The Three Global Rankings of Universities. The why, the how, and the what are they informing?

Juan Dip[1] and Gustavo Ferro[2]

Abstract

Several rating agencies evaluate universities, give scores, and offer rankings. Three global indices have the most reach: ARWU, QS, and THE. They measure the same phenomenon, with different methods, variables, and weights; however, their results are correlated. We aim to answer two questions: Which indicators most strongly explain the variance in global rankings, and what are their implicit importance weights? In other words, our first concern is to explore the "as if" aspect of the story: what are their implicit or underlying weights? As a second concern, we evaluate the reproducibility of these rankings using publicly available data, employing Benefit-of-the-Doubt (BoD) models based on Data Envelopment Analysis (DEA), including weight restrictions and a multi-directional extension (MDir_RBoD) to measure indicator-specific efficiency. Analyzing datasets for 2024 reveals that ARWU emphasizes publications, THE highlights industry impact and research quality, while QS focuses on internationalization and sustainability. Unlike prior studies that broadly assessed overall performance, our method offers detailed diagnostics of each indicator's contribution and areas for improvement. Using the MDir_RBoD model, we provide practical benchmarks for institutions and policymakers, enhancing the understanding of efficiency in university ranking strategies.

Keywords: University Rankings, Benefit of the Doubt, Efficiency, Indicator Improvements.

JEL Codes: I23, C02

Introduction

Universities produce teaching, codified research, and "third mission" (extension) outcomes. In doing so, they use human capital, physical capital (buildings, labs, libraries), "raw materials" (students whose minds are enriched by knowledge and experiences), administrative services, and intangible inputs such as a rich history of attainments, public recognition to the collective or the individual efforts of their professors or scientists, prestige and social and political influence they enjoy, etc. Besides teaching performance and the quantity and quality of research, the attributes considered could be awards, the degree of internationalization, graduate employment and achievements, productive sector linkages, and others (Ferro & D'Elia, 2020).

Most notably, global rankings such as the Academic Ranking of World Universities (ARWU), Quacquarelli Symonds (QS), and Times Higher Education (THE) combine, with different weights, universities' three main activities into composite indicators, thereby significantly influencing their reputation, student choices, and policy development (including budgetary decisions). However, these rankings are criticized for lacking transparency, arbitrary weighting, and/or biases favoring research-oriented institutions (Hamann & Ringel, 2023). Even when rankings had received multiple criticisms, as Hazelkorn & Mihut (2021) highlight, "*they reaffirmed a common-sense understanding about universities and their international hierarchical positioning, which impacted their respective countries' geopolitical standing*[3]". *"…(P)art of the appeal of rankings is*

[1] Universidad Nacional de Misiones (UNaM).Facultad de Ciencias Económicas. Email: juan.dip@fce.unam.edu.ar, Av. Fernando Elías Llamosas 9458 Posadas (CP N3304) Misiones –Tel: +54 (376) 4480006

[2] Universidad del CEMA (UCEMA) and CONICET. Email: gferro05@yaho.com.ar; gaf97@ucema.edu.ar. Avenida Córdoba 374, (1044) Buenos Aires (Argentina), phone +54-11-63143000, extension 780.

[3] Chankseliani (2023) explores geopolitical biases in impact among the most productive communication scholars from 11 countries and three world regions. The results indicate a strong US dominance in citation-based impact.

that they reinforce narratives of long-standing hierarchies within universities and media rather than challenging them." (Stack, 2016).

As in other service sector production functions, better outcomes are achieved when more and higher-quality inputs are used. However, not all institutions transform inputs into outputs equally, due to efficiency differences and random events that can affect performance. The outcomes of universities are degrees attained by students, codified knowledge in the form of research publications and patents for inventions, technology transfer to industry -when knowledge is transferred to innovation, that is, new or better products or services obtained thanks to applying abstract knowledge to concrete problem solving-, and spillovers to communities (cultural, social or political externalities, by definition, hard to measure).

Conversely, as in other service sectors, the quality of the outcomes, whether real or perceived, can vary significantly. The quality of outcomes and incomes partly determines the institution's overall quality. The institution's quality, in turn, determines its ability to attract talented students, outstanding professors, and researchers, and these factors feed back into the institution's overall quality.

Measuring quality is not straightforward in specific productions; sometimes, a proxy is needed. As Beath and Katsoulacos (1991) pointed out, *"…what differentiates products are the characteristics that they each possess… product differentiation involves making a… product either really or apparently different from that of its rival*". "*Rankings are associated with branding (sic), reputation management, and greater government and industry funding*" (Stack, 2016). In higher education, quality is a complex measure vulnerable to external influences, with substantial lags between academic quality changes and reputation changes. In some cases, the latter may be more closely tied to prestige than quality. Thus, "*rankings increasingly become reputation, rather than reputation being an independent indicator that rankings can use to assess changes in quality*" (Bowman & Bastedo, 2011).

One type of quality proxy is an independent analyst's score or an agency's rating. The agency would develop an algorithm to collapse information concerning cardinal attributes or Likert scales[4] into a score and put it in an ordinal ranking. Third parties' acceptance of the validity of the ranking implies that the quality proxy is accepted and influences talent allocation among institutions. As the state's role changed concerning post-WWII consensus, public involvement shifted to emphasize measuring, assessing, comparing, and benchmarking performance and productivity. This led to new forms of accountability focused on measurable outcomes, increasingly tied to funding through performance-based funding or performance agreements. Global rankings are a product of an increasingly globalized economy and an internationalized higher education landscape. Quality control and accountability are the primary purposes (Hazelkorn & Gibson, 2017).

According to Hazelkorn (2015), the literature on rankings can be roughly divided into methodological concerns and theoretical understanding. This paper is concerned with the first one, and we summarize only the theoretical discussions about the ranking impact and implications on universities and stakeholders. University rankings are contested since the weights given to each variable are debatable. Additionally, as with other third-party performance measures, some critics question the objectivity of the rankings or the construction of a metrics-based reality, or a "metricality," a logic that could eventually impregnate or distort academic life (Giacalone, 2009). Critics sometimes raise concerns about potential geographical, linguistic, or other biases that may favor some universities (or scientific disciplines) over others

[4] To measure attitudes, opinions, or perceptions, before a statement, respondents are presented with some issues. They can agree or disagree by answering using an ordinal scale that denotes the intensity of their agreement or disagreement. Respondents select the option that best fits their opinion. The scales provide cardinal values of ordinal opinions.

(Barsky, 2012). We do not discuss the latter, as our concern is technical (i.e., knowing the "how"). As analysts, our purpose is to decode and reconstruct rankings.

Researchers have primarily examined ranking methodology in two ways: first, *whether it is valid* for evaluating and comparing institutional quality, and second, *what the numbers represent*, particularly whether they reflect meaningful differences between institutions or merely statistical error. The relationship between rankings, resources, and reputation is sometimes autocorrelated (Kim & Bastedo, 2021). This paper is concerned with what the numbers represent within the methodological discussion.

There are regional and subject rankings. Sometimes, they are published in clusters: engineering, biomedicine, natural sciences, social sciences, and the arts and humanities. The most extensive global rankings analyzed in this paper measure the same phenomenon using different methods, variables, and weights; however, their results are correlated. We examine the three most popular rankings of 2024 (THE, n = 1,904; ARWU, n = 989; and QS, n = 602) and identify the common elements (N = 361). Considering the universities that appear simultaneously in all three rankings (n = 361), the correlation is 0.788 between THE and QS, 0.852 between THE and ARWU, and 0.704 between QS and ARWU.

Previous research on ranking methodology and the interpretation of numbers primarily focused on comparing weights, running correlations across rankings, or tracing overall performance patterns. Although these approaches are valuable, they seldom investigate indicator-level efficiency, which measures how institutions perform on each metric compared to a best-practice frontier. This micro-level perspective is vital for practical benchmarking, as aggregate scores often obscure the specific indicators contributing to inefficiency.

This paper addresses two central questions:

1. Which indicators most strongly explain the variance in global rankings, and what are their implicit importance weights?

2. To what extent can these rankings be replicated using publicly available indicators, and how can efficiency analysis identify improvement opportunities for individual universities and at the level of single indicators?

We combine Benefit-of-the-Doubt (BoD) models with weight restrictions (BoDCONST) to derive efficiency-based importance weights and extend this analysis using the Multi-directional Robust BoD (MDir_RBoD) model. This robust, multi-directional framework quantifies indicator-specific efficiency gaps and prescribes directional improvements for each university, providing granular insights beyond traditional composite analyses.

Analysis of the 2024 datasets from ARWU, QS, and THE reveals consistent discrepancies between official and derived weights: in ARWU, productivity metrics such as publications constitute the primary indicators of efficiency; in THE, industry impact is the most influential factor, followed by research quality; and in QS, internationalization metrics, alongside sustainability, are significant. Furthermore, our methodology visualizes efficiency at the indicator level, providing detailed insights valuable for institutional benchmarking and policy formulation.

The paper makes contributions to the literature on university rankings. While prior research has mainly explored global university ranking methods, few studies have analyzed the efficiency of individual indicators within a comprehensive framework. This paper addresses that gap by using the Multi-directional Robust Benefit-of-the-Doubt (MDir_RBoD) model, which allows us to evaluate not only the overall efficiency of universities but also the specific contributions and potential for improvement of individual indicators, such as teaching, research quality, and international outlook. This approach provides a detailed diagnostic tool, enabling institutions to identify which aspects affect their current ranking and how much each indicator needs to

improve to reach the efficiency frontier. These insights extend beyond traditional composite scores, providing actionable benchmarks for policymakers and university administrators, and significantly contributing to the literature on higher education rankings.

Following this introduction, Section 2 reviews the literature on university rankings. Section 3 outlines methodology, data, and models for calculating importance weights. Section 4 presents the analysis results. Section 5 discusses the study's achievements and limitations. Section 6 concludes and suggests future research.

1. Literature review

The first global ranking of universities (The Academic Ranking of World Universities, ARWU) was developed in 2003 by Shanghai Jiao Tong University, China, and is updated annually. Since 2009, the ARWU has been published and copyrighted by Shanghai Ranking Consultancy. ARWU was followed by the joint venture "Times Higher Education-Quacquarelli Symonds" World University Rankings (or THE-QS) in 2004, which, in turn, split into two rankings after 2009. The development of the QS-THE initiative follows the Review of Business-University Collaboration, commissioned by the UK government and executed by Lambert's commission. It focused on encouraging universities to promote more direct industry links, partly through "*… the development of a league table (ranking) of the world's best research-intensive universities*." QS collected data, and THE was responsible for the structure and commentary on the rankings. At the time, QS was recognized for its student placement services and recruitment, primarily for MBA programs. In contrast, Times Higher Education was known for reporting on educational issues and its connections to policymakers and educators, primarily in the UK. In 2010, QS and THE started their respective rankings, the latter, in partnership with Thomson Reuters. ARWU, QS, and THE are sometimes called "The Big Three" (Stack, 2016). Most of the top ten globally ranked institutions are in southern England, California, and the northeastern American states of New York, New Jersey, Connecticut, and Massachusetts.

"*The design of the indicator system reflects the rankers' conception of the universities and the original purpose of ranking… ARWU originated from identifying world-class universities and finding the gap between Chinese universities and world-class universities, particularly in terms of academic and research performance. In addition, the availability of internationally comparable data was considered*" (Wu & Liu, 2018). Instead, the core mission philosophy of QS ranking is "*to enable motivated people around the world to fulfil their potential by fostering international mobility, educational achievement and career development… A student-centered approach … it seeks to address and inform"*. It is "*a student-oriented ranking organization*" (Sowter et al., 2018). THE ranking focuses on research-intensive universities with a broad curriculum and an undergraduate teaching role, limiting the scope to institutions with more than 1,000 qualifying publications over 5 years, undergraduate teaching, and a broad curriculum *"… generating a balanced scorecard that… gives an overall estimation of the performance of universities.*" (Ross, 2018).

In the ARWU ranking, data are primarily generated externally, and only the full-time equivalent faculty is somewhat manipulable by universities. QS based its information on scientific publications associated with the Scopus database, part of the Elsevier editorial group. At the same time, THE ranking was associated with Thomson Reuters as the primary source of academic statistics; nowadays, it is associated with Clarivate Analytics (Barsky, 2012). "*The competition between ranking agencies has helped each to develop a distinguishing factor of its ranking… (QS)… (*specialized*)… in graduate employability. It includes employer reputation in its… (*ranking*)…, and… has a non-research-based Graduate Employability Ranking… (THE)…(*defines*)… the university–industry relationship, as measured by industry income, as a distinguishing feature of its ranking… (ARWU) has maintained its Scientometric-only construction since its inception."*

(Cantú-Ortiz and Fangmeyer Jr., 2018). The ranking organizations were developed under different contexts and have varying strategic goals and primary audiences: THE was developed by a media company, QS by a company specializing in study-abroad networks and business education, and ARWU by a university (Stack, 2016). The QS is particularly entrepreneurial (Stack, 2021).

The Chinese government started its strategic policy of building world-class universities with the "985 Project". Firstly, nine top Chinese universities were selected for that end (Peking, Tsinghua, Shanghai Jiao Tong -SJTU-, Fudan, Zhejiang, Nanjing, University of Science and Technology of China, Xi'an Jiao Tong, and Harbin Institute of Technology). SJTU assigned its Office of Strategic Planning to make the strategic plan. This should answer several questions: How do you define a "world-class university"? How many world-class universities exist worldwide? What are the positions of the top Chinese universities? How can top Chinese universities reduce their gap with world-class universities? From 1999 to 2001, SJTU benchmarked top Chinese universities with four groups of U.S. universities. The results of those quantitative and qualitative analyses were used in the planning process. After another two years of benchmarking work, the first ARWU ranking was released in 2003 (Wu & Liu, 2018).

The rise of rankings coincides with an explosion in academic production and its spread from a small set of developed countries to the entire world. According to the World Bank, the volume of global publications grew at a rate of 4.95 percent yearly between 2000 and 2018, while the world's GDP increased at 2.5 percent per annum. The group of countries leading in science production has diversified. In 1980, world science was almost a duopoly between the US and Western Europe, especially the UK. The Soviet science system had only episodic collaboration with the Western World. More than forty years later, science was highly connected and liked among scientists, institutions, and countries. Codified science outputs exploded after 1990, because of the concurrent phenomenon of the internet surging and the end of the Cold War. Production became more diversified by country of origin. China's share of global science rose from 5 to almost 21 percent between 2000 and 2018, when resources devoted to science multiplied by more than 10 times, while in the US and the EU, they doubled. In the pre-Internet world, Continental European Countries, the English-speaking ones, and Japan were responsible for almost 90 percent of total production.

Researchers' training has also diversified, as China and India are now the two major producers of doctoral graduates. Cross-border collaboration has increased, including co-authorship and citation, which are primarily global (Marginson, 2021; Chankseliani, 2023). Marhl et al. (2025) find significant shifts in the global academic landscape in Engineering, Life Sciences, Medical Sciences, Natural Sciences, and Social Sciences. The changes among the four primary regions under consideration affect the relative positions of China, Europe, the USA, and the rest of the world. Notably, the proportion of Chinese universities in this elite group has escalated from 11.5% in 2017 to 20.2% in 2023. In contrast, Europe has experienced a decline from 34.5% in 2017 to 30.4% in 2023. The USA's share has decreased from 26.8% in 2017 to 22.6% in 2023. The rest of the world's share has remained relatively stable, around 27%.

The ARWU surge coincides with China's explosive growth and reflects the country's aspiration to compete with other powers in the knowledge arena. ARWU was developed as a strategic tool to compare the position of Chinese universities with established world-class universities in developed countries, and to lobby for funding. ARWU highlighted the gap between Chinese universities and those in other countries and helped identify the criteria for catch-up (Hazelkorn & Mihut, 2021). The ARWU rankings are based on Euro-American-centric data, the universities' research publications in English, scientific recognition in the form of valued awards, and publications in STEM disciplines. Just in 2013, publications from the Social Sciences Citation Index were added (Stack, 2016). Chinese universities are publicly regulated and rely on their ranking and research assessments to be eligible for public funding. Due to competitive pressure

and high rewards, Chinese researchers have been encouraged to publish in high-impact journals. In the last two decades, the journal impact factor (JIF) has helped catapult Chinese scientific publication output to the highest in the world. However, quality levels have been denounced as insufficient, with events of fraud and "paper mills". Other undesirable practices have been identified, including citation cartels ("citation bartering"), coercive citations, citation stacking, and AI-generated or AI-rewritten texts. Wang et al. (2024) provide three arguments to explain these practices: institutional evaluation regimes (incentives), differing epistemic cultures (tolerance for certain practices), and the influence of the commercial publishing industry (the cost of publishing in specific journals to achieve impact).

ARWU is the only ranking based almost exclusively on objective attributes (see next section for adjustments made in each ranking across time). While some rankings use only hard data, others rely on a combination of hard and soft data. Soft data are qualitative and depend on ideas, knowledge, experience, and opinions. Hard data is suitable for obtaining more objective results, but can be limited to specific criteria, universities, and regions (Olcay & Bulu, 2016). For instance, bibliometric databases have excluded some scientific output not published in peer-reviewed English-language journals. The indicators used in ARWU are less debatable than survey data, which can be affected by the halo[5] and leniency[6] effects. To avoid manipulation, no data is collected directly from the ranked universities. Every university can verify its position and compare indicator scores with peer institutions. There have been no significant changes to the ARWU methodology since 2004, enabling comparisons across years (Wu & Liu, 2018). However, since 2015, ARWU has changed approximately every two years.

The Matthew effect, Matthew principle, or cumulative advantage, called by the biblical parable of talents, denotes a tendency to accrue success in direct proportion to one's initial level (Merton, 1968; Merton, 1988). In university rankings, there is some inertia in the best-positioned institutions because they were the canon on which subsequent rankings were based. "*A high ranking means more opportunities for faculty and students – privilege tends to beget more privilege*." (Stack, 2016). Anchoring theory suggests that rankings may influence institutional reputation, and this effect may be powerful when new rankings are introduced. Bowman and Bastedo (2011) find that anchoring effects substantially influence future reputational assessments.

QS includes reputational measures, based on subjective opinions of experts or prospective employers that the consultancy firm picks. The data used to compile the world university rankings are sourced from the QS Intelligence Unit's data acquisition team, the Scopus database, and the QS Academic and Employer Surveys. Universities feed the full-time equivalent number of students, faculty members, international students, and international faculty (Sowter et al., 2018). This ranking has been revised over time (see next section).

The ranking attempts to encompass more university functions beyond research. THE invites universities to submit data overall and in eight broad subject areas. The second data source is an annual Academic Reputation Survey, which randomly samples academics from the Scopus database, ensuring a balanced representation by country. The final dataset is the Scopus database, extracting information on citations, academic productivity, and the degree of international collaboration. Many variables are represented as ratios to the student body or faculty size.

[5] A cognitive error consistent in formulating a positive first impression of a person, company, country, brand, or product, influenced by previous judgments of status, performance, or personality, not resorting to all the evidence at hand. Instead, opinions can rest on positive or negative prejudices, ideology, or social perceptions.

[6] A tendency to evaluate all ratings as probably higher than deserved, resulting in inflated ratings rather than accurate performance assessments.

ARWU ranking exhibits more permanence and consistency, since its weights have undergone the fewest changes since its first edition. QS includes 1,000 universities out of 20,000 worldwide, while ARWU contains 2,000 (Altbach, 2021). Also, in QS and THE rankings, some amendments were made at the time, changing the weights repeatedly. QS, especially, and THE ranking have also been deemed a plebiscitary or survey-based system. Nevertheless, both rankings emphasize that the information is surveyed from qualified peers with balanced origins (Sowter et al., 2018, and Ross, 2018).

ARWU's model is inspired by the investigation of universities in the US, where a significant proportion of the new generation of Chinese scientists studied. They are fewer than 150 out of more than 4000 American universities, where most knowledge advances are produced (Barsky, 2012)[7]. As the first movers, they set the canonical characteristics of high-end universities, which were later employed as the international quality measure. These rankings can also be a useful accountability tool[8]. Rankings are helpful: stakeholders seek to learn about relevant or more advantageous options. They offered to guide students, parents, governments, and other stakeholders in attracting talent to the faculty and classroom levels (Hazelkorn, 2015). Likewise, institutions want to benchmark themselves against peers and compete for faculty and talented students; rankings provide a helpful guide. They are also tools for selecting places to study and work. Policymakers can use rankings to allocate budgets or regulate the industry. Internationalization contributed to the popularity of rankings, since over 5 million people study abroad (Altbach, 2021). Rankings interest stakeholders because they show information in a comprehensible, simple, and synthetic way (Malea & Perez-Esparrels, 2021).

Locke (2021) conceptualizes six main ways in which different types of universities are influenced and react to rankings: 1) They are used for strategic positioning, marketing, and setting goals; 2) They are employed to redefine activities and alter perceptions; 3) They show evolving responses, from initial rejection to efforts to produce desired outcomes or to use good positioning in rankings; 4) They have an impact on staff morale and status; 5) They leveraged internal change, promoting accountability and competition; 6) They generate attempts to manage their influence.

However, they are not absent from controversy since they "*make perceptions of prestige and quality explicit*", setting the norm to which other universities are compared. "*Mediatization plays an active role in constructing discourses about what stands for a "good", "excellent" and "bad" university… Universities often use media logic to "brand" themselves... If universities do participate in HEI rankings, they are reinforcing a system based on intersecting media, commercial, and political logics and practices that have little to do with academic norms of research and inquiry; if they do not participate, they risk losing public funding, students, and donors in an increasingly competitive and globalized environment*." (Stack, 2016). Since access to institutional data is not straightforward, and there are issues of information heterogeneity, global rankings have primarily focused on bibliometric and citation data.

Other countries and regions have promoted alternative rankings with different weights to improve their position, a response that can be perceived as protective or reactive. Other rankings include, in a non-exhaustive list, Webometrics (Spanish National Research Council, Spain), National Taiwan University Rankings (formerly Performance Ranking of Scientific Papers

[7] There are different types of universities, differing in objectives (teaching-oriented, research-oriented, industry-oriented), degrees offered (all degrees, only graduate degrees, only postgraduate degrees), and disciplines to study (all fields, humanistic, technical) (Malea & Perez-Esparrels, 2021).

[8] Other transparency, accountability, and comparability instruments include Accreditation (as an institution certified to award qualifications), Assessment, Quality Assurance (QA), and Evaluation (concerned with practices and processes), Benchmarking (systematic performance peer comparison), Classification and Profiling (according to mission and type), College Guides, and Social Networking (descriptive information to stakeholders).

for Research Universities, HEEACT), Leiden Ranking (Centre for Science & Technology Studies, University of Leiden), SCImago Journal and Country Rank (SJR) (Spain), University Ranking by Academic Performance (URAP) (Informatics Institute of Middle East Technical University, Turkey), U-Multirank (European Commission, Brussels), and Best Global Universities Ranking (USNWR, US), etc.

Only U-Multirank, developed by the European Union, differs in vision and methodology from the three most popular rankings (Hazelkorn & Mihut, 2021). Prado (2023) examines the U-Multirank initiative, which begins by disregarding the idea that the best universities can be defined in a general sense. Instead, the initiative's best university depends on each student's personal choice, and the ranking can be elaborated according to the interests and constraints of each stakeholder. The initiative evaluated educational performance in five dimensions: teaching, research, knowledge transfer, internationalization, and regional involvement. The stakeholder can evaluate each dimension separately or group by sub-items, according to their interests. Each category can contain between 4 and 11 indicators. Data is obtained from sources common to all universities, reported by third parties, and directly from universities. The latter information is less precise, more subject to manipulation, and often unavailable. As an advantage, it is not based on subjective opinions. Since many indicators can cast doubt on the possibility of correlation between variables, Pardo (2023) finds a 35% positive correlation in his research sample. U-Multirank faced trouble generating timely and homogeneous information (Holmes, 2021).

Other rankings consider universities' unique roles in different cultures, such as Russia, India, and Latin America (Hazelkorn, 2015). Russian and former Soviet universities ranked low, prompting Lomonosov Moscow University to create the Reitor Ranking, which includes criteria like patents, textbooks, faculty honors, and memberships (Peümop, Barsky, 2012). National research evaluations and global rankings differ significantly in data collection, usage, interpretation, and format. They may use similar bibliometric sources, but in contrast to global rankings, national rankings, like national research evaluations, may be able to generate and have access to significant nationally specific data, which can involve peer review as an element of the evaluation in terms of review panels, etc. (Hazelkorn & Gibson, 2017). The rankings focus on research productivity, using data collected by firms acting as collectors and aggregators, which account for peer-reviewed publications and impact measured by citations (Altbach, 2021). There are two public databases, Scopus (Elsevier) and the Web of Science (Clarivate Analytics), providing raw data on research for rankings. Google Scholar Citations has not been used yet. All rankings that measure research include data on publications, except for QS. Citations serve as proxies for research impact in the three rankings. Teaching and learning in rankings are still indirect and limited. As proxies, rankings used faculty-to-student ratio, degree of internationalization (deemed correlated positively with quality; it tends to be negatively correlated with the population of the recipient country), income sources (other than tuition), or surveys collecting opinions of employers. Patent counts are not used directly (THE use them indirectly by counting income from industry) (Holmes, 2021). Previous experiences in the pre-global ranking era did not rank schools but situated them across categories according to their missions. For instance, that was the policy of the Carnegie Classification of Institutions of Higher Education in the US between 1970 and 2005. This avoided the zero-sum criteria that ranking involves (Altbach, 2021).

Besides the differences in attributes, the universities at the top of the rankings (of which the top 100 represent roughly 0.5 percent of the world’s universities and 0.4 percent of university students) are highly correlated. Given the time lag between investments, achievements, and prestige building, they have permanence. They are established institutions in the United States, Western European countries, and other developed countries in Europe, Oceania, and East Asia, with abundant resources, including tuition and government funding, private donations, and

research grants. The best-ranked universities have English-speaking faculty and students, are highlighted in science for attracting external funding, publish their results in internationally rated journals, and offer graduate programs. In undergraduate studies, faculty tasks of research and teaching are more substitutive than complementary, while in graduate studies, human capital development and knowledge production are complementary rather than competitive. Thus, the international rankings weigh research more heavily than teaching, and postgraduate students more than undergraduate students. They are also the most internationalized since international student flows are composed chiefly of Asians and other nationals from developing countries to North America, Europe, and Oceania destinations (Hazelkorn, 2015).

The concept of a "Third mission" is challenging to proxy, and attempts have been made to measure it through variables such as gender equity, sustainability, open-access publishing, or online education (Holmes, 2021). Teaching was almost ignored in rankings until recently. The criteria were included in THE and QS rankings through proxies, which do not measure teaching or assess quality or impact (Altbach, 2021).

According to Li et al. (2011), as cited by Hazelkorn (2015), a significant amount of cross-country variation in university performance can be attributed to income, population size, research and development spending, and the national language. The rankings consider research outputs, privileged STEM disciplines over social sciences, arts, and humanities, and focus on published papers over books or other codified knowledge outputs. Accounting is also heavily concentrated on English-language publications (Altbach, 2021). Open-source publications and grey literature are commonly excluded from rankings. They are more likely to be regionally impactful, written in languages other than English, and are more frequent in the humanities and social sciences (Hazelkorn & Mihut, 2021).

Kim and Bastedo (2021) document studies that tested the relationships between rankings, trying to discern whether they provide good proxies for institutional quality. Aguillo et al. (2010) classified existing international rankings as "*student-oriented rankings*", including THE-QS, and "*research-oriented rankings*", encompassing ARWU. After the two partners split, QS can be classified as "*student-oriented*," while THE is a mixture of both. This study utilized statistics to calculate the size of the overlap between different rankings, finding similarities between the rankings that focus on different aspects (student and research). The size and input resources from the previous year positively and significantly predict ARWU raw scores in the following year. Subjective factors, particularly reputation surveys, highly influence ranking measurements (Aguillo et al., 2010). Other authors simulated ARWU, excluding awards, and found that when the simulated ranking results were compared for the top 20 universities, they were more stable than the original (Kim & Bastedo, 2021).

Citations measure academic performance, counting for "impact." An article's scientific quality can be captured by its relevance, evidence, novelty, design, and method. Nevertheless, factors such as the length of the abstract, the number of references, the prestige or gender of the author, nationality, university of origin, etc., could influence impact (Mammola et al., 2022).

Finally, the Benefit of the doubt (BoD method, described in the next section) is applied to determine weights in university performance research or teaching areas (De Witte & Hudrlikova, 2013; Szuwarzyński, 2019; El Gibari et al., 2022; Jain & Gulati, 2024) or when utilizing university ranking data (Johnes, 2018; Dip, 2021; Matulová, 2023). However, in contrast to previous studies that broadly assess overall performance, our efficiency-based methods offer detailed diagnostics of each indicator's contribution, its efficiency score, and areas for improvement, utilizing available data obtained from university ranking websites.

2. Method and data

3.1 Method

We work with ARWU, THE, and the QS global indices. We propose three methodologies based on efficiency criteria to elucidate the relative importance of the different weights using Data Envelopment Analysis (DEA). DEA is a nonparametric technique used to evaluate the relative efficiency of comparable "decision-making units" or DMUs—in this case, universities—by simultaneously considering multiple inputs and outputs.

As formalized by Cherchye et al. (2007), the BoD (Benefit of Doubt) approach uses the DEA framework to construct composite indices (e.g., rankings) when arbitrary weighting of the different indicators is not desired. The model maximizes observed performance based on the outputs (desirable indicators, such as publications, citations, internationalization, etc.). Inputs remain fixed and equal to 1, focusing the evaluation on each unit's (university's) capacity to "produce" outputs given those same fictitious resources. Thus, each university can assign weights (relative importance) to the different outputs that make it appear most efficient, provided those weights are also feasible for all other units. The model seeks the set of weights that maximizes a university's efficiency, with no other unit achieving an efficiency greater than 1 using those same weights.

Mathematically, the I_{c0}, (the composite performance indicator) is obtained by solving the following optimization problem:

$$I_{c0} = \mathrm{Max}_{w_{0,i}} \sum_{i=1}^{m} w_{0,i} y_{0,i}$$

$$\text{s.t.} \sum_{i=1}^{m} w_{0,i} y_{j,i} \leq 1 \quad \text{con } j = 1, \dots, n; \quad w_{0,i} \geq 0, \quad \text{con } i = 1, \dots \dots, m \quad (1)$$

Where $w_{0,i}$ is the weight of the individual indicator *i* of each DMU; $y_{0,i}$ is the value of each indicator *i* for the DMU$_{0;}$ *n* is the number of DMUs and *m* is the number of individual indicators. Equation (1) is solved by maximizing *n* times for each DMU to obtain a set of I_{c0} for the whole sample. We call this model BoD. The optimal solution of equation (1) gives us a composite indicator representing a measure of efficiency for each DMU, where $0 \leq I_{c0} \leq 1$. A value of 0 indicates the worst performance of the DMU compared to the other DMUs in the sample; conversely, a value of 1 indicates that the DMU has the best performance.

If each optimal weight is multiplied by the value of each indicator, we obtain the "Pie Shares", ($w_{0,i} y_{j,i}$), which indicates how much the dimension "i" contributes to the overall composite score of each university. If one were to execute model (1) in its current state, one might anticipate the occurrence of the zero-weight phenomenon with considerable frequency. To avoid this problem, we propose placing restrictions on these "Pie Shares", following Van Puyenbroeck & Rogge (2017). They propose constructing multiplicative CIs as geometric mean quantity index numbers. It introduces an indirect, two-step procedure to combine multiplicative aggregation with BoD weighting. The foundation is the geometric mean quantity index, which aggregates ratios of sub-indicators (DMU values relative to a reference standard, y_{ri}/y_{rB}) using "importance exponents or weights" (ω_{ri}^{*}), where y_{rB} are values of the base performance indicator and the respective sub-indicator values y_{ri} with r = 1,…,s.

Thus, the $CI_i = \prod\left[(y_{ri}/y_{rB})^{\omega_{ri}^{*}}\right]$ (2).

We start with a Linear BoD Model (1) to derive the weights: a standard linear BoD model is solved for each university. This identifies the optimal shadow prices that maximize the CI value for the evaluated unit, subject to an upper bound (typically 100%). Secondly, to calculate BoD Sub-indicator Shares ("Budget Shares") (ω_{ri}^{*}), we calculate them from the optimal weights (w_{ri})

values obtained in the first step and complete the analogy with economic quantity indices, where budget shares are used as weights:

$$\omega_{ri}^{*} = (w_{ri}^{*} * y_{ri})/\Sigma(w_{ri}^{*} * y_{ri})(3).$$

The sum of these (ω_{ri}^{*}) for each university i is equal to 1. The model can also incorporate weight restrictions for specific sub-indicators (e.g., $0.005 \leq \omega_{ri}^{*}$), to avoid the zero-weight problem. This model, with restrictions, is called BoDCONST.

Finally, to evaluate the presence of outliers and determine how far universities are from their target values, we follow Vidoli et al (2024). They introduce the Multi-directional Robust Benefit-of-the-Doubt (MDir_RBoD) model. Its primary objective is to provide a composite quality measure resilient to outliers and identify potential improvement directions for each component indicator. MDir_BoD seeks the maximum possible increase in a single indicator (y_{ri}) required to reach the efficiency frontier, while keeping other indicators fixed. Additionally, we obtain the directional vector, which signifies the disparity between the ideal (frontier) value and the observed value for each indicator. This metric measures the proportion by which each simple indicator must be increased to reach the frontier. Thus, the MDir_RBoD model offers a weighting scheme that favors each unit and "most favors the desirable direction" by identifying potential improvements, rather than simply following past production or a specific direction (e.g., input- or output-oriented). Moreover, it enhances this approach by comparing it with a robust frontier resilient to outliers and abnormal units. We calculate two indicators: 1) the proportion by which each of the simple indicators must be increased to reach the frontier $\beta \in [0,1]$ and 2) the relative robust multi-directional scores vector for the simple indicator e_q,

$$e_q = \frac{\mathbf{y}_q}{\mathbf{y}_q + \beta^{*} \mathbf{g}_q^{PI}} \quad (4)$$

where y_q is the simple indicator and g_q^{PI} is the vector of units specific q directions of the simple indicator q. (See for details, Vidoli et al., 2024). We call this BoDRMDIR.

Some methodological adjustments are necessary to ensure the reliability and consistency of efficiency measurements to apply Benefit-of-the-Doubt (BoD) models to the THE, QS, and ARWU rankings. Notably, when working with THE ranking, any missing values are filled in with a small positive number, following the DEA assumption that all data points must be positive. Once the missing values are addressed, each indicator is rescaled to restore its original range of 0 to 100, thereby maintaining the internal consistency of the ranking.

It is important to note that the approach proposed by Cherchye et al. (2007) does not require normalization, as it lets each university choose weights freely to maximize its efficiency score. In contrast, the model developed by Van Puyenbroeck & Rogge (2017) does benefit from normalization. This is because they restrict the possible weights or use methods that compare the direction of improvement across indicators.

Finally, the rankings from each BoD model are created, considering ties in efficiency scores. These adjusted rankings are then used to calculate the relevant correlation coefficients, comparing the consistency between different methodological approaches.

3.2 Data

The paper draws upon publicly accessible data from the 2024 editions of the three prominent global university rankings: ARWU, THE, and QS[9]. THE ranking introduced substantial changes in

[9] As the database for ARWU only computes the score and assigns a position in the ranking to the first 100 universities, to replicate the ranking for the following 900 universities (they have only attributed a position), the methodology was applied using the individual data of each university. The replication procedure yielded negligible differences (the correlation for the first 100 universities between the original ranking and the recalculated one is 0.9999). We did the same with the other two rankings to provide a

2015 by migrating to Scopus as a data source, followed by adjustments in 2019-2023 that reduced the weighting of reputation and incorporated metrics for open research and socioeconomic impact. Meanwhile, ARWU's ranking maintained its traditional focus on research productivity but added indicators of international collaboration and high-quality publications from 2015 to 2021. In contrast, the QS World University Rankings has made more frequent and disruptive changes, highlighting the inclusion of sustainability as an evaluative criterion (5% of the weighting) in 2024, aligning itself with global agendas such as the SDGs. These methodological differences underline the heterogeneity in the conceptualization of university quality: while THE and ARWU prioritize bibliometric indicators, QS is based on surveys and incorporates emerging social dimensions. Thus, introducing key methodological changes in 2024 as a reference year is justified, mainly in the QS World University Rankings.

ARWU and QS exclusively furnish individual performance metrics for a restricted number of institutions (e.g., the top 100 or 600 universities). Consequently, the remaining universities are categorized into position intervals (generally in bands of 50). Given this constraint, the rankings generated through the Benefit-of-the-Doubt (BoD) models, along with their respective correlation coefficients, are analyzed according to the following criteria: 1) the entire dataset of universities is used for estimating the importance weights and efficiency scores, ensuring that the BoD approach captures all relevant performance structures; however, 2) the comparative analysis of institutional positions focuses only on those universities for which detailed score information is publicly available in the original rankings. Specifically, the ARWU-based analysis includes the top 100 universities (for the following 900 only provides intervals); for QS, the full set of the top 602 institutions is used; and for THE, the dataset comprises 1,904 universities, based on the released score data.

The following tables present the descriptive statistics of the variables in each 2024 ranking (entire dataset).

Table 1: ARWU´s descriptive statistics.

No	Indicator	Stats / Values	Distinct Values	Valid	Missing
1	Alumni [numeric]	Mean (sd): 4.5 (10.1). min < med < max: 0 < 0 < 100. IQR (CV): 0 (2.2)	57	989 (100.0%)	0 (0.0%)
2	Award [numeric]	Mean (sd): 3.8 (11.5). min < med < max: 0 < 0 < 100. IQR (CV): 0 (3)	75	989 (100.0%)	0 (0.0%)
3	Highly Cited Publication [numeric]	Mean (sd): 10.8 (11). min < med < max: 0 < 9.2 < 100. IQR (CV): 14.6 (1)	59	989 (100.0%)	0 (0.0%)
4	Nature & Science [numeric]	Mean (sd): 10.6 (10.6). min < med < max: 0 < 7.3 < 100. IQR (CV): 9.4 (1)	259	989 (100.0%)	0 (0.0%)

standard rescaled score. The original databases had standardized values equal to zero (for example, if the university does not have any positive value of a variable) or NA for missing data. DEA-BoD does not allow working with zeroes or NA. To apply the technique, universities with missing data (NA) were eliminated, and the sample decreased, for example, for ARWU to 989 observations, THE to 1904, and QS to 602. The zeroes were rescaled employing a minimum absolute value different from zero and replacing zeroes with a percentage of that value. DEA-BoD was applied with the re-scaled database without weight constraints, yielding new weights. Using comparable metrics allows us to build the “intersection” set of 361 universities in the three rankings to examine correlations between them.

5	Publications [numeric]	Mean (sd): 35.2 (13.7). min < med < max: 9.3 < 31.8 < 100. IQR (CV): 15.7 (0.4)	414	989 (100.0%)	0 (0.0%)
6	Per Capita Academic Performance [numeric]	Mean (sd): 19.7 (8.8). min < med < max: 8.6 < 17.6 < 100. IQR (CV): 8.8 (0.4)	267	989 (100.0%)	0 (0.0%)

Source: Authors. ARWU (2024).

Table 2: THE´s descriptive statistics.

No	Indicator	Stats / Values	Distinct Values	Valid	Missing
1	Teaching [numeric]	Mean (sd): 29.1 (14). Min < Med < Max: 9.4 < 25.8 < 99. IQR (CV): 15.1 (0.5)	473	1904 (100.0%)	0 (0.0%)
2	Research Environment [numeric]	Mean (sd): 23.4 (16.7). Min < Med < Max: 4.6 < 17.2 < 100. IQR (CV): 18.1 (0.7)	506	1904 (100.0%)	0 (0.0%)
3	Research Quality [numeric]	Mean (sd): 52.2 (25.1). Min < Med < Max: 3.4 < 52.4 < 99.7, IQR (CV): 42.8 (0.5)	809	1904 (100.0%)	0 (0.0%)
4	Industry Impact [numeric]	Mean (sd): 47.1 (26.2). Min < Med < Max: 15.6 < 41 < 100. IQR (CV): 45.9 (0.6)	688	1904 (100.0%)	0 (0.0%)
5	International Outlook [numeric]	Mean (sd): 49.9 (21.8). Min < Med < Max: 16.1 < 45.3 < 98.8. IQR (CV): 33.4 (0.4)	725	1904 (100.0%)	0 (0.0%)

Source: Authors. THE (2024).

Table 3: QS´s Descriptive statistics**.**

No	Indicator	Stats / Values	Distinct Values	Valid	Missing
1	Academic Reputation [numeric]	Mean (sd): 37.5 (26). min < med < max: 2.8 < 28.8 < 100. IQR (CV): 34.7 (0.7)	410	602 (100.0%)	0 (0.0%)
2	Employer Reputation [numeric]	Mean (sd): 36.4 (28.7). min < med < max: 1.6 < 27.8 < 100. IQR (CV): 38.9 (0.8)	402	602 (100.0%)	0 (0.0%)
3	Faculty to Student [numeric]	Mean (sd): 40 (31.5). min < med < max: 1.7 < 31.3 < 100. IQR (CV): 52.9 (0.8)	417	602 (100.0%)	0 (0.0%)
4	Citation [numeric]	Mean (sd): 45.2 (31). min < med < max: 1 < 41.7 < 100. IQR (CV): 50.3 (0.7)	436	602 (100.0%)	0 (0.0%)
5	International Faculty [numeric]	Mean (sd): 51 (36). min < med < max: 1.3 < 46.5 < 100. IQR (CV): 75.8 (0.7)	376	602 (100.0%)	0 (0.0%)
6	International Students [numeric]	Mean (sd): 41.3 (34). min < med < max: 1.1 < 30.6 < 100. IQR (CV): 63.2 (0.8)	401	602 (100.0%)	0 (0.0%)

7	International Research Network [numeric]	Mean (sd): 47 (32.9). min < med < max: 1 < 46.1 < 100. IQR (CV): 62.9 (0.7)	417	602 (100.0%)	0 (0.0%)
8	Employer [numeric]	Mean (sd): 31.5 (25.4). min < med < max: 4 < 21.8 < 100. IQR (CV): 29 (0.8)	369	602 (100.0%)	0 (0.0%)
9	Sustentability [numeric]	Mean (sd): 49.6 (32.3). min < med < max: 1 < 48.7 < 100. IQR (CV): 59.9 (0.7)	330	602 (100.0%)	0 (0.0%)

Source: Authors. QS (2024).

Appendix 1 presents a Table of correlations between the attributes of the three indices for the 2024 rankings.

3. Results

First, we analyze the "importance weights" calculated using equations (1), (2), and (3). We focus on the "importance weights" for the ARWU Ranking.

Table 4 illustrates the percentage contribution of each indicator (importance weights) to the "BoD" efficiency score when there are no restrictions. This aligns with the traditional BoD calculation presented in equation (1). The shadows of gray coding in the table represent each indicator's relative contribution across institutional tiers (Top 10, 50, and 100 universities):

- Darkest gray: Primary efficiency driver (maximum contribution).
- Medium gray: Secondary determinant.
- Lightest gray: Tertiary contributing factor.

The calculation for the entire sample is also included in Table 4.

Table 4: Importance Weights – BoD- ARWU

	Alumni	Awards	Highly Cited Publication	Nature & Science	Publications	Per Capita Academic Performance
Top 10	0.088	0.319	0.000	0.004	0.383	0.132
Top 20	0.048	0.216	0.000	0.004	0.498	0.098
Top 50	0.020	0.102	0.000	0.002	0.599	0.040
Top 100	0.009	0.046	0.001	0.001	0.555	0.045
ALL	0.002	0.006	0.000	0.000	0.318	0.036
ARWU weights	20	20	20	20	20	10

Source: Authors. ARWU (2024)

A table analysis reveals a significant dominance of Publications in the rankings. They hold a substantial 38.3% share in the top 10 and even up to 59.9% in the top 50, revealing their pivotal role in driving efficiency, particularly in non-elite universities. On the other hand, Awards exhibit a certain stratified relevance. While they contribute 31.9% in the top 10, their impact diminishes to only 4.6% in the top 100. This disparity makes awards a key differentiator, primarily in elite institutions. Alumni and Awards are affected by a depreciation rate. Since Awards are usually achieved at mature ages, depreciation affects the alumni more than the Awards variable. Moreover, the weight of the Awards is greater than that of the Alumni.

Furthermore, Highly Cited Publications and Nature & Science appear to make a minimal contribution. Despite their official weight of 20%, these indicators contribute less than 0.5%. Consequently, the ARWU's fixed weights (20% for Publications, Highly Cited Publications, and

Nature & Science) may not accurately reflect real patterns. For instance, they underestimate the significance of Publications and overestimate the relevance of Highly Cited Publications and Nature & Science.

As highlighted above, BoD can generate weights equal to zero. To counteract this result, we calculated BoDCONST (Eqs. 2, 3). For ARWU, the (ω_{ri}^{*}) were constrained to a minimum value of 0.05 and a maximum upper value of 0.5.

Table 5 shows the results with BoD when the importance weights are constrained.

Table 5. Importance Weights – BoDCONST- ARWU

	Alumni	Award	Highly Cited Publications	Nature & Science	Publications	Per Capita Academic Performance
ALL	0.0675	0.0641	0.0985	0.0856	0.4511	0.2333
Top 10	0.1700	0.4400	0.0500	0.0750	0.1200	0.1450
Top 20	0.1100	0.2675	0.1000	0.1350	0.2775	0.1100
Top 50	0.0840	0.1520	0.1200	0.1240	0.3680	0.1520
Top 100	0.0745	0.1230	0.1025	0.1190	0.3990	0.1820
ARWU weights	20	20	20	20	20	10

Source: Authors. ARWU (2024)

Both the BoDCONST and unrestricted (BoD) models reveal consistent patterns in the main drivers of university efficiency. Both approaches point to PUB as the most relevant indicator, particularly for non-elite institutions (Top 50-ALL), where it maintains substantial contributions (36.8–45.1% in BoDREST vs. 38.3–59.9% in unrestricted BoD). Furthermore, both models agree on the low real contributions of Highly Cited Publications and Nature & Science, which at best reach 12% in BoDCONST (due to artificial restrictions). They are marginal (<0.4%) in the unrestricted model.

The second ranking we evaluate is THE ranking. Table 6 shows the "importance weights" for the BoD (unrestricted).

Table 6: Importance Weights – BoD- THE

	Teaching	Research Environment	Research Quality	Industry Impact	International Outlook
Top 10	0.0015	0.0763	0.0645	0.7929	0.0643
Top50	0.0010	0.0434	0.0709	0.7194	0.1607
Top 100	0.0006	0.0305	0.1428	0.5713	0.2422
Top 500	0.0003	0.0078	0.2904	0.3254	0.2675
All	0.0072	0.0027	0.2272	0.1930	0.2094
THE weights	29.5	29	30	4	7.5

Source: Authors. THE (2024)

Concerning specific patterns in institutional efficiency measured by the THE ranking, Industry Impact is the most prominent indicator at almost all levels analyzed (79.3% in the Top 10 versus 19.3% in the full sample). On the other hand, Research Quality exhibits a differentiated behavior, increasing its contribution in universities with lower selectivity (6.4% in the Top 10 versus 29.0% in the Top 500).

Both results contrast with the official THE weights, particularly the extreme underestimation of Industry Impact (4% official weight versus 19.79% in the BoD model) and the overestimation of Teaching (29.5% official versus <1% actual).

The evidence suggests that the current weighting scheme does not adequately reflect the determinants of efficiency, especially the critical importance of industry-academia linkages at elite institutions. To calculate BoDCONST, the (ω_{ri}^{*}) were constrained to a minimum value of 0.05 and a maximum of 0.8 for THE ranking case. Table 7 shows the results.

Table 7. Importance Weights – BoDCONST- THE

	Teaching	Research Environment	Research Quality	Industry Impact	International Outlook
Top 10	0.0718	0.1008	0.2950	0.3493	0.1831
Top 50	0.0610	0.0757	0.2840	0.4029	0.1765
Top 100	0.0571	0.0666	0.3011	0.3989	0.1763
Top 500	0.0530	0.0542	0.3643	0.2919	0.2366
All	0.0708	0.0523	0.3865	0.2270	0.2634
THE weights	29.5	29	30	4	7.5

Source: Authors. THE (2024)

The BoDCONST model reveals a slightly different weight distribution than the BoD, possibly indicating a greater balance between indicators. While Industry Impact maintains its predominance (22.7-34.9%), its contribution decreases significantly compared to BoD, especially at elite institutions. Simultaneously, Research Quality increases its importance in the Top 10 (29.5% vs. 6.45%), and traditionally marginal indicators such as Teaching reach the minimum importance weight (5.7-7.2%). Compared with the official THE weights, the restricted model substantially corrects the underestimation of Industry Impact and the overestimation of Teaching.

Finally, we analyze the QS Ranking, beginning by determining the contribution of each indicator to the efficiency score via unrestricted BoD. Table 8 shows the results.

Table 8: Importance Weights – BoD- QS

	Academic Reputation	Employer Reputation	Faculty / Student	Citations	International Faculty	International Students	International Research	Employers	Sustainability
Top 10	0.0177	0.0092	0.1287	0.0735	0.4236	0.1780	0.0517	0.0212	0.0960
Top 50	0.0336	0.0264	0.1388	0.1114	0.3457	0.1465	0.0651	0.0198	0.1090
Top 100	0.0239	0.0451	0.1254	0.1161	0.3047	0.1392	0.0824	0.0216	0.1305
All	0.0166	0.0618	0.1184	0.1152	0.1895	0.0950	0.1178	0.0185	0.1298
QS weights	30	15	10	20	5	5	5	5	5

Source: Authors. QS (2024)

The BoD results for the QS ranking reveal significant discrepancies with their official weights. Internationalization indicators, particularly faculty and students, emerge as the primary drivers of efficiency, collectively accounting for between 30.45% (for the full sample) and 60.16% (for the Top 10) of the actual weight, compared to the modest 10% officially assigned. In contrast, reputation indicators, which in the QS scheme carry 45% of the weight, show a marginal contribution in the BoD model. Their weighting in the BoD model is only 2.69% in the Top 10 and 7.84% in the full sample, indicating an overestimation of their contribution compared to the weight assigned by THE World University Ranking. Consequently, using the BoD, the current QS ranking would underestimate the significance of international components, overvalue reputational components, and maintain a weighting of Citations (20%) that significantly exceeds its actual observed contribution (7-12%).

To calculate BoDCONST for QS, the (ω_{ri}^{*}) were constrained to a minimum value of 0.05 and a maximum upper value of 0.5. Table 9 shows the results:

Table 9. Importance Weights – BoDCONST- QS

	Academic Reputation	Employer Reputation	Faculty / Student	Citations	International Faculty	International Students	International Research	Employers	Sustainability
Top 10	0.1415	0.0500	0.0577	0.1382	0.1086	0.1216	0.1604	0.0681	0.1538
Top 50	0.0992	0.0557	0.0877	0.1281	0.1444	0.1623	0.1051	0.0588	0.1588
Top 100	0.0802	0.0638	0.0845	0.1213	0.1681	0.1399	0.1141	0.0555	0.1727
All	0.0635	0.0900	0.1190	0.1351	0.1507	0.1125	0.1273	0.0622	0.1398
QS weigths	30	15	10	20	5	5	5	5	5

Source: Authors. QS (2024)

The BoDCONST model identifies emerging indicators as the key factors of efficiency, particularly for the top 10 institutions:

1) The International Research Network holds a higher relevance (16.04% BoDCONST vs. 5% QS), and Sustainability is also important (15.38% BoDCONST vs. 5%).
2) The contribution of reputational indicators, particularly Academic Reputation (14.15% BoDCONST vs. 30% QS), is relatively low. Initially, the QS systematically underestimates internationalization indicators while overestimating traditional components like academic reputation.
3) Both BoD models agree on the need to rebalance the ranking by:

   i) Increasing the weight of Sustainability (≥15%) and International Research (≥10%).

   ii) Reducing Academic Reputation (≤15%).

   iii) Implementing different asymmetric restrictions (e.g., 0.05-0.3 for reputation vs. 0.1-0.4 for internationalization).

An interesting aspect of the rankings is the efficiency scores for each indicator (eq), which reveal performance. We use BoDRMDIR to evaluate those scores, presenting them in Table 10.

Table 10. Efficiency scores for each indicator. BoDRMDIR - ARWU

	Alumni	Award	Highly Cited Publications	Nature & Science	Publications	Per Capita Academic Performance
Top 10	0.6580	0.7302	0.6256	0.6853	0.7205	0.6726
Top 20	0.3419	0.3714	0.3251	0.3729	0.5010	0.3893
Top 50	0.2140	0.2328	0.2196	0.2404	0.4027	0.2546
Top 100	0.1448	0.1580	0.1839	0.2011	0.3642	0.2138
ALL	0.0648	0.0539	0.1311	0.1296	0.3296	0.1548

Source: Authors

Table 10 shows that the efficiency scores (e_q) obtained through the Multi-directional Robust Benefit of the Doubt (BoDRMDIR) model reveal significant variations in the performance of each indicator. The analysis shows that Alumni (e_q) Top 10 = 0.6580 vs. ALL = 0.0648 and Awards (eq Top 10 = 0.7302 vs. ALL = 0.0539) display the most notable gaps between elite institutions and the average, emphasizing their highly selective nature. In contrast, Publications shows a more balanced performance, with an e_q Top 10 of 0.7205 compared to 0.3296 for ALL, confirming its robustness as a cross-cutting metric.

Table 11 shows that BoDRMDIR efficiency scores highlight significant differences in institutional performance across the individual indicators of THE ranking.

Table 11. Efficiency scores for each indicator. BoDRMDIR – THE

	Teaching	Research Environment	Research Quality	Industry Impact	International Outlook
TOP 10	0.9818	0.9842	0.9859	0.9700	0.9763
TOP 50	0.7897	0.8129	0.9482	0.9198	0.9498
TOP 100	0.6368	0.6730	0.9083	0.8804	0.8910
TOP 500	0.4735	0.4943	0.8320	0.7717	0.7154
ALL	0.2507	0.2236	0.5257	0.3887	0.4298

Source: Authors

Elite universities (in the Top 10) show near-perfect efficiency (>0.97) across all categories, particularly excelling in Research Quality (0.9859) and Industry Impact (0.9700), which remain key distinctions within the Top 50 (0.9482 and 0.9198, respectively). However, a sharp decline in performance appears beyond this level, with Teaching (0.2507) and Research Environment (0.2236) showing the most significant drops across the institution. International Outlook maintains relatively stable efficiency across tiers (0.9763 for Top 10 to 0.4298 for ALL).

Table 12. Efficiency scores for each indicator.
BoDRMDIR – QS

	Academic Reputation	Employer Reputation	Faculty / Student	Citations	International Faculty	International Students	International Research	Employers	Sustainability
Top 10	0.9984	0.9926	0.9860	0.9951	0.9987	0.9934	0.9875	0.9987	0.9944
Top 50	0.9832	0.9595	0.9375	0.9687	0.9438	0.9509	0.9452	0.9512	0.9634
To 100	0.8337	0.7939	0.7660	0.8680	0.9219	0.8848	0.8983	0.7749	0.9237
All	0.4038	0.4029	0.4470	0.4941	0.5314	0.4438	0.5074	0.3410	0.5278

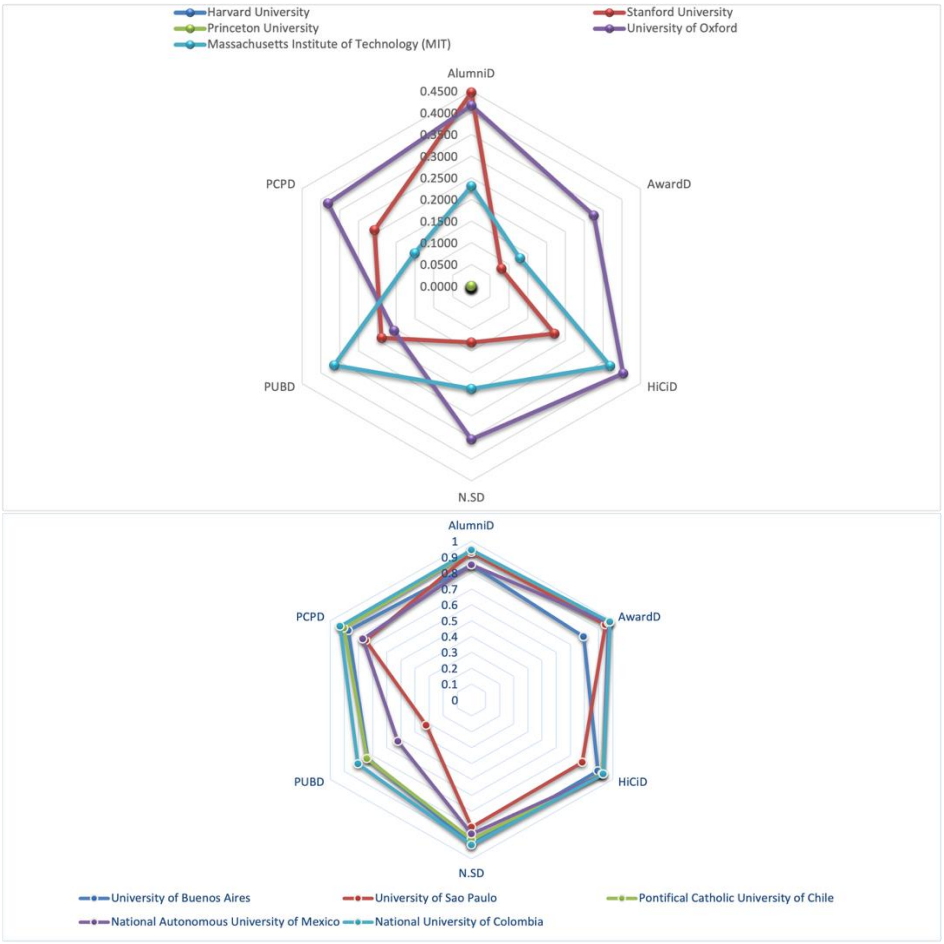
Source: Authors

Table 12 shows that the BoDRMDIR efficiency scores for individual indicators reveal some variation in performance across QS indicators.

Elite institutions (top 10) show near-perfect efficiency (>0.98) in all categories. International Faculty (0.9987) and Employer Reputation (0.9987) stand out as the strongest differentiators for top-tier universities, while Citations (0.9951) and Sustainability (0.9944) also perform exceptionally well. However, beyond the top 50, most indicators decline by 30-40 percentage points, notably affecting Employer Reputation (0.9595 to 0.4029) and Academic Reputation (0.9832 to 0.4048). Interestingly, the International Research Network (0.5074) and Sustainability (0.5278) maintain higher efficiency scores across institutions.

Regarding the improvement trajectory for each university, we select the most significant universities in each ranking and the most prominent universities in Latin America to illustrate the proportion by which each of the simple indicators must be increased to reach the frontier.

Figure 1. Indicator improvement trajectories. ARWU ranking



Source: Authors

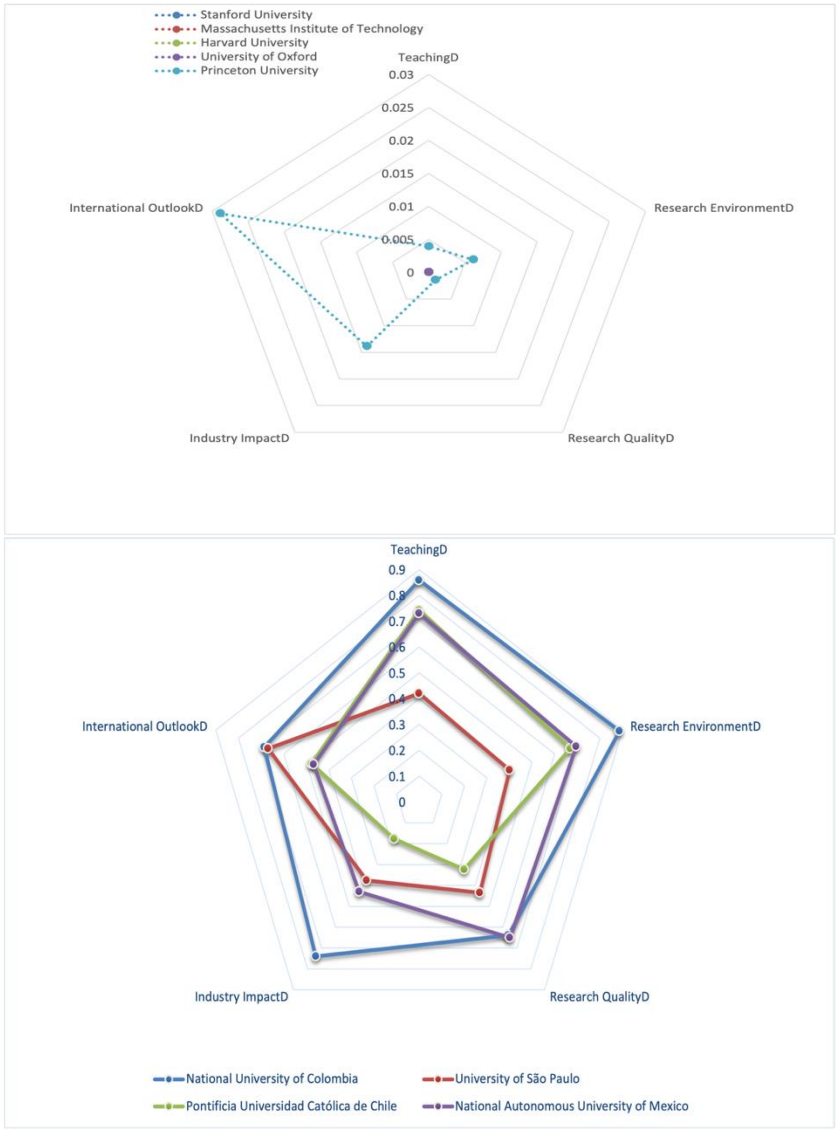
The radar charts (Figure 1) illustrate how universities must improve across ARWU indicators to reach the efficiency frontier; a value of 0.30 indicates that a 30% improvement is required. The upper panel features elite global universities, including Harvard, MIT, Stanford, Oxford, and Princeton, which still exhibit inefficiencies in certain areas. Oxford and Stanford, for example, lag in Alumni and Per Capita Performance (PCPD), needing significant improvements (30–45%). Harvard and Princeton are closer to full efficiency (approaching 0), requiring only minor adjustments. This indicates high rankings do not always mean efficiency in all areas.

The lower panel highlights leading Latin American universities, such as the University of Buenos Aires, the University of São Paulo (USP), and UNAM (Mexico). These face large gaps in Award, Alumni, and Highly Cited Researchers, often more than doubling their current performance. While some, such as USP and UNAM, perform better in Publications and PCPD, they struggle overall with global research impact.

Leading global institutions show minimal inefficiencies overall (Figure 2). The main gap is in Princeton's International Outlook, which needs only a 2–3% increase. The required improvements for all other indicators—teaching, Research Environment, Research Quality, and Industry Impact — are negligible (often below 1%), suggesting these universities are operating near the efficiency frontier. Their performance is thus highly balanced and close to optimal in the context of the THE ranking metrics. In contrast, the Latin American universities—UNAM, the USP, the National University of Colombia, and the Pontificia Universidad Católica de Chile (the University of Buenos Aires did not participate in the 2024 THE ranking)- exhibit considerably larger gaps across multiple indicators. All institutions would need to improve substantially in Teaching, Research Environment, and Research Quality, where required enhancements range from 40% to 90%. For instance, the University of São Paulo and UNAM show the greatest

inefficiencies in Teaching and Research Quality. At the same time, the National University of Colombia shows relatively minor gaps in Industry Impact and International Outlook (60%).

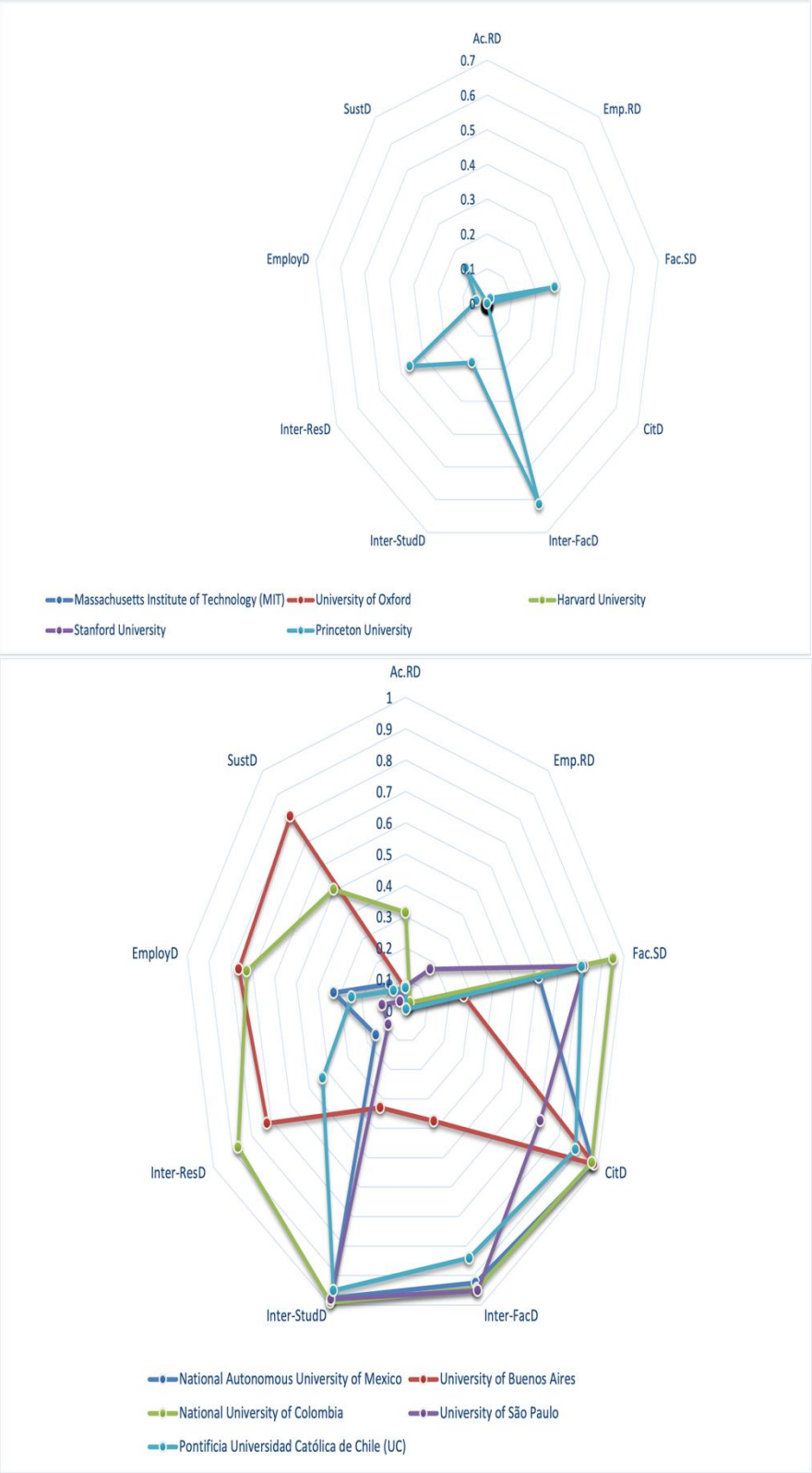
Figure 2. Indicator improvement trajectories. THE Ranking



Source: Authors

The upper panel highlights world-leading universities performing excellently across most indicators (Figure 3). However, Princeton University indicates a significant need for improvement (between 20% and 60%), especially in internationalization metrics such as International Faculty, International Students, and International Research. In contrast, other areas like Academic Reputation and Faculty-Student Ratio are nearly optimal. The lower panel shows Latin American universities, including UBA and USP, where substantial gaps remain across several criteria—citations, faculty ratios, and reputation—with some shortfalls exceeding 80% of the target benchmarks. While institutions such as USP, UBA, and UNAM achieve relatively strong results in academic reputation, ongoing challenges persist in areas like research impact (citations) and internationalization.

Figure 3. Indicator improvement trajectories. QS Ranking



Source: Authors

4. Discussion of results

In the ARWU ranking (1), Publications (scientific productivity) would be the actual driver of institutional efficiency, and (2) High Cited Research plus Nature and Science (research quality indicators) would have a limited impact on practice. The results reveal an apparent discrepancy

between the fixed ARWU weights (10-20% per indicator) and the empirical evidence from the BoD models: while the ARWU significantly overestimates the role of Highly Cited Research and Nature & Science (with 20% assigned vs. <5% contribution) and underestimates Publications (20% assigned vs. 38-60% observed), the BoD models show that these weights should be dynamic and stratified, adjusting to the institutional level (e.g., Publications requiring 35-55% according to stratum, and Highly Cited Research and Nature & Science deserving ≤5%).

Despite their differences, for THE ranking, both approaches agree in identifying Industry Impact as the primary driver of efficiency at all institutional levels, albeit with varying intensities (34.9 - 79.3% in the Top 10). Furthermore, the models converge in pointing out the relatively low relevance of Teaching and Research Environment, which systematically occupy the last positions in contribution, regardless of the restrictions applied. Both schemes also reflect the progressive increase in Research Quality with lower institutional prestige (Top 10: 6.4–29.5% and Top 500: 29.0–36.4%), confirming its role as a differentiating variable in non-elite universities. We can observe three fundamental patterns: 1) the governance of industry-academia collaborations, 2) the comparative insignificance of teaching in rankings based on BoD, and 3) the inherent stratification of research excellence criteria. Moreover, restrictions modify the magnitude of these structural relationships but not their direction, further reinforcing our findings.

The differences in weights assigned to the QS 2024 indicators based on BoD and technical weights can be attributed to the fact that the analysis is limited to only 600 universities out of 1,498. This sample may be overrepresented by elite institutions, potentially inflating the weights assigned to indicators such as sustainability. Additionally, the contribution of the Faculty/Student Ratio indicator could be underestimated for non-elite universities. Consequently, it is advisable to interpret the results as representative only of the analyzed segment (600 universities) and not necessarily extrapolated to the entire global higher education institution set.

For ARWU, BoDRMDIR reveals that: (1) traditional indicators (Alumni, Awards) exhibit diminishing e_q values outside the Top 50; (2) productivity metrics (Publications, Publications per Capita) maintain relatively stable performance; and (3) there is a clear need for differentiated weighting by institutional tier. The model shows that although Highly Cited Research (eq = 0.6256–0.1311) and Nature & Science (eq = 0.6853–0.1296) are correlated, they follow different paths depending on institutional prestige.

As previously noted, this analysis indicates specific strategic priorities for universities, depending on their institutional tier. Elites should maintain their advantages in industry impact and research quality. At the same time, emerging institutions would benefit from focusing on an international outlook and selective improvements in research quality, as these offer more achievable efficiency gains than competing on teaching infrastructure.

The results suggest that these newer indicators may offer more practical avenues for institutional improvement than traditional reputation metrics. Following the same analysis, we recommend that universities focus on different areas depending on their level. For example, elite institutions should prioritize reputation and internationalization, while emerging universities should emphasize sustainability and international research networks, which are more effective across various tiers. According to Vidoli et al. (2024), even if generalization is not always practical, a university excelling in nearly all indicators but performing poorly in one will not have a lower overall score than a university with moderate performance across all indicators. This pattern becomes more evident as the number of indicators increases.

Assessing Spearman rank correlations is essential as they show the consistency between various university ranking methodologies. Table 10 indicates that the Spearman correlation coefficients display clear patterns in the relationship between BoD methods and global rankings.

Table 13- Spearman Rank Correlations

Method	ARWU	THE	QS
BoD	0.48	0.91	0.55
BoDCONST	0.91	0.98	0.86
BoDRMULTI	0.18	0.96	0.91

Source: Authors.

The standard BoD method shows a strong correlation with THE (ρ = 0.91), a moderate correlation with QS (ρ = 0.55), and a low correlation with ARWU (ρ = 0.48). This suggests it aligns more closely with THE, likely due to shared subjective criteria. Meanwhile, BoDCONST shows nearly perfect correlations with THE (ρ = 0.98) and strong links with QS (ρ = 0.86) and ARWU (ρ = 0.91). This suggests that the added constraints help the method match rankings based on reputation and quality of research. Conversely, BoDRMULTI, while highly correlated with THE (ρ = 0.96) and QS (ρ = 0.91), has a very weak correlation with ARWU (ρ = 0.18).

In summary, although BoDCONST aligns with traditional research rankings, BoDRMULTI may better reflect the diversity of university missions, and the standard BoD sits in the middle ground with limitations for both purposes.

5. Conclusions

The paper improves the analytical framework of global university rankings by integrating efficiency-based weighting methodologies (BoD, BoDCONST) with multi-directional robust diagnostics (MDir_RBoD). The findings reveal implicit discrepancies between the ranking methodologies' official and effective weighting schemes. Besides, the results show granular indicator-level efficiency profiles that facilitate targeted improvement strategies for individual universities. Thus, they provide policy-relevant insights for benchmarking and ensuring equitable performance assessments. By shifting the analytical focus from overall scores to diagnostics tailored to specific indicators, this method provides actionable insights for universities seeking to improve their rankings and for policymakers developing, for example, performance-based funding models.

A key limitation of this analysis is that most publicly available ranking data provide detailed final ranking indicators only for the top 100 universities (ARWU) or a slightly larger number (600 -QS). This restricts the scope of our analysis. Extending the methodology to broader institutions could yield deeper insights into mid-tier and emerging universities if more granular data becomes publicly available. Future research could improve this paper in several ways. For example, introducing new methods to compare rankings over time requires careful harmonization due to changes in ranking methodologies (e.g., indicator definitions or weight allocations). Developing adjustment techniques to account for these shifts would enable meaningful year-to-year comparisons. Finally, applying the approach to discipline-specific indicators and integrating qualitative dimensions could help mitigate biases in bibliometric and reputational metrics, providing a more complete view of university missions.

References

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

Altbach, P. G. (2021). Foreword: reflections on rankings. In Ellen Hazelkorn and Georgiana Mihut, “Research Handbook on University Rankings. Theory, Methodology, Influence and Impact.” Elgar Handbook in Education. Edward Elgar Publishing.

Barsky, O. (2012). Acerca de los rankings internacionales de las universidades y su repercusión en Argentina. Debate Universitario 11/2012. Universidad Abierta Interamericana. ISSN: 2314–1530.

Beath, J. & Katsoulacos (1991). The economic theory of product differentiation. Cambridge University Press.

Bowman, N. A. & Bastedo, M. N. (2011). Anchoring effects in world university rankings: exploring biases in reputation scores. *Higher Education* (2011) 61: 431–444. https://doi.org/10.1007/s10734-010-9339-1

Cantú-Ortiz, F. & Fangmeyer Jr., J. (2018). University Performance in the Age of Research Analytics. In Cantú-Ortiz, F. J. (Editor) (2018). Research Analytics Boosting University Productivity and Competitiveness through Scientometrics. CRC Press Taylor & Francis Group.

Chankseliani, M. (2023). Who funds the production of globally visible research in the Global South? *Scientometrics* (2023) 128: 783–801. https://doi.org/10.1007/s11192-022-04583-4

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to 'Benefit of the doubt' composite indicators. *Social Indicators Research,* 82, 111–145. https://doi.org/10.1007/s1120 5-006-9029-7.

De Witte, K., & Hudrlikova, L. (2013). What about excellence in teaching? A benevolent ranking of universities. *Scientometrics*, 96, 337–364. https ://doi.org/10.1007/s1119 2-013-0971-2.

Dip, J.A. (2021). What does U-Multirank tell us about knowledge transfer and research? *Scientometrics*, 126, 3011–3039 (2021). https://doi.org/10.1007/s11192-020-03838-2

El Gibari, S., Gómez, T. & Ruiz, F (2022). Combining reference point-based composite indicators with data envelopment analysis: application to the assessment of universities. *Scientometrics* 127, 4363–4395. https://doi.org/10.1007/s11192-022-04436-0

Ferro, G. and D'Elia, V (2020). Higher Education Efficiency Frontier Analysis: A Review of Variables to Consider. *Journal of Efficiency and Responsibility in Education and Science* 13(3), 140-153, 09/2020. https://doi.org/10.7160/eriesj.2020.130304

Giacalone, R. A. (2009). Academic rankings in research institutions: A case of skewed mind-sets and professional amnesia. *Academy of Management Learning & Education,* 8(1), 122–126. http://www.jstor.org/stable/40214578

Green, M. (2012). Measuring and assessing internationalization. https://nafsa.org/uploadedfiles/nafsa_home/resource_library_assets/publications_library/measuring and assessing internationalization.pdf

Green, M. F., Marmolejo, F., & Egron-Polak, E. (2012). The SAGE handbook of international higher education.

Hamann, J. & Ringel, L. (2023). The discursive resilience of university rankings. *Higher Education*, 86, 845–863. https://doi.org/10.1007/s10734-022-00990-x

Hazelkorn, E. (2015). Rankings and the Reshaping of Higher Education. The Battle for World-Class Excellence. 2nd edition, Palgrave Macmillan.

Hazelkorn, E. & Gibson, A. (2017). Global science, national research, and the question of university rankings. *Palgrave Communications,* 3(21). https://doi.org/10.1057/s41599-017-0011-6

Hazelkorn, E. & Mihut, G. (2021). Introduction: putting rankings in context -looking back, looking forward. In Ellen Hazelkorn and Georgiana Mihut, "Research Handbook on University Rankings.

Theory, Methodology, Influence and Impact." Elgar Handbook in Education. Edward Elgar Publishing.

Holmes, R. (2021). What do global rankings try to measure? Global university ranking methodologies since 2003. In Ellen Hazelkorn and Georgiana Mihut, "Research Handbook on University Rankings. Theory, Methodology, Influence and Impact." Elgar Handbook in Education. Edward Elgar Publishing.

Jain, I. & Gulati, R. (2024). How has higher education developed in India? Evidence using a "benefit-of-the-doubt" based multidimensional index approach. *Journal of Applied Research in Higher Education*. https://doi.org/10.1108/JARHE-02-2024-0107

Jing Wang, J. Halffman, W., & Horbach, S. P. J. M. (2024). Listing quality: Chinese journal lists in incoherent valuation regimes. *Science and Public Policy,* 2024, 51, 55–66. https://doi.org/10.1093/scipol/scad052

Johnes, J. (2018). University rankings: What do they really show? *Scientometrics,* 115, 585–606. https://doi.org/10.1007/s11192-018-2666-1

Kim, J. & Bastedo, M. (2021). How do we know what we know? Empirical methodologies for studying college rankings. In Ellen Hazelkorn and Georgiana Mihut, "Research Handbook on University Rankings. Theory, Methodology, Influence and Impact." Elgar Handbook in Education. Edward Elgar Publishing.

Li, M., Shankar, S. & Tang, K. K. (2011). Why does the USA Dominate University League Tables? *Studies in Higher Education*, 36: 923–937. https://doi.org/10.1080/03075079.2010.482981

Locke, W. (2021). Researching and understanding the influence of rankings on higher education institutions: logics, methodologies and conceptualisations. In Ellen Hazelkorn and Georgiana Mihut, "Research Handbook on University Rankings. Theory, Methodology, Influence and Impact." Elgar Handbook in Education. Edward Elgar Publishing.

Maisano, D. A., Mastrogiacomo, L. & Franceschini, F. (2023). Empirical evidence on the relationship between research and teaching in academia, *Scientometrics* (2023) 128: 4475–4507 https://doi.org/10.1007/s11192-023-04770-x

Malea, E. & Perez-Esparrells, C. (2021). Multidimensional taxonomy of university rankings. In Ellen Hazelkorn and Georgiana Mihut, "Research Handbook on University Rankings. Theory, Methodology, Influence and Impact." Elgar Handbook in Education. Edward Elgar Publishing.

Mammola, S., Piano, S., Doretto, A., Caprio, E., and Chamberlain, D. (2022). Measuring the influence of non-scientific features on citations. *Scientometrics* (2022) 127: 4123 4137 https://doi.org/10.1007/s11192-022-04421-7

Marginson, S. (2021). The reality underneath the rankings: trends in global science. In Ellen Hazelkorn and Georgiana Mihut, "Research Handbook on University Rankings. Theory, Methodology, Influence and Impact." Elgar Handbook in Education. Edward Elgar Publishing.

Marhl, M., Markovič, R., Grubelnik, V. and Perc, M. (2025). The changing world dynamics of research performance. *Scientometrics* (2025) 130: 469–488 https://doi.org/10.1007/s11192-024-05199-6

Matulová, M. (2023). Ranking of European Universities by DEA-Based Sustainability Indicator. *Journal on Efficiency and Responsibility in Education and Science*, 16(4), 287–298. https://doi.org/10.7160/eriesj.2023.160403

Merton, R. K. (1968). The Matthew Effect in Science. *Science.* 159 (3810): 56–63. https://doi.org/10.1126/science.159.3810.56

Olcay, G. A. & Bulu, M. (2016). Is measuring the knowledge creation of universities possible? A review of university rankings. *Technological Forecasting and Social Change* (2016), http://dx.doi.org/10.1016/j.techfore.2016.03.029

Prado, A. F. B. A. (2023). Estudo do ranking acadêmico multidimensional U-Multirank. *Estudos em Avaliação Educacional*, 34, Artigo e09117. https://doi.org/10.18222/eae.v34.9117_port

Ross, D. (2018). Times Higher Education World Universities Ranking. In Cantú-Ortiz (Editor), Research Analytics Boosting University Productivity and Competitiveness through Scientometrics. CRC Press Taylor & Francis Group.

Sowter, B., Reggio, D., & Hijazi, S. (2018). QS World University Rankings. In Cantú-Ortiz (Editor), Research Analytics Boosting University Productivity and Competitiveness through Scientometrics. CRC Press Taylor & Francis Group.

Stack, M. (2016). Global University Rankings and the Mediatization of Higher Education. Palgrave Macmillan.

Stack, M. (Editor) (2021). Global University Rankings and the Politics of Knowledge. University of Toronto Press.

Szuwarzyński, A. (2019). Benefit of the doubt approach to assessing the research performance of Australian universities. *Higher Education Quarterly,* 73, 235–250. https://doi.org/10.1111/hequ.12184 .

Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with Benefit-of-the-Doubt weights. *European Journal of Operational Research,* 256(3), 1004-1014. https://doi.org/10.1016/j.ejor.2017.06.061

Vidoli, F., Fusco, E., Pignataro, G., Guccio, C. (2024). Multi-directional Robust Benefit of the Doubt model: An application to measuring the quality of acute care services in OECD countries. *Socio-Economic Planning Sciences*. https://doi.org/10.1016/j.seps.2024.101877

Wu, Y. & Liu, N. C. (2018). Academic Ranking of World Universities (ARWU): Methodologies and Trends. In Cantú-Ortiz (Editor), Research Analytics Boosting University Productivity and Competitiveness through Scientometrics. CRC Press Taylor & Francis Group.

Appendix 1: Correlations between the attributes of the three global indices for the 2024 rankings.

Table A1: Correlations between attributes of the three indices greater than two-thirds

N = 361	Alumni	Award	Highly Cited Publications	Nature And Science	Publications	Per Capita Academic Publications	**ARWU**
Teaching	0.669	0.664	0.690	0.831	0.684	0.641	0.833
Research Environment.	0.635	0.659	0.729	0.836	0.693	0.678	0.843
Research Quality	0.323	0.363	0.587	0.512	0.378	0.504	0.524
Industry Impact	0.191	0.249	0.373	0.383	0.436	0.379	0.406
International. Outlook	0.217	0.230	0.190	0.206	-0.111	0.334	0.175
THE	0.639	0.663	0.772	0.848	0.662	0.723	0.852
Academic Reputation	0.598	0.588	0.621	0.743	0.645	0.540	0.750
Employer Reputation	0.465	0.458	0.404	0.531	0.429	0.392	0.534
Faculty To Students	0.378	0.352	0.320	0.404	0.285	0.308	0.402
Citations	0.254	0.299	0.502	0.445	0.449	0.521	0.485
International Faculty	0.139	0.166	0.197	0.167	-0.056	0.323	0.155
International Students	0.258	0.255	0.187	0.204	-0.034	0.313	0.204
International Research Network	0.354	0.341	0.384	0.452	0.363	0.313	0.443
Employer	0.514	0.532	0.486	0.575	0.457	0.409	0.596
Sustainability	0.331	0.327	0.337	0.406	0.230	0.264	0.376
QS	0.609	0.594	0.579	0.704	0.528	0.545	0.704

	Teaching	Research Environment.	Research Quality	Industry Impact	International. Outlook	**THE**
Academic Reputation	0.820	0.854	0.338	0.412	0.199	0.797
Employer Reputation	0.687	0.701	0.139	0.358	0.153	0.616
Faculty To Students	0.569	0.447	0.045	0.308	0.024	0.424
Citations	0.485	0.555	0.496	0.437	0.189	0.600
International Faculty	0.056	0.206	0.475	0.102	0.852	0.346
International Students	0.200	0.300	0.403	0.070	0.835	0.412
International Research Network	0.281	0.393	0.493	0.056	0.391	0.461
Employer	0.632	0.634	0.309	0.267	0.196	0.622
Sustainability	0.293	0.372	0.395	0.094	0.424	0.431
QS	0.813	0.809	0.334	0.412	0.327	0.788

Source: Authors