model of the process. Reputational signals are essential to understanding the impact of influential rankings, their limitations and ways they can potentially be improved, and also for universities themselves, as reputation is one of their critical assets.

This study presents a model of the inter-ranking relationships from the perspective of reputation along with an empirical study on two of the most influential rankings: the Academic Ranking of World Universities (ARWU) and the Times Higher Education World University Rankings (THE). Data from these two rankings between 2010 and 2018, and the application of ordinal regressions, provide evidence that both rankings are mutually influential, generating reciprocal reputational effects over time.

Reputational signals and reputational biases in global rankings

Several authors posit that rankings are unable to act as effective tools for assessing the quality of academic institutions, stating that what they really do is measure reputation consolidated by universities over time (Shin 2011; Stella and Woodhouse 2006). Toutkoushian and Webber (2011) and Shin (2011) believe the problem stems from the peer assessments used in many rankings, which do not give any indication of whether what is being measured is actually productivity or reputation. Some global rankings, including those of the USNWR, THE and QS rankings are constructed using these types of data (surveys), and thus suffer from reputational bias.

On the other hand, the literature on this topic has revealed that rankings affect the assessments of institutional reputation made by faculty and institutional leaders via an anchoring effect (Bastedo and Bowman 2009, 2011). With the passage of time, rankings that are totally or partially based on academic, recruiter or student opinions become unconnected with the changes in quality through the action of an echo effect (Leiter 2006; Stake 2006), which basically reflects changes in the prestige of institutions (Bowman and Bastedo 2011).

According to anchoring theory, ''people use the starting value to inform their judgments, and then they adjust (insufficiently) this value when making their final judgment, even when the starting value is entirely random'' (Bowman and Bastedo 2011, p. 433). The first university rankings may have established anchoring for certain universities, whose effects have lasted over time and are still relevant today. Brooks (2005) comments that the majority of US universities in the top 25 in 1925 (year of the first ranking) have remained in that tier for most subsequent rankings. This ranking, and the subsequent ones, especially the 1983 *US News "America's Best Colleges"* ranking, created anchoring effects that modified the perceptions of the people whose opinions were used in successive rankings. Today, every time a new opinion-based ranking is generated, the anchoring effects present in the previous rankings have a potential degree of influence. This effect has been demonstrated in the THE-QS ranking by Bowman and Bastedo (2011).

This effect has also been observed in business schools with a longer history. These institutions have had the chance to enjoy a sizeable advantage, with regard to first mover advantages and reputation, based on their inclusion in the first selection of b-schools that appeared in business week (BW) in 1986. As stated by Morgeson and Nahrgang (2008), this list was made up of b-schools that enjoyed the highest reputation at that moment in time, most likely explained by the age of the institution. Those 20 b-schools formed the origins of the first BW ranking that appeared 2 years later, and have almost invariably appeared in subsequent bi-annual BW rankings.

The echo effect or echo-chamber effect (Leiter 2006; Stake 2006) is an effect that explains how rankings, due to their capacity to synthesize and simplify, and their powerful communication channels via the media (Dyck and Zingales 2002), influence the opinions that are subsequently used to generate new rankings. This effect appears when alternative viewpoints, opinions and arguments are significantly absent from any community, particularly a university community (Kohm and Wardle 2011). It is then that the assessment of the reputation of institutions can be confused with an echo effect, where initial reactions can be reinforced and unendingly spread by the media (Rondeau 2017).

Anchoring effects have played a crucial role in forming perceptions based on the first rankings, but also influence current perceptions, as each new ranking anchors the perceptions of observers for each university. When this is seen as a dynamic process and echo effects are added, a "reputation-ranking-reputation" circle appears.

This *circle* would have a more solid foundation for rankings based on opinions, such as the THE ranking, rather than for those based on research performance objective data. However, as reputation influences research performance through a variety of biases, it would also influence the rankings not based on opinions, such as the ARWU.

The Matthew effect and the halo effect justify reputational bias that rankings based on research performance might undergo. The halo effect is a cognitive bias in which the assessment of one characteristic impacts the judgment of others. According to this bias, the raters (faculty, students, recruiters…) attribute their positive perceptions of a high-reputation university to its schools, departments, faculty, outputs… Authors affiliated to highly reputed universities might enjoy a halo effect (Dey et al. 1997; Fowles et al. 2016;

Medoff [2006]). This effect would increase their chances of publishing in the best journals and of obtaining more citations due to editor, referee and citation biases. This effect acts as a research performance potentiator, as a moderating variable, but lacks value in itself without a quantitatively and qualitatively acceptable scientific output.

The risk of a paper's rejection derives from multiples biases (Lee et al. [2013]). Two of the most significant are the prestige and affiliation biases, which affect editors and referees, because where authors' identities are known, their institutional affiliations may influence judgements (Clark and Wright [2007]). According to Miller ([2006], p. 425), current affiliations "provide clues, albeit imperfect ones, as to the competency of a manuscript's author(s), and these clues can be used as a means to avoid the burden of deeply evaluating research or as unconscious or conscious shortcuts around any uncertainty about the value of a submission. […] In fields with blind review, […] status based on institutional affiliations are less troublesome, at least at the reviewer level (they can still be operative at the editor level)". In the same line, Judge et al. ([2007], p. 495) state that "even at journals that use blind peer review, author status and affiliation may influence editors, who may 'front-load' articles by prominent scholars as a way of legitimizing the journals".

Some evidence exists on these biases. Peters and Ceci ([1982]) studied the editor and referee biases by resubmitting published articles written by high-status authors from prominent institutions under fictitious author names and affiliations. They found that resubmitted papers were rejected higher than the journal's rejection rate on the basis that the studies contained serious methodological defects. McGillivray and De Ranieri ([2018]) found that corresponding authors from less prestigious institutions are more likely to choose double-blind review, in order to avoid affiliation bias. They also found that manuscripts from authors from less prestigious institutions are accepted with a lower rate than those from more prestigious institutions. Ross et al. ([2006]) provide evidence of bias in the open review of abstracts, proposing that the system favors authors from prestigious academic institutions. Garfunkel et al. ([1994]) also verified that the likelihood of recommendation for acceptance and of selection for publication of brief reports appeared to correlate with the prestige of the institution, although they obtained inconclusive results for more extensive manuscripts. Baghestanian and Popov ([2014]) found a positive relationship between affiliation and the probability of publishing in a top economic journal. Judge et al. ([2007]) find a correlation between affiliation and citation on a set of articles published in the top 21 management journals. Lou and He ([2015]) find that affiliation and uncitedness are negatively correlated. And Oswald ([2008]) tested pro-Harvard bias and Yotopoulos ([1961]) pro-Harvard and pro-Chicago bias.

In relation to paper evaluation systems, Blank ([1991]) observes a referee bias towards highly reputed universities under single-blind refereeing (vs double-blind) from an experiment conducted at *The American Economic Review*. Double-blind peer review has been proposed as a means to avoid referee bias (Ross et al. [2006]), however, this type of review model has sometimes been shown to fail to hide authors' identity (McGillivray and De Ranieri [2018]). For example, Falagas et al. ([2006]) found that the half of the cases, authors were in fact recognizable.

Another bias comes from citations, an element considered in global rankings such as ARWU and THE.[1] Medoff ([2006]) found that papers authored by researchers from elite institutions obtained higher citation rates, even after controlling for author and journal

---

[1] See Van Raan ([2005]) for a discussion on the major problems in the application of publication and citation data in the context of evaluation.

quality and characteristics of the article. More recently, Meyer et al. (2018) investigate the number of citations received by papers published in six well-known management science journals and find that the status of the first author's institution was a strong predictor.

On the other hand, the editor and referee biases amplify citation bias by the bias that top journals introduce into the citation process. Larivière and Gingras (2010) shows that the journal in which papers are published have a strong influence on their citation rates, as duplicate papers published in top journals obtain twice as many citations as their identical counterparts published in journals with lower impact factors.

Figure 1 summarizes reputational effects among different types of rankings (with or without a reputational survey) and between rankings from different years. It is important to note that universities are the unit of analysis of the model, although the model shows reputational biases that occur at the individual level and their effects are subsequently aggregated in the rankings.

A key aspect of the model shown in Fig. 1 is *speed of feedback* (dotted line). Rankings influence the reputation of university institutions from the moment they are published (Argenti 2000; Kane 1993; Safón 2007; Wedlin 2007), since they are widely disseminated through various media channels—who are paying increasing attention to higher education rankings (González-Riaño et al. 2014)—and via the institutions themselves through their websites and press releases[2] (ARWU 2018; Luca and Smith 2015). Consequently, a lagged effect of between one and 2 years can be expected, as the most recent information is the one that is uppermost in the minds of rankers (Keller et al. 2008). The studies by Bowman and Bastedo (2011) and Stake (2006) support the idea of a rapid return of between one and 2 years for reputational surveys to complete the circuit and feed back into rankings. The circuit derived from the publication and citation process also feeds back quickly, particularly since manuscript management systems and online publication have notably accelerated the evaluation and publication stages for papers, and hence citations appear shortly after. Studies such as those by Wang et al. (2015) patently show that these stages have become markedly shorter, showing citations from papers that were published only a few months before. In any case, older studies, such as that of Moed et al. (1985), show that the maximum level of citations occurs in the third year of publication.

The institutional version of the Matthew effect allows for the introduction of an important element in the model: the attraction of doctoral students, top researchers and funds. According to this effect, made popular by Merton (1968), the most prestigious universities receive recognition and disproportionate awards, to the notable detriment of the less highly reputed institutions. This effect foresees that top research universities will concentrate ''the lion's share of federal support and super-star researchers (e.g., Nobel laureates)'' (Bentley and Blackburn 1990, p. 329), which allows them to obtain higher levels of productivity. This effect has been demonstrated by Dey et al. (1997).

Figures 2 and 3 show the variables used by the ARWU and the THE rankings that are most directly involved in shaping reputational bias. 60% of ARWU and 69% of THE are affected by this bias. The publication of ARWU and THE rankings, in August and September respectively, generate immediate effects on the opinions incorporated in the THE ranking the following year (this ranking also takes into account opinions from the previous

---

[2] This statement refers to the communication of the results of rankings, which complement other types of communication that also influence perceptions, e.g. speeches given by vice-chancellors or rectors (Rodriguez-Pomeday and Casani 2016) or the degree of news coverage of the university (Arpan et al. 2003), among others.
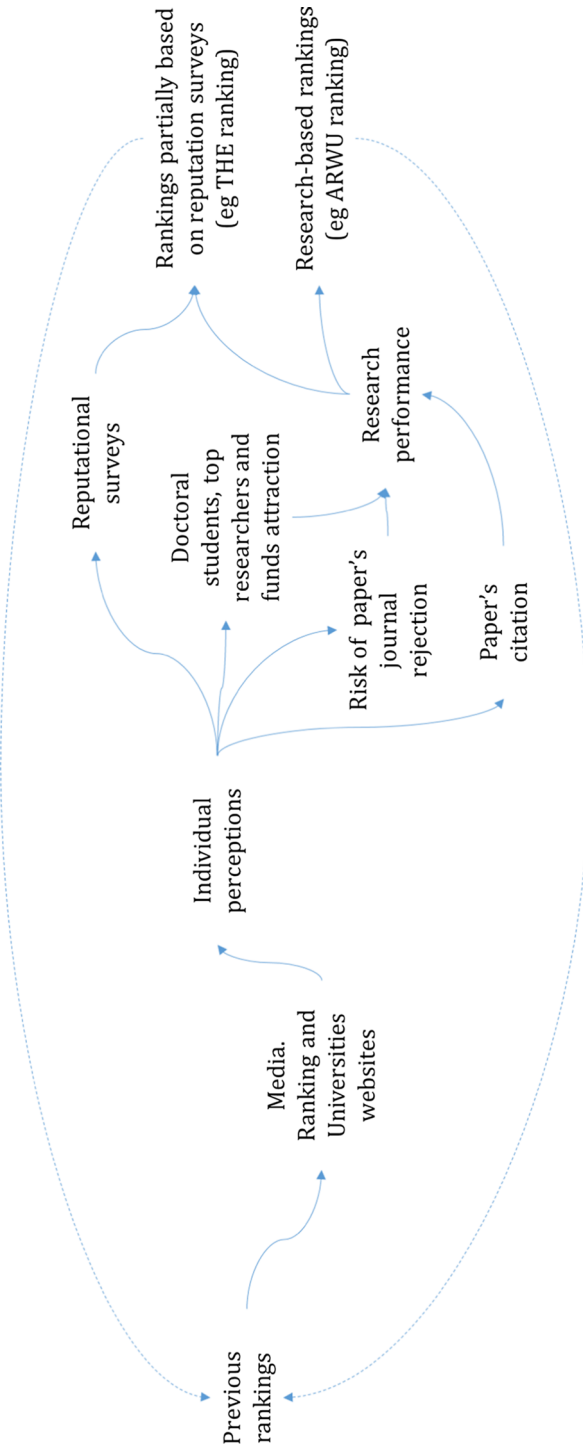
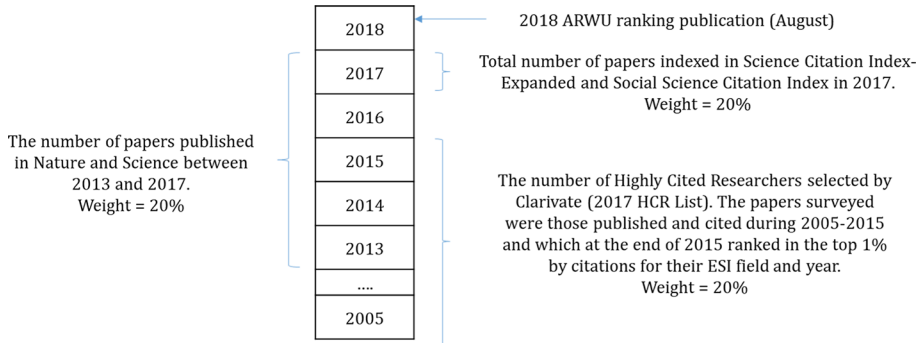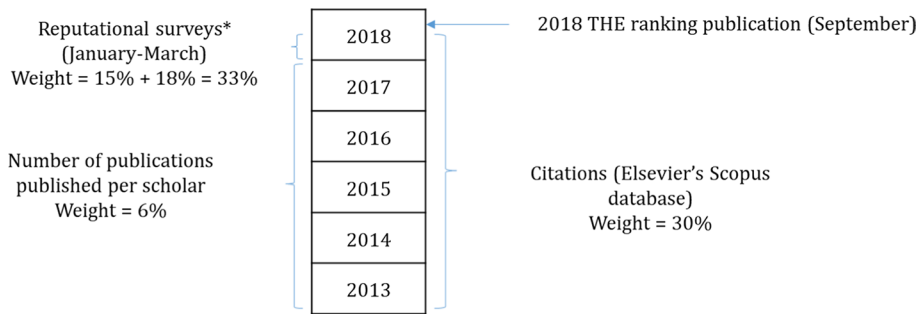**Fig. 1** Inter- and intra-ranking reputation effects model

Fig. 2 Indicators from the 2018 ARWU ranking that potentially channel reputational biases

*The 2018 data are combined with the results of the 2017 survey.

Fig. 3 Indicators from the 2018 THE ranking that potentially channel reputational biases

year). The influence of rankings in the production and citation of articles through editor, referee and citation biases takes more time, although the initial signs could be observed in the production and citation of articles just 2 years later.

The evolution over time of a university's ranking also depends upon a *compensatory effect* that tends to arise in rankings in accordance with changes in the preceding years, thus neutralizing any improvements or declines in one particular year with subsequent opposite effects. These effects inherently skew the effectiveness of the circuit described in Fig. 1.

Rankings are periodical products, and normally appear yearly. Changes from 1 year to the next express, according to their promoters, relative changes in the performance of a university in one or several of their three missions: research, teaching and knowledge transfer. However, despite the existence of a strong correlation between the lists produced by the same ranking over time (Aguillo et al. 2010; Fowles et al. 2016; Shehatta and Mahmood 2016), commonly interpreted as evidence of internal coherence, annual variations in university positions are, in fact, rather erratic.

Several studies have concluded that rankings contain a good deal of noise in their annual variations, particularly in rankings based on reputational surveys (Sorz et al. 2015), and that their changes tend to be reversible over time. Dichev (2001) posits that between 70 and 80% of the variation in USNWR changes is due to transitory and reversible noise. Gnolek et al. (2014) determine that up to four positions in the USNWR annual rank changes should be considered noise. Dichev (2001) shows that USNWR rankings have a strong tendency

to revert in the next two rankings. These results show that, for US universities, about 30% of a given change in the rankings reverses in the very next ranking, and another 23% unravels in the ranking 2 years on. Bookstein et al. (2010) also show evidence of the tremendous amount of noise present in the annual variations in the THE-QS ranking, derived, though not exclusively, from surprising yearly changes in normally stable variables, such as staff-student ratio.[3] Saisana et al. (2011) obtain similar findings when analyzing the ARWU and THE rankings from 2008 and conclude that, from a statistical perspective, a sizeable number of universities occupy unstable positions in the rankings (43 universities in the ARWU top 100 and 61 universities in the THE top 100). Sorz et al. (2015) analyze the THE and ARWU rankings from 2010 to 2014 and find inconsistent fluctuations in the rank and score for universities below position 50. Pinar et al. (2019) find that ARWU and THE rankings are very sensitive to weight variations, especially for middle- and low-ranked universities.

These results support the idea of a longitudinal equilibrium in the rankings (Fowles et al. 2016), and challenge the hypothesis that rankings capture variations in university performance over time.

Methods

Sample

At present, there are over 20 global rankings (Hazelkorn 2017). This study focuses on the two most important global rankings in terms of readership and influence (Kauppi 2018; Piro and Sivertsen 2016; Sorz et al. 2015) i.e. the ARWU and THE rankings. The former does not use reputation surveys, while the latter does.

ARWU is the longest world ranking available. Between 2003 and 2016, it listed 500 universities. In 2017, it increased the list to 800 and in 2018 to as many as 1000. Its methodology relies on six research-focused indicators: number of alumni or faculty winning Nobel Prizes or Fields Medals, number of highly cited researchers, number of articles published in Nature or Science, number of articles in the Science Citation and Social Sciences Citation indices, and per capita performance of the institution. This ranking is largely size-dependent, although it partially corrects this effect with the dimension *per capita performance of the institution*. It focuses only on research and is mainly bibliometric, although the bibliometric dimension is not complete, as it does not contemplate citation analysis of all the papers of a university. Only in the indicator *highly cited scientists* are citations involved, and this indicator is a measure of past performance. The indicators based on *Nobel Prizes and Fields Medals* reflect past rather than current performance and disadvantage recently established universities (Waltman et al. 2012). A detailed explanation of the methodology can be found at www.shanghairanking.com/ARWU-Methodology-2018.html.

The ARWU was closely followed by the THE–QS World University Rankings, which annually published a joint ranking from 2004 to 2009. In 2010, the THE broke away from QS and launched its own independent ranking using a new methodology. In 2010, it listed 200 universities; between 2011 and 2014, 400 universities; in 2015, the number rose to 800; in 2016, it reached 978, in 2017, the number was 1260 and in 2018, it listed a total of

---

[3] This data is provided by universities themselves and is easily manipulated (Waltman et al. 2012), which explains the high fluctuation from year to year. Rankings based entirely on bibliometric data do not suffer from this type of problem.

1258 universities. The ranking focuses on research and teaching dimensions, and in both university tasks expert opinion is involved in the calculation of ranking scores. Reputational surveys make up 33% of the total ranking. The ranking is generated using the following variables: citations impact, income per academic, reputational survey (teaching), Ph.D. awards per academic, Ph.D. awards to bachelor's awards, undergraduates admitted per academic, papers academic and research staff, research income (scaled), reputation survey (research), ratio of international to domestic students, ratio of international to domestic staff, proportion of internationally co-authored research papers, and research income from industry (per academic staff). Due to the normalization process of the variables, the ranking is largely size-independent. A detailed explanation of the method can be found at www. timeshighereducation.co.uk.

With a view to avoiding the effects of the methodological change in the THE in 2010 and the changes in the size of the listings, this study analyzes the THE and ARWU rankings between 2010 and 2018, using the first 500 universities in each ranking. As later explained, this research works on the basis of a maximum lag of 3 years. The ranking with the shortest span is THE, which accounts for a total of 9 years. By working with annual variations, 1 year is lost, and by introducing variables with a lag of up to 3 years, 3 years cannot be included in the sample. Therefore, the total potential sample is of 5 years for 500 universities representing 2500 observations. However, as these two rankings do not rank exactly the same set of universities for their different methodologies and inclusion criteria,[4] the sample is reduced to 2210 observations.[5]

Variables

If the empirical evidence is compatible with the reputation bias, an association between the variations of rankings from different years can be expected. The year-to-year changes in the ARWU and THE rankings[6] are the dependent variables (DV). They were captured by comparing a university's ranking in year t with that of year $t-1$ ($t=2014$–2018). This data was transformed into categorical variables that indicate declining ($=-1$), improving ($=+1$), or maintaining the same position ($=0$) from the previous year.

With a view to capturing reciprocal influences between the two rankings, variations in the other ranking with a lag of 1, 2 and 3 years are used as independent variables (IV). The IVs take into account changes in the years 2011 (change in comparison with 2010) to 2017 (change from 2016), and are measured in the same way as DVs. As shown by Figs. 2 and 3, these lags allow us to capture the effect of the variation of a university in a ranking on

---

[4] According to the THE methodology "Universities can be excluded from the World University Rankings if they do not teach undergraduates, or if their research output amounted to fewer than 1,000 relevant publications between 2013 and 2017 (with a minimum of 150 a year). Universities can also be excluded if 80 per cent or more of their research output is exclusively in one of our 11 subject areas" (timeshighereducation. com). The complete inclusion criteria appear in the document *Methodology for overall and subject rankings for the Times Higher Education World University Rankings 2019 (September 2018)*, published in https ://www.timeshighereducation.com/sites/default/files/the_2019_world_university_rankings_methodolog y_pwc.pdf

[5] The THE ranking was preferred over QS due to its wider audience, but also because of the greater overlap with ARWU. Shehatta and Mahmood (2016) observe that ARWU and QS 2015 results are the lowest overlapping in terms of universities among the six global rankings studied.

[6] Overall rankings are taken. These rankings are those that are disseminated by the media and have reputational effects.

Table 1 Descriptive statistics (full sample)

Variable	Mean	SD	1	2	3	4	5	6	7
1. THE (t) Ch.	− .06	.82							
2. THE ($t-1$) Ch.	− .04	.82	**−.11**						
3. THE ($t-2$) Ch.	− .04	.82	−.04	**−.12**					
4. THE ($t-3$) Ch.	.03	.82	−.01	−.06	**−.13**				
5. ARWU (t) Ch.	.00	.65	**.09**	**.06**	**.08**	−.01			
6. ARWU ($t-1$) Ch.	.00	.63	.01	**.08**	**.05**	**.07**	**−.13**		
7. ARWU ($t-2$) Ch.	.02	.60	**.05**	.01	**.09**	.04	**.08**	**−.09**	
8. ARWU ($t-3$) Ch.	.00	.56	.03	**.07**	.02	**.07**	**.08**	**.09**	−.06

Significant Pearson's correlations (bilateral) in bold. $N=2210$. Ch. = change; $t=2014$, 2015, 2016, 2017 and 2018

the perceptions reflected in reputational surveys, article publication, and to some extent article citation. Two more arguments justify the choice of three lags. Firstly, bearing in mind that the THE ranking is only 9 years long, applying a greater number of lags results in too much of the sample being lost. Secondly, previous studies do not observe a relation between ranking changes in t and ranking changes in $t-4$ (Dichev 2001).

As mentioned in Sect. 2, several studies have observed greater stability towards the top part of the rankings (especially among the top 50), and conversely, a higher likelihood of change (and noise) in the middle and lower parts. Reputational effect is also more notable amongst the top universities. In order to control for this effect, the analysis was applied to the full sample and to the subsample made up of the top 50 in each ranking for $t-1$.

In order to control for the compensatory and reputational effects derived from the same ranking (intra-ranking effects), variations were introduced in the same ranking in the three previous years. These two effects have an opposite influence on DV, and thus conclusions can only be drawn in comparative terms.

Finally, to control for effects previous to the 3 years considered here or reputational effects from rankings other than the THE and ARWU, or of any other kind that might affect the study, the variations from other rankings in that same year were also introduced.

Analysis

Due to the nature of the DVs, ordinal regressions were chosen and 2 models were applied to each ranking. The first uses the full sample and the second the subsample made up of the top 50. For each model, two regressions are applied: the first introduces the control variables and the second adds the IVs. According to the distribution of the DV in each sample, the most appropriate link function was applied, these being the logit, probit and cauchit functions.

Results

Table 1 shows the descriptive statistics and the correlations. Most of the correlations are significant, whilst there is no indication of multicollinearity. The first-order autocorrelations are negative and significant, which indicates that the compensatory effect is present in

the full sample studied in both rankings, and that it is also more intense than the intra-ranking reputational effect. The reputational effect can be observed, however, in the inter-ranking correlations. In the case of THE, the significant correlations can be seen in the two-year lag in the ARWU ranking ($r_{\text{THE}(t),\ \text{ARWU}(t-2)} = .05$, $p < .05$, and $r_{\text{THE}(t-1),\ \text{ARWU}(t-3)} = .07$, $p < .05$). With regard to ARWU, the correlations are significant and positive with one or two lags, except in the case of $r_{\text{ARWU}(t-2),\ \text{THE}(t-3)}$, which is not significant at the .05 level ($p = .07$).

Table 2 shows the regressions carried out. *Panel A* contains regressions where the DV is change in the current (t) THE ranking. *Panel B* shows regressions that analyze change in the current ARWU ranking.

The regression in *Panel A* carried out with the full sample (Model 1) is significant ($p < .001$). The predictive capacity, measured using the pseudo R^2, is low. However, this result is in line with those of other related research. For example, Dichev (2001) reports adjusted R^2 between .04 and .12 for regression results for the relation between current and lagged changes for the US Top 25 rankings of national universities. In our case, the sample includes all universities, where even more noise can be expected.

Previous THE changes have a decreasing influence on current THE changes. This effect means that changes in 1 year tend to peter out in subsequent and become less intense. Variation in the previous year has the greatest impact, while changes 3 years prior to the ranking is not significant. There is no evidence to suggest that changes in the previous THE rankings changes influence intra-ranking reputational effects, as no positive, significant parameters are apparent. The effect may well exist, but if it does, the effect would be cancelled out by the negative compensatory effect.

Previous ARWU ranking changes are the IVs. Model 1b analyzes them for the full sample. The model is significant, but the effect size of the IVs on DV is very low. The coefficients for the previous ARWU changes in $t-1$ ($\beta = .08$, $p < .05$), $t-2$ ($\beta = .11$, $p < .01$) and in $t-3$ ($\beta = .07$, $p < .10$) were positive and significant (Model 1b), and a log-likelihood ratio test (Aldrich and Nelson 1984) shows that entering all IV together results in a significant increase in fit with Model 1a. The greatest effect can be observed in the variable with two lags.

Model 2 focuses attention solely on the top 50 universities. The results are similar to those from Model 1, though with reduced significance due to the small sample size. Model 2b verifies that the compensatory effects continue to be present for $t-1$ ($\beta = -.21$, $p < .05$) and for the inter-ranking effect in $t-2$ ($\beta = .15$, $p < .10$). As in Model 1, the effect size of the control and dependent variables are very low.

Panel B in Table 2 studies the current ARWU changes as a DV. Model 3 analyzes the full sample. The regressions are significant ($p < .001$) and the predictive capacity of the control model and of the full model, measured with the pseudo R^2, is low, though slightly greater than that observed in Models 1 and 2. Model 3b improves the fit indicators of the control model and is higher on the log-likelihood ratio test.

Changes from the previous year are compensated the following year, which supports the theory of the compensatory effects. However, previous ARWU changes that occurred two or 3 years prior have a positive effect on current ARWU changes, which lends weight to the reputational intra-ranking effect.

In *Panel 2*, the IVs are the lagged variations from the THE ranking. The variations that occur one or 2 years before significantly influence the ARWU ranking (Model 3b), but not those that occurred 3 years before ($\beta = .10$, $p < .001$, $\beta = .11$, $p < .001$, and $\beta = .02$, $p > .05$, respectively). These results strengthen the idea of an inter-ranking reputational effect, although effect size is small.

Table 2 Ordinal regression results for the relation between current and lagged changes for the THE and ARWU rankings

Variable	Model 1 Full sample		Model 2 Top 50	
	1a. Controls	1b. Full model	2a. Controls	2b. Full model
Panel A. Dependent variable: THE (t) Ch.				
THE (t−1) Ch.	−.24***	−.26***	−.22*	−.21*
THE (t−2) Ch.	−.13***	−.15***	−.03	−.03
THE (t−3) Ch.	−.05	−.07	.03	.04
ARWU (t) Ch.	.19***	.19***	.20*	.15†
ARWU (t−1) Ch.		.08*		.02
ARWU (t−2) Ch.		.11**		.15†
ARWU (t−3) Ch.		.07†		.11
Pseudo R²				
Nagelkerke	.030	.036	.040	.057
McFadden	.012	.015	.017	.024
Chi square	59.37***	72.36***	9.01†	13.05†
Link function	Logit	Logit	Cauchit	Cauchit
N	2210	2210	251	251

Variable	Model 3 Full sample		Model 4 Top 50	
	3a. Controls	3b. Full model	4a. Controls	4b. Full model
Panel B. Dependent variable: ARWU (t) Ch.				
ARWU (t−1) Ch.	−.15***	−.17***	−.05	.01
ARWU (t−2) Ch.	.09***	.07**	.28**	.26**

Table 2 (continued)

Variable	Model 3 Full sample		Model 4 Top 50	
	3a. Controls	3b. Full model	4a. Controls	4b. Full model
ARWU ($t-3$) Ch.	.11***	.10***	.18*	.10
THE (t) Ch.	.10***	.11***	.09	.14
THE ($t-1$) Ch.		.10***		.07
THE ($t-2$) Ch.		.11***		.32**
THE ($t-3$) Ch.		.02		−.07
Pseudo R^2				
Nagelkerke	.045	.062	.057	.094
McFadden	.020	.028	.024	.040
Chi square	87.21***	120.26***	12.14*	20.41**
Link function	Probit	Probit	Cauchit	Cauchit
N	2210	2210	235	235

†$p<.10$; *$p<.05$; **$p<.01$; ***$p<.001$. Ch.=change; $t=2014, 2015, 2016, 2017$ and 2018

Model 4 studies the subsample made up of the top 50 universities. The regressions are significant, as is the log-likelihood ratio test. The results do not support the theory of the reversibility of the rankings, as no significant negative effects can be observed in the previous ARWU changes (Model 4a). However, there is evidence of significant, positive inter and intra ranking effects on ARWU and THE changes 2 years before ($\beta = .26$, $p < .01$, and $\beta = .32$, $p < .01$, respectively).

Our results identify a positive association between the current ARWU and THE changes in all the models except for Model 4. Evidently, there is no causal relation between the two variables, but the fact that they are related means that, regardless of the other variables analyzed, both rankings are capturing common elements that are not considered in the regressions. One possibility is that research performance prior to the last 3 years is being captured by those variables.

Additional test

Taking a maximum delay of 3 years has allowed for a sample of 2210 observations but has limited capturing the effects derived from the citations, especially in the case of ARWU (see Fig. 2), since the first citations they consider are at least 3 years old. The tests shown in Table 2 were repeated for the full sample taking a delay of 5 years (now $N = 1326$), with only positive results for 'THE $(t-4)$ Change' over 'ARWU (t) Change' ($p = .029$), although with less effect size than 'THE $(t-2)$ Change'.

Discussion and conclusion

This study examines the existence of intra- and inter-ranking reputational effects in two very different global rankings: ARWU and THE rankings. The main findings can be summarized as follows.

The reversibility of the rankings over time and the stability of the listings towards the top have suggested the theory of a longitudinal equilibrium (Fowles et al. 2016). The results of this study support that theory. In line with previous research (Dichev 1999 and 2001), our results show that annual changes in the two rankings analyzed compensate ranking changes from earlier years, where the compensatory effect is greater in the following year.

The strongest compensatory effect can be observed in the THE ranking, a result that is also coherent with previous literature, such as the study by Sorz et al. (2015), who find that, due to their methodology, the ARWU ranking has much less noise compared to the THE ranking.

In this study, the compensatory effects are reduced among the top universities. No significant effect can be seen for ARWU (Model 4b) and for THE, it can only be observed for the one-year lagged variable (Model 2b). This result is coherent with the literature that has reported greater levels of noise in the middle and lower parts of the rankings (e.g. Pinar et al. 2019; Sorz et al. 2015).

This study theoretically justifies that rankings suffer from reputational bias derived from the anchoring, echo, halo and Matthew effects. These effects explain why rankings are affected by the reputation of universities, due partly to the "reputation-ranking-reputation" circle shown in Fig. 1. Our results are compatible with this theory, not only in rankings

that include reputational surveys, such as the THE ranking, but also for those that do not include reputational surveys, such as ARWU.

The speed of feedback of the 'reputation-ranking-reputation' model is measured here by introducing lagged variables. Reputational biases seem to exist, which have a relatively immediate effect of between one and 2 years, but also appear with less intensity in the third and fourth years. Such results are coherent with the echo-chamber effect theory offered by Stake (2006).

Our results have provided positive evidence of intra- and inter-ranking effects, but we should not forget that the effect size is markedly low. This means that there is sufficient, albeit precarious, support, whilst it cannot be ruled out that the results may have causes other than those proposed here. New research should continue to investigate the issues highlighted in this paper, adding new tests and rankings to the analysis. Of particular interest is the study of rankings based exclusively on bibliometric data, such as the Leiden ranking (Waltman et al. 2012). It is also necessary to enlarge the number of control variables, thereby proposing more demanding models. The current results show positive levels of significance, but very low effect sizes. Broader tests will have to challenge the levels of significance reached.

Conclusions

This study proposes an intra- and inter-ranking relationship model based on varying reputational biases that affect global rankings, regardless of whether they include peer assessments or not. The model contributes to the literature on university reputation and global rankings and integrates a variety of previous studies on the topic.

The model is tested using two highly influential rankings, with very different methodologies: the ARWU and THE rankings. The former only considers objective indicators of research performance, while the latter is more complex and includes reputation surveys with a weighting of 33% out of the total ranking score. Up until now, no study has analyzed both rankings from the perspective of their reputational interrelations.

This research obtains empirical evidence of a longitudinal equilibrium, and also of the theory that reputational effects have an effect on all universities, not only on the top-ranking institutions.

An interesting question arises for scientometrics derived from these results: is it possible to construct a ranking based on bibliometric data free from reputational bias? Or rather, free from old reputation. To the extent that the publication of papers and the generation of citations is affected by reputational biases, is it possible to generate bibliometric data free from old reputation? Additionally, while there are rankings based on opinions, and others exclusively based on recent bibliometric data, such as the Leiden Ranking (Waltman et al. 2012), can the level of reputation bias for different rankings be obtained? It is to be expected that rankings that are less dependent on old reputation have less reputational bias than rankings based on opinions, such as the THE or QS rankings.

Our study also has implications at the university level. Administrators should be aware of reputational effect when attempting to improve their positions in the rankings or defending those already achieved, as part of the changes that occur in rankings are systematic, not random. Several studies explain how certain corporate strategies, such as mergers (Bolbanabad et al. 2017; Docampo et al. 2015; Harman and Harman 2008), or certain strategies related to how to organize research (Pang 2018) or how to manage it (Safón 2013), or

even some dubious ethical practices such as highly cited researchers' secondary affiliations (see the case of King Abdulaziz University in Bhattacharjee 2011; Messerly 2014), may improve positions in the global rankings. Our model explains how gains in the rankings throw out signals to the market that are subsequently amplified and feedback, creating a sustaining force for universities that achieve significant improvements in the global rankings. Declines in rankings also have consequences, either because universities lose some of their *reputational stock*, or because competitors that are "growing" overtake those that have lost their 'appeal'. It should be remembered that there is not a place for everyone at the top of the rankings (top 10, top 50, top 100). This is clearly evident in the case of US universities, who have previously enjoyed reputational effects. Between 2010 and 2018, 12 US universities disappeared from the top 100 of the THE ranking and 8 from the top 100 of the ARWU ranking. Their places have been taken up by other universities, several of which are Chinese. It is possible that such universities are benefitting from the reputational circle. The proposed model advises institutions not to lose sight of reputational strategies when they come to establish their objectives, and these are essential both to defensive and offensive strategies.

Finally, the dynamic vision of reputation and its reciprocal influence on rankings should not forget the large amount of noise that ARWU and THE rankings incorporate into their annual changes. Any effort to benefit from virtuous reputational circles demands a sizeable sustained effort to overcome such noise.

References

Aguillo, I., Bar-Ilan, J., Levene, M., & Ortega, J. (2010). Comparing university rankings. *Scientometrics, 85*(1), 243–256.

Aldrich, J., & Nelson, F. (1984). *Linear probability, logit and probit models. Sage university paper series on quantitative applications in the social sciences*. Beverly Hills: Sage.

Argenti, P. (2000). Branding b-schools: Reputation management for MBA programs. *Corporate Reputation Review, 3*(2), 171–178.

Armstrong, J. S., & Sperry, T. (1994). Business school prestige-research versus teaching. *Interfaces, 24*(2), 13–43.

Arpan, L. M., Raney, A. A., & Zivnuska, S. (2003). A cognitive approach to understanding university image. *Corporate Communications: An International Journal, 8*(2), 97–113.

ARWU. (2018). *About ARWU*. http://www.shanghairanking.com/aboutarwu.html Accessed 19 December 2018.

Baghestanian, S., & Popov, S. V. (2014). Alma matters: Determinants of early career success in economics. Technical report, Goethe University.

Bastedo, M. N., & Bowman, N. A. (2009). US news & World report college rankings: Modeling institutional effects on organizational reputation. *American Journal of Education, 116*(2), 163–183.

Bentley, R., & Blackburn, R. (1990). Changes in academic research performance over time: A study of institutional accumulative advantage. *Research in Higher Education, 31*(4), 327–353.

Bhattacharjee, Y. (2011). Saudi universities offer cash in exchange for academic prestige. *Science, 334*, 1344–1345.

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Evidence from the American economic review. *American Economic Review, 81,* 1041–1067.

Bolbanabad, A. M., Mosadeghrad, A. M., Arab, M., & Majdzadeh, R. (2017). Impact of merger and acquisition on university performance. *Archives of Iranian Medicine, 20*(8), 518–524.

Bookstein, F. L., Seidler, H., Fieder, M., & Winckler, G. (2010). Too much noise in the times higher education rankings. *Scientometrics, 85*(1), 295–299.

Bowman, N. A., & Bastedo, M. N. (2011). Anchoring effects in world university rankings: Exploring biases in reputation scores. *High Education, 61,* 431–444.

Brooks, R. (2005). Measuring university quality. *The Review of Higher Education, 29*(1), 1–21.

Çakır, M. P., Acartürk, C., Alaşehir, O., & Çilingir, C. (2015). A comparative analysis of global and national university ranking systems. *Scientometrics, 103*(3), 813–848.

Clark, C., & Wright, M. (2007). Reviewing journal rankings and revisiting peer reviews: Editorial perspectives. *Journal of Management Studies, 44*(4), 612–621.

Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review, 32*(3), 377–390.

Dey, E. L., Milem, J. F., & Berger, J. B. (1997). Changing patterns of publication productivity: Accumulative advantage or institutional isomorphism? *Sociology of Education, 70*(4), 308–323.

Dichev, I. (1999). How good are business school's rankings? *Journal of Business, 72*(2), 201–213.

Dichev, I. (2001). News or noise? *Research in Higher Education, 42*(3), 237–266.

Docampo, D., Egret, D., & Cram, L. (2015). The effect of university mergers on the Shanghai ranking. *Scientometrics, 104*(1), 175–191.

Dyck, I. & Zingales, L. (2002). *The corporate governance role of the media*. CRSP working paper, no. 543. Retrieved October 19, 2005. from ssrn.com/abstract=335602.

Falagas, M. E., Zouglakis, G. M., & Kavvadia, P. K. (2006). How masked is the "masked peer review" of abstracts submitted to international medical conferences? *Mayo Clinic Proceedings, 81*(5), 705.

Fombrun, C. (1996). *Reputation: Realizing value from the corporate image*. Boston: Harvard Business School Press.

Fowles, J., Frederickson, H. G., & Koppell, J. G. (2016). University rankings: Evidence and a conceptual framework. *Public Administration Review, 76*(5), 790–803.

Garfunkel, J. M., Ulshen, M. H., Hamrick, H. J., & Lawson, E. E. (1994). Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA, 272*(2), 137–138.

Gnolek, S. L., Falciano, V. T., & Kuncl, R. W. (2014). Modeling change and variation in US news & World report college rankings: What would it really take to be in the top 20? *Research in Higher Education, 55*(8), 761–779.

González-Riaño, G., Repiso Caballero, R., & Delgado López-Cózar, E. (2014). The impact of university rankings in the Spanish press (Repercusión de los rankings universitarios en la prensa española). *Revista Española de Documentación Científica, 37*(3), e055.

Harman, G., & Harman, K. (2008). Strategic mergers of strong institutions to enhance competitive advantage. *Higher Education Policy, 21*(1), 99–121.

Hazelkorn, E. (Ed.). (2016). *Global rankings and the geopolitics of higher education: Understanding the influence and impact of rankings on higher education, policy and society*. Abingdon: Taylor & Francis.

Hazelkorn, E. (2017). *Rankings and higher education: Reframing relationships within and between states*. Centre for Global Higher Education working paper no. WC1H 0AL. London, the UK: UCL Institute of Education.

Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited—Article, author, or journal? *Academy of Management Journal, 50*(3), 491–506.

Kane, K. F. (1993). MBAs: A recruiter's-eye view: Important criteria for MBA positions. *Business Horizons, 36*(1), 65–71.

Kauppi, N. (2018). The global ranking game: Narrowing academic excellence through numerical objectification. *Studies in Higher Education, 43*(10), 1750–1762.

Keith, B. (2001). Organizational contexts and university performance outcomes: The limited role of purposive action in the management of institutional status. *Research in Higher Education, 42*(5), 493–516.

Keller, K., Apéria, T., & Georgson, M. (2008). *Strategic brand management*. Upper Saddle River: Prentice Hall.

Kohm, L. M., & Wardle, L. D. (2011). The echo-chamber effect in legal education: Considering family law casebooks. *St. Thomas Journal of Law & Public Policy, 6*(1), 104–140.

Larivière, V., & Gingras, Y. (2010). The impact factor's matthew effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology, 61*(2), 424–427.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17.

Leiter, B. (2006). Commentary: How to rank law schools. *Indiana Law Journal, 81,* 47–52.

Locke, W., Verbik, L., Richardson, J., & King, R. (2008). *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England*. Bristol: Higher Education Funding Council for England.

Lou, W., & He, J. (2015). Does author affiliation reputation affect uncitedness? In *Proceedings of the 78th ASIS&T annual meeting: Information science with impact: Research in and for the community* (p. 103). American Society for Information Science.

Luca, M., & Smith, J. (2015). Strategic disclosure: The case of business school rankings. *Journal of Economic Behavior & Organization, 112,* 17–25.

McGillivray, B., & De Ranieri, E. (2018). Uptake and outcome of manuscripts in Nature journals by review model and author characteristics. *Research Integrity and Peer Review, 3*(5), 1–12.

Medoff, M. H. (2006). Evidence of a harvard and Chicago Matthew effect. *Journal of Economic Methodology, 13*(4), 485–506.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56–63.

Messerly, M. (2014). *Citations for sale: Saudi university boosts rankings by aggressively recruiting world's top researchers.* The daily Californian. http://www.dailycal.org/2014/12/05/citations-sale. Retrieve 1 December 2018.

Meyer, M., Waldkirch, R. W., Duscher, I., & Just, A. (2018). Drivers of citations: An analysis of publications in "top" accounting journals. *Critical Perspectives on Accounting, 51,* 24–46.

Miller, C. C. (2006). Peer review in the organizational and management sciences: Prevalence and effects of reviewer hostility, bias, and dissensus. *Academy of Management Journal, 49*(3), 425–431.

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics, 110*(2), 967–990.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*(3), 131–149.

Morgeson, F. P., & Nahrgang, J. D. (2008). Same as it ever was: Recognizing stability in the business week rankings. *Academy of Management Learning & Education, 7,* 26–41.

Olcay, G. A., & Bulu, M. (2017). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change, 123,* 153–160.

Oswald, A. J. (2008). *Can we test for bias in scientific peer-review?* IZA DP No. 3665, Forschungsinstitut zur Zukunft der Arbeit Institute for the Study of Labor.

Pang, L. (2018). *How Tsinghua became a world class research university: A case study on the impact of rankings on a Chinese higher education institution* (doctoral dissertation).

Peters, D. P., & Ceci, S. J. (1982). A naturalistic study of psychology journals: The fate of published articles resubmitted. *Behavioral and Brain Sciences, 5,* 187–252.

Pinar, M., Milla, J., & Stengos, T. (2019). Sensitivity of university rankings: implications of stochastic dominance efficiency analysis. *Education Economics, 27*(1), 75–92.

Piro, F. N., & Sivertsen, G. (2016). How can differences in international university rankings be explained? *Scientometrics, 109*(3), 2263–2278.

Rauhvargers, A. (2011). *Global university rankings and their impact.* Brussels: European University Association.

Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005). Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Academy of Management Journal, 48*(6), 1033–1049.

Rodriguez-Pomeday, J., & Casani, F. (2016). Legitimating the world-class university concept through the discourse of elite universities' presidents. *Higher Education Research and Development, 35*(6), 1269–1283.

Rondeau, K. V. (2017). The impact of world ranking systems on graduate schools of business: Promoting the manipulation of image over the management of substance. *World Journal of Education, 7*(3), 62–73.

Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., et al. (2006). Effect of blinded peer review on abstract acceptance. *JAMA, 295*(14), 1675–1680.

Safón, V. (2007). Factors that influence recruiters' choice of business schools and their MBA graduates: Evidence and implications for business schools. *Academy of Management Learning and Education, 6*(2), 217–233.

Safón, V. (2009). Measuring the reputation of top US business schools: A MIMIC modeling approach. *Corporate Reputation Review, 12*(3), 204–228.

Safón, V. (2012). Can the reputation of an established business school change? *Management in Education, 26*(4), 169–180.

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.

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

Shehatta, I., & Mahmood, K. (2016). Correlation among top 100 universities in the major six global rankings: Policy implications. *Scientometrics, 109*(2), 1231–1254.

Shin, J. C. (2011). Chapter 2. Organizational effectiveness and university rankings. In J. C. Shin, R. K. Toutkoushian, & U. Teichler (Eds.), *University rankings the changing academy—The changing academic profession in international comparative perspective 3.* Berlin: Springer.

Smyth, D. J. (1999). The determinants of the reputations of economics departments: Pages published, citations and the Andy Rooney effect. *American Economist, 43*(2), 49–58.

Sorz, J., Wallner, B., Seidler, H., & Fieder, M. (2015). Inconsistent year-to-year fluctuations limit the conclusiveness of global higher education rankings for university management. *PeerJ, 3*(e1217), 1–14.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*(3), 355–374.

Stake, J. E. (2006). The interplay between law school rankings, reputations, and resource allocation: Ways rankings mislead. *Indiana Law Journal, 81,* 229–270.

Stella, A. & Woodhouse, D. (2006). *Ranking of higher education institutions*. AUQA Occasional Publications Number 6. Australian Universities Quality Agency.

Sweitzer, K., & Volkwein, F. (2009). Prestige among graduate and professional schools: Comparing the U. S. news' graduate school reputation ratings between disciplines. *Research in Higher Education, 50*(8), 812–836.

Toutkoushian, R. K., & Webber, K. (2011). Chapter 7. Measuring the research performance of postsecondary institutions. In J. C. Shin, R. K. Toutkoushian, & U. Teichler (Eds.), *University rankings, the changing academy—The changing academic profession in international comparative perspective 3*. Berlin: Springer.

Van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 133–143.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E., Tijssen, R., Eck, N., et al. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419–2432.

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics, 103*(2), 555–564.

Wedlin, L. (2007). The role of rankings in codifying a business school template: Classifications, diffusion and mediated isomorphism in organizational fields. *European Management Review, 4*(1), 24–39.

Yotopoulos, P. A. (1961). Institutional affiliation of the contributors to three professional journals. *American Economic Review, 51,* 665–670.