Principles and Practices in National Rankings: Assessing the Alignment between the Berlin Principles and Brazil's Folha University Ranking

Marco Felipe Zanchetta Moreno Guidio Biondo[1], Thiago Henrique Almino Francisco[1], Giancarlo Moser[2], Vanessa Moraes de Andrade[1] & Mouhamadou Moustapha Seck[2]

[1] Universidade do Extremo Sul Catarinense, Brazil

[2] Universidade do Sul de Santa Catarina, Brazil

Correspondence: Marco Felipe Zanchetta Moreno Guidio Biondo, Universidade do Extremo Sul Catarinense, Brazil.

Received: August 20, 2025 Accepted: October 5, 2025 Online Published: October 11, 2025

doi:10.5430/ijhe.v14n5p65 URL: https://doi.org/10.5430/ijhe.v14n5p65

Abstract

National university rankings have gained increasing prominence as instruments for quality signaling, institutional benchmarking, and policy influence. However, their methodological rigor and capacity to reflect diverse institutional missions remain contested. This study critically examines the extent to which the Ranking Universitário Folha (RUF), Brazil's most visible national ranking, aligns with the Berlin Principles on Ranking of Higher Education Institutions. Drawing on a longitudinal document analysis (2012–2024) and semi-structured interviews with higher education experts, the study evaluates RUF's compliance with each of the 16 Berlin Principles, with special attention to Principle 3 concerning the recognition of institutional diversity. The findings reveal strong alignment in areas such as methodological transparency, data verifiability, and multi-criteria evaluation, but persistent misalignment in acknowledging institutional mission differentiation and social engagement—dimensions central to Brazil's higher education ecosystem. A comparative discussion with international rankings (e.g., THE, QS, U.S. News) and policy frameworks contextualizes RUF's structural limitations. The article concludes with six actionable propositions to recalibrate RUF toward a more inclusive, balanced, and development-oriented ranking model. These findings contribute to critical debates on the future of rankings and their role in shaping equitable and context-sensitive higher education systems.

Keywords: university rankings, institutional diversity, higher education policy

1. Introduction

University rankings have become an influential feature of the global higher education landscape, shaping public perceptions and institutional behavior (Hazelkorn, 2015). These league tables promise a simplified comparison of universities, but their methodologies and consequences have been widely debated. In response to concerns about quality and fairness, the International Ranking Expert Group (IREG) formulated the Berlin Principles on Ranking of Higher Education Institutions in 2006. The Berlin Principles consist of 16 guidelines intended to promote good practices in ranking design, data collection, and transparency (International Ranking Expert Group [IREG], 2006). They emphasize, among other values, clarity of purpose, methodological transparency, validity of indicators, and recognition of the diversity of institutional missions. In essence, the Berlin Principles seek to align rankings with academic values and minimize distortions (Barron, 2017; Chen & Liu, 2008).

Meanwhile, many countries have developed their own national rankings to complement or counterbalance global league tables. Brazil's Ranking Universitário Folha (RUF), launched in 2012 by the newspaper Folha de S. Paulo, is one prominent example. RUF is the first comprehensive Brazilian university ranking, evaluating all nationally accredited universities on multiple criteria (Folha de S. Paulo, 2024). As a media-driven ranking, RUF gained rapid visibility in Brazil and has been published annually (with a brief pause from 2020–2022) as a major reference on institutional performance. RUF's methodology was inspired by international rankings and adapted to local data availability, using five broad indicators: Research, Teaching, Market (employability), Innovation, and Internationalization, each composed of specific measures with assigned weights (Folha de S. Paulo, 2024; Fausto, Calero-Medina, & Noyons, 2016). Over the past decade, RUF has become influential in Brazilian higher education discourse, often cited by university administrators and policymakers as an indicator of institutional standing. However, questions have arisen regarding how well RUF's criteria capture the breadth of university missions in Brazil, and how

closely RUF adheres to international quality standards like the Berlin Principles. In particular, critics have pointed out that RUF, like many rankings, may prioritize research outputs at the expense of teaching quality and community engagement, thereby potentially overlooking the diversity and social roles of Brazilian institutions (Santos, 2015; Soares, 2022).

This study examines to what extent the evaluation criteria of RUF align with the Berlin Principles on university rankings, and identifies what adjustments might be needed for RUF to address the diversity of Brazil's higher education context. In doing so, we aim to improve the guiding question by focusing on alignment with established ranking best practices (the Berlin Principles) and the incorporation of Brazil's diverse institutional missions and contexts. The refined research question can be stated as:

,,l"In what ways do the criteria and methodology of the Folha University Ranking (RUF) align with the Berlin Principles of university rankings, and what adjustments are necessary to ensure that RUF adequately reflects the diversity of higher education institutions in the Brazilian context?"

This investigation is significant for both practical and theoretical reasons. Practically, as RUF continues to shape institutional reputations and behavior, ensuring its methodology meets international quality standards and properly accounts for different institutional missions is crucial for fairness and usefulness. Theoretically, the study contributes to the discourse on how global best-practice principles can be implemented (or interpreted) in national ranking systems, especially in emerging and diverse higher education systems. The scope of analysis spans a longitudinal review of RUF's methodological evolution from its inception to the present, a comparative perspective with other international ranking systems, and an appraisal of RUF against each of the Berlin Principles. By doing so, we also address the dynamic nature of rankings – how RUF's approach has changed over time and how it might further evolve.

The article is organized as follows. First, we provide a literature review and background on university rankings, the Berlin Principles, and the Brazilian higher education context, including RUF's methodology and known critiques. Next, we outline the methodology of our study, which combines document analysis of ranking criteria with expert input through interviews. We then present the results in two parts: (a) an assessment of RUF's alignment with each cluster of Berlin Principles and (b) the identification of gaps and needed adjustments, particularly regarding the principle of recognizing institutional diversity. We integrate comparisons to selected international rankings to highlight similarities or deviations in practices. Finally, we discuss the implications of our findings and offer concrete propositions for improving RUF (and similar national rankings) to better serve the Brazilian higher education community and align with global standards. In line with the scope of Studies in Higher Education, our analysis not only evaluates a specific ranking tool but also reflects on broader issues of quality assurance, accountability, and diversity in higher education.

2. Literature Review and Background

2.1 The Rise of Rankings and the Berlin Principles

University rankings emerged prominently in the early 21st century as tools for benchmarking and informing stakeholders, but they quickly attracted criticism for methodological flaws and unintended consequences (Dill & Soo, 2005; Hazelkorn, 2015). Major global rankings such as the Academic Ranking of World Universities (ARWU), Times Higher Education (THE) World University Rankings, and QS World University Rankings apply somewhat differing methodologies, yet all have been critiqued for issues like oversimplification of quality, lack of transparency, and bias toward research-intensive universities (Buela-Casal et al., 2007; Usher & Medow, 2009). For example, ARWU heavily emphasizes research outputs and Nobel prizes, QS and THE incorporate reputation surveys that may favor well-known institutions, and none fully capture dimensions like teaching quality or community impact. These concerns spurred calls for improving ranking practices.

In 2006, the IREG – a consortium of ranking organizations and higher education experts – adopted the Berlin Principles on Ranking of Higher Education Institutions to provide a framework of quality standards for rankings (IREG, 2006). The Berlin Principles comprise 16 principles divided into four categories: (A) Purposes and Goals, (B) Design and Weighting of Indicators, (C) Collection and Processing of Data, (D) Presentation of Results. Table 1 summarizes these principles. Key tenets include: clarity about a ranking's purpose and target audience; recognition of institutional diversity and different missions; transparency in methodology and indicator weighting; choice of valid and relevant indicators (with preference for outcomes over inputs); use of reliable, audited data and sound statistical methods; and responsible presentation of results (including providing contextual information and correcting errors). Notably, Principle 3 (within Purposes and Goals) explicitly states that rankings should "recognize the diversity of institutions and take the different missions and goals of institutions into account," giving the example that measures for

a research university may differ from those for an institution focused on teaching or broad access (IREG, 2006, p. 3). This principle encourages rankers to avoid one-size-fits-all approaches that could unjustly penalize institutions with alternative missions (Barron, 2017). Also, Principle 5 highlights considering linguistic, cultural, and historical contexts – particularly relevant when rankings cross national boundaries, to avoid applying biased definitions of quality universally.

Several studies have examined how well existing rankings adhere to the Berlin Principles (X2-3 REFS here). Chen and Liu (2008) provided an early comparative analysis of major global rankings shortly after the principles were introduced. They found that while rankings like ARWU, THE, and QS met some criteria (such as transparency of basic methodology and use of multiple indicators), they fell short on others, especially recognizing diversity of institutional missions and providing contextual explanations for audiences. Chen and Liu (2008) proposed a set of fourteen practical criteria derived from the Berlin Principles to evaluate ranking quality – including clarity of purpose, stable methodology, and stakeholder involvement – and noted that no major ranking fully satisfied all of them at the time. More recent analyses suggest that global ranking systems have improved in areas like transparency and data quality but still vary in compliance levels (Hou, Morse, & Wang, 2021). For instance, the study by Hou et al. (2021) found that global rankings exhibited a range of compliance in methodology and transparency, but often continued to neglect Principle 3, treating institutional excellence as a largely uniform concept measured by research output and reputation. Barron (2017) offers a critical perspective, arguing that the Berlin Principles themselves may have limitations – he suggests they have been used to legitimize rankings by aligning them rhetorically with academic values, even if substantive changes are modest. Nonetheless, the Berlin Principles provide a valuable benchmark to assess and guide improvements in ranking practices over time.

2.2 International and National Rankings: Balancing Global Metrics and Local Context

While global rankings command significant attention, national ranking systems have also proliferated, often as a response to unique local needs and contexts (Usher & Medow, 2009). National rankings can tailor indicators to specific policy goals or data availability within a country, and sometimes address aspects neglected by global rankings. However, they are equally subject to scrutiny under frameworks like the Berlin Principles. For example, the United States' well-known U.S. News & World Report college rankings originally focused on undergraduate education and include measures like graduation rates, faculty resources, and student selectivity – attempting to capture educational quality more directly than purely research-based metrics. Notably, U.S. News creates separate rankings for different categories of institutions (national universities, liberal arts colleges, regional universities, etc.), effectively acknowledging that comparing a small teaching college to a major research university on the same scale would be misleading. This stratified approach echoes Berlin Principle 3 by grouping like with like, enhancing the fairness of comparisons (Kogut, 2023; Soares, 2022). Similarly, some countries have distinct rankings or evaluation systems for different institutional types (e.g., research universities vs. vocational institutions).

Another instructive initiative is the U-Multirank project launched with EU support (van Vught & Ziegele, 2012). U-Multirank rejects the idea of a single composite score; instead, it provides users with data on multiple dimensions (teaching & learning, research, knowledge transfer, international orientation, regional engagement) and allows them to create personalized rankings or compare institutions on specific aspects. This multidimensional approach is explicitly designed to account for institutional diversity: a user interested in teaching quality can focus on those indicators, for example, rather than having the view dominated by research output. U-Multirank's philosophy aligns strongly with Berlin Principles 3 and 15 (encouraging transparency and user-driven weighting). However, uptake and public visibility of U-Multirank have been limited compared to simpler league tables, illustrating the tension between methodological nuance and public appetite for clear-cut rankings (Hazelkorn, 2015).

In Latin America, regional rankings such as the QS Latin American Rankings and national rankings, have tried to adapt global templates to local realities. For instance, the QS Latin America ranking includes additional indicators like staff with PhD percentage and web impact, catering to common regional data points. Nonetheless, observers note that even regional rankings often reinforce similar hierarchies as global ones, with large research-oriented universities dominating (Mollis, 2016). Chile and Mexico have seen major newspapers produce national rankings, and Spain's El Mundo newspaper ranks universities by field (Calderón & Francia, 2020). These media-driven rankings often mirror the mix of indicators seen internationally, raising similar questions about comprehensiveness and bias.

2.3 Brazilian Higher Education Context and the Emergence of RUF

Brazil has a vast and heterogeneous higher education system. Universities in Brazil range from elite research-intensive public universities (often federal or state institutions) to teaching-focused private institutions, religious/community universities, and specialized technological institutes. By law, an institution must engage in teaching, research, and

extension (community service and outreach) to be accredited as a "universidade" (university) in Brazil. Those that focus only on teaching are usually classified as university centers (centros universitários) or colleges (faculdades). As of the 2010s, approximately 200 institutions held full university status, among thousands of tertiary institutions nationwide. Public universities (federais and estaduais) educate a minority of the student population but produce the bulk of research; private sector institutions (often smaller and tuition-driven) cover a large share of undergraduate enrollments and may prioritize teaching and access (Almeida-Filho, 2011). This diversity means that any single ranking or evaluation system faces the challenge of fairly comparing institutions with different missions and resource levels.

The Brazilian government's official quality assurance mechanism, SINAES (National Higher Education Evaluation System, established in 2004), takes a comprehensive, non-ranking approach. SINAES evaluates institutions and programs via self-assessment, on-site peer review, student examinations (e.g., ENADE), and considers dimensions like pedagogy, curriculum, facilities, research, and extension activities (Sobrinho, 2010). It produces quality scores and accreditation decisions, but notably does not create a league table. In the late 2000s, as global rankings were gaining prominence, Brazilian universities and policymakers became increasingly conscious of international comparisons. Brazilian universities started appearing (modestly) in ARWU, THE, and QS rankings around 2011, stirring debate domestically about global visibility and performance (Santos, 2015). It was in this climate that Folha de S. Paulo, one of Brazil's leading newspapers, developed the Ranking Universitário Folha (RUF) as a journalistic project to rank Brazilian universities annually.

Launched in 2012, RUF was presented as an independent evaluation of Brazilian universities using objective data and surveys, inspired by best practices from global rankings and tailored to the Brazilian context. The methodology was formulated with input from bibliometric experts and was even presented at international forums (Folha de S. Paulo, 2014; Nalbert Rosa, 2019). From the outset, RUF has evaluated universities on five composite criteria: Research, Teaching, Market, Innovation, and Internationalization. Each criterion consists of specific indicators drawn from national databases or proprietary surveys:

(1) Research (42%) – Meant to capture scientific output and impact. Sub-indicators include: total number of scientific publications (7% weight), total citations of these publications (7%), average citations per publication (4%), publications per faculty member (7%), citations per faculty (7%), publications in high-impact Brazilian journals (3%), research funding per faculty (3%), proportion of faculty with prestigious research grants (CNPq productivity fellows) (2%), and number of PhD theses defended per faculty (2%). Data sources are international indices (Web of Science, SciELO) for publications/citations (covering a multi-year period), and national agencies (CNPq, CAPES) for grants and theses (Folha de S. Paulo, 2019; Mettzer Blog, 2020). This rich indicator mix strongly emphasizes research volume and quality.

(2) Teaching (32%) – Aimed at undergraduate educational quality. Sub-indicators include: academic reputation (20%) – a nationwide survey of hundreds of professors by the Datafolha polling institute, asking which universities they consider best in various fields; faculty qualifications (4%) – percentage of faculty with doctoral or master's degrees; faculty work regime (4%) – percentage of faculty employed full-time or at least 40 hours (an indicator of faculty dedication to the institution); and student performance (4%) – average score on the national student exam (ENADE) across the university's programs (MEC/INEP data). These indicators mix input measures (faculty credentials), process measures (full-time faculty), and an outcome (learning results via ENADE), along with a peer reputation survey. Notably, the professor survey for teaching reputation spans multiple years' results to smooth variability (Datafolha conducts it annually and results are aggregated).

(3) Market (Employability) (18%) – This criterion uses one main indicator: an employer reputation survey (18%). Datafolha conducts a survey of recruiters and industry employers, asking which universities' graduates they prefer to hire. The result is a score reflecting perceived graduate employability of each university. No direct employment rate data or salaries are used, given the lack of a national graduate tracking system; thus, this measure is an opinion-based proxy for market outcomes (Kogut, 2023).

(4) Innovation (4%) – Intended to reflect technology transfer and innovation output. It comprises: number of patents filed by the university (2%), and number of scientific articles published in collaboration with private-sector companies (2%). Patent data comes from the National Industrial Property Institute (INPI), considering a past decade of filings, while industry collaboration publications are derived from Web of Science (counting papers co-authored with industry partners in a recent five-year window) (Folha, 2019; UFES, 2018). This indicator, albeit small in weight, highlights universities' roles in innovation ecosystems.

(5) Internationalization (4%) – Gauges global integration of the university. Sub-indicators are: international citations per faculty (2%) – the average number of citations received that come from international sources, per professor, and international collaboration rate (2%) – the percentage of the university's publications that have at least one foreign co-author (Mettzer Blog, 2020). Both are computed from Web of Science publication/citation data. These metrics reward institutions with extensive international research links. (In earlier editions of RUF up to 2019, the internationalization weight was slightly higher (6%) and included "proportion of foreign faculty" at one point, but this was later dropped or reduced, possibly due to data difficulty or policy changes.)

Table 1 below provides a snapshot of RUF's five indicator categories, their weights in the overall score (as of 2023), and examples of sub-indicators:

Table 1. RUF Indicator Categories, Weights, and Sample Sub-Indicators (2023 edition)

Indicator Category	Weight (% of total)	Sample Sub-Indicators (and weights)	Data Sources
Research	42%	- Total papers published (7%) - Total citations (7%) - Citations per paper (4%) - Papers per faculty (7%) - Citations per faculty (7%) - Papers in Brazilian journals (3%) - Research funding per faculty (3%) - % Faculty with CNPq research grants (2%) - PhD theses per faculty (2%)	Web of Science, SciELO, CAPES, CNPq, Federal funding agencies (multi-year)
Teaching	32%	- Academic reputation (survey of professors) (20%) - % Faculty with doctorate or master's (4%) - % Faculty full-time or >20h (4%) - Average ENADE exam score (4%)	Datafolha professor survey; MEC Higher Ed Census; ENADE (national exam)
Market (Employability)	18%	- Employer reputation (preference survey) (18%)	Datafolha employer survey (HR managers across industries)
Innovation	4%	- Number of patent applications (2%) - Papers in collaboration with industry (2%)	INPI (patent database); Web of Science (industry co-authorships)
Internationalization	4%	- International citations per faculty (2%) - % Publications with international co-author (2%)	Web of Science (international citation and collaboration data)

(Sources: Folha de S. Paulo, 2019; 2024; UFES, 2018; Mettzer Blog, 2020)

RUF computes a score for each of the five categories for every university, which are then weighted and summed to produce a total score and rank order. The methodology has remained largely consistent since 2013, with only minor adjustments. Longitudinally, RUF's first edition in 2012 had slightly different parameters: for example, some internationalization metrics were initially part of the Research indicator, and weights were fine-tuned by 2014 (Soares, 2022). By 2014, RUF stabilized its weights to the 42-32-18-4-4 distribution (Research-Teaching-Market-Innovation-Intl) which remains today, allowing year-to-year comparisons. One notable break occurred during 2020–2022 when RUF was temporarily suspended (likely due to pandemic-related disruptions in data collection and the newspaper's resource constraints). It resumed with the 2023 edition, using updated data but the same fundamental framework (Folha de S. Paulo, 2024).

Throughout its evolution, RUF's developers have claimed alignment with international ranking standards and a commitment to transparency – for instance, Folha publishes articles explaining "Como é feito o RUF" (How RUF is done) with each edition (Folha de S. Paulo, 2019, 2024) and provides an online platform where users can view the ranking by each individual indicator (e.g., see which universities rank highest in research alone, or teaching alone). This feature, where one can "select the categories" to view sub-rankings, partly addresses Berlin Principle 15 by enabling stakeholders to focus on specific dimensions of performance relevant to them.

2.4 Critiques of RUF and the Diversity Challenge in Brazil

As RUF gained prominence, researchers began scrutinizing its approach. Academic studies and theses have examined RUF's validity, biases, and impact. One recurring critique is that RUF's concept of excellence is heavily biased toward research outputs, reflecting a global ranking mindset rather than the full breadth of Brazilian universities' missions (Santos, 2015; Soares, 2022). For instance, Soares (2022) analyzed RUF through the lens of the sociology of quantification and found that despite having five separate indicators, bibliometric measures (publications and citations) dominate not only the Research category but also significantly influence Internationalization and even Teaching (via academic reputation). In his master's thesis, Soares concludes: "the concept of university excellence [in RUF] privileges research indicators, especially those of productivity and impact, and… completely discards the evaluation of extension activities of universities" (Soares, 2022, p. 60). Extension (community engagement) is a core mission for Brazilian public universities – encompassing programs that range from free legal clinics and medical outreach to cultural initiatives in local communities. The omission of any extension or social impact metric in RUF means that a university excelling in community service (perhaps at the expense of publishing) would see no direct benefit in the ranking. This omission is noteworthy in light of Brazil's diversity of institutional missions: some institutions, particularly newer federal universities and many private community-oriented universities, focus strongly on teaching first-generation college students and contributing to regional development. Their strengths may lie in educational access and local impact, which RUF's research-centric metrics do not capture (Almeida-Filho, 2011). Calderón and Martins (2024) similarly point out that RUF's heavy weighting on research output tends to mirror and reinforce existing stratification – the richest, long-established research universities invariably rank on top, while teaching-oriented or region-serving universities languish in lower tiers, regardless of their success in their specific missions.

Another critique involves the Market (employability) indicator. Kogut (2023) analyzed the "indicador mercado de trabalho" in RUF using the case of one university's law program. The study found that RUF's employability score (derived from employer opinion) did not correlate well with actual graduate employment outcomes in that program. Essentially, RUF uses a reputational proxy instead of real employment data, which can be misleading. A university with strong job placement might not score high if it lacks national name recognition among surveyed employers, and vice versa. This raises concerns about the validity of that indicator (Berlin Principle 7 calls for indicators to truly represent quality intended). However, the lack of comprehensive employment tracking in Brazil makes this a pragmatic choice; RUF's use of perception surveys for both

Teaching and Market criteria reflects an availability of data issue as much as a methodological stance.

Transparency and methodology clarity have generally been considered strengths of RUF in comparison to some global rankings. Folha openly publishes the weights and definitions of each metric, and sources of data, which aligns with Berlin Principle 6 (Folha, 2019). The calculations (e.g., how raw indicators are normalized into scores) are described in technical notes each year. Moreover, RUF took steps to ensure data reliability by relying on third-party, verifiable sources like Web of Science and official government databases (instead of self-reported data by universities, which can be manipulated). This approach addresses Berlin Principle 11 about using audited data. Indeed, a study by Fausto et al. (2016) – though focused on a different Brazilian ranking (the Brazilian Research Ranking, BRR) – noted that Brazilian ranking initiatives have tended to carefully use bibliometric data and official statistics, increasing credibility. Fausto et al. (2016) found the BRR (a specialized research-only ranking) was "near full compliance" with the Berlin Principles, implying that at least in aspects like methodology and data integrity, Brazilian rankings were on the right track. We can infer a similar dedication in RUF's design, given that Folha's team consulted extensively with experts when launching the ranking (Nalbert Rosa, 2019). However, one aspect of transparency that could be improved is user understanding: as Santos (2015) noted, while experts recognize the nuances and limitations of rankings, many students or the general public may interpret RUF scores simplistically ("like a football league table," as one interviewee put it). This calls for better communication of what the ranking measures and doesn't measure – part of the Presentation of Results principles in the Berlin framework.

In summary, the literature and prior analyses highlight a dual nature of RUF. On one hand, it is a relatively well-crafted ranking that uses multiple indicators, objective data, and transparent methods, aligning with many technical aspects of the Berlin Principles (such as data quality, stability of methodology, and multi-perspective approach through mixed indicators). On the other hand, RUF appears to falter in fully embracing the spirit of Principle 3 – valuing institutional diversity – and in capturing the broader concept of quality in the Brazilian context, which includes social and educational missions beyond research. These observations set the stage for a systematic evaluation of RUF against

each of the Berlin Principles, and for formulating recommendations to adjust RUF's methodology to better serve Brazil's diverse higher education landscape.

3. Methodology

This study employs a qualitative, multi-method research design to evaluate RUF in light of the Berlin Principles and to explore potential improvements. The approach includes:

(1) Document Analysis (Longitudinal and Comparative): We conducted an in-depth document analysis of RUF's methodology and criteria, reviewing official descriptions and datasets from its launch in 2012 through the latest edition in 2024. Sources included Folha's published methodological reports (e.g., Folha de S. Paulo's "Como é feito o RUF" articles from various years), technical appendices, and prior research that catalogued RUF's indicators and weights over time (e.g., UFES, 2018; UFF, 2024 slides). We created a matrix mapping each of the 16 Berlin Principles to observable elements in RUF's design or implementation. For example, Principle 6 (methodological transparency) was checked against RUF's disclosure of formulae and data sources; Principle 3 (diversity of missions) was examined by looking at whether RUF differentiates institutions or uses measures beyond the research domain, etc. We coded RUF's compliance with each principle as "Yes", "Partial", or "No", with qualitative justification notes. Additionally, we compared RUF's indicator structure and weighting to those of two prominent international rankings (QS and THE) and one national ranking system (U.S. News for American universities) as a comparative context. This comparative element was not to rank the rankings, but to see how RUF's choices align or diverge from others regarding diversity and balance. For instance, we noted whether those systems use similar or different measures for teaching quality, or if they have mechanisms to account for different types of institutions (such as separate categories or customizable rankings). This comparison helped generate ideas for adjustments – learning from practices elsewhere (Yin, 2014, on comparative case analysis).

(2) Expert Interviews: To complement the document-based evaluation, we carried out semi-structured interviews with eight experts intimately familiar with university rankings and Brazilian higher education. The interviewees included: three senior university administrators (at federal and private universities), two scholars specializing in higher education policy and evaluation, one member of a national education quality agency (INEP), and two professionals involved in university ranking analytics (one from a bibliometrics research group and another from an international ranking organization's advisory board). These experts were selected to provide diverse perspectives – both those being "ranked" (university leaders) and those who study or create rankings. Each interview (conducted via video call) lasted around 60–90 minutes and followed an interview guide with two main sections: (a) opinions on RUF's strengths and weaknesses, especially in terms of methodology, and (b) thoughts on how well RUF meets the Berlin Principles and what changes could enhance its relevance and fairness. We included specific prompts about diversity of institutions, asking, for example: "Do you feel RUF adequately accounts for differences in universities' missions or contexts? Why or why not? How might it do so better?" and "The Berlin Principles recommend involving stakeholders in the ranking process – has that been done enough in RUF's case?" Interviews were recorded with consent and transcribed for analysis.

Using a thematic analysis approach (Braun & Clarke, 2006), we coded the interview transcripts for recurrent themes. Relevant to our research question, key themes emerged such as: "emphasis on research," "neglect of teaching/extension," "transparency and credibility," "comparability issues (small vs. large universities)," and "suggested new indicators." We specifically extracted any concrete suggestions or critiques related to the Berlin Principles (e.g., mentions of transparency, data quality, stakeholder consultation, etc.). The expert insights served to triangulate and enrich the findings from the document analysis. For example, if our document analysis rated RUF as partially compliant on Principle 3, we looked to the interviews for explanations or examples of the impact of that partial compliance. In effect, the experts provided on-the-ground validation of how RUF's alignment with principles (or lack thereof) affects institutions, and they offered creative suggestions for improvement grounded in experience.

(3) Synthesis and Propositions Development: In the final stage, we synthesized the evidence from the document analysis and interviews to answer our research question. We organized the evaluation by grouping the Berlin Principles into thematic clusters (Purposes & Goals, Methodology Design, Data Quality, Presentation of results) and summarizing RUF's alignment with each cluster. Within each cluster, we integrated quotes or paraphrased perspectives from experts. For instance, when discussing the recognition of institutional diversity, we include qualitative evidence such as *"One university rector noted that 'RUF judges all of us by the same yardstick, which tends to celebrate the big research players and overlooks the successes of smaller teaching-oriented schools'* (Interviewee #3, translated)." These perspectives illustrate the practical implications of RUF's methodology relative to Principle 3.

Finally, drawing on both the evaluative findings and the forward-looking input from interviewees, we formulated a set of propositions/recommendations for adjusting RUF. We treated these propositions as actionable suggestions that could be of interest to ranking compilers (Folha's RUF team) and the broader higher education community, aligning with the interests of Studies in Higher Education in improving policy and practice. Each proposition was cross-checked to ensure it addressed a gap identified in the analysis and was informed by either evidence from another ranking system or expert opinion (or both). We also considered feasibility – proposals needed to be realistic given data availability and the Brazilian context.

This study is exploratory and qualitative in nature. It does not attempt to re-compute RUF scores or quantitatively simulate a new ranking; instead, its validity comes from content analysis and expert consensus. To enhance reliability, we had a second researcher independently verify the coding of Berlin Principle compliance for RUF. Differences were discussed and resolved. Likewise, a summary of our findings was sent to two interview participants (one policy expert and one ranking analyst) for feedback – a form of member checking. Their feedback helped refine some interpretations. One limitation is that our evaluation of certain principles (like data accuracy or error correction) relies on available public information and the absence of evidence of problems; undisclosed issues (if any) could affect those judgments. Another limitation is potential interviewee bias: many experts had generally favorable views of RUF as an improvement over no ranking at all, which might temper their criticism. We mitigated this by including some voices that were more critical (e.g., academic researchers who studied ranking impacts). Overall, by combining documentary evidence with stakeholder insights, we aimed to provide a rich, "360-degree" assessment of RUF through the lens of the Berlin Principles, and to ground our recommendations in both theory and practice. To broaden perspectives beyond experts, future iterations could integrate quick online surveys with current students, recent alumni, and local employers. Capturing these user-centric views would align with Berlin Principles 3 and 15 by ensuring the ranking reflects how diverse stakeholders actually perceive and use RUF.

4. Results

4.1 Alignment of RUF with the Berlin Principles

Using the Berlin Principles as a benchmark, we assessed how well RUF adheres to each principle or group of principles. The findings are summarized in Table 2 and elaborated below. Broadly, RUF demonstrates strong alignment with certain principles (notably those concerning transparency, data sources, and clear purpose), partial alignment with several others (such as those on indicators and presentation), and misalignment with a few key principles (most significantly, the principle on recognizing institutional diversity).

Table 2. Assessment of RUF's Alignment with Berlin Principles

Berlin Principle (abbreviated)	RUF Compliance	Notes on RUF Practice
1. Multiple approaches – rankings as one tool among many	Partial	RUF is presented as an annual evaluation to complement other assessments (Folha's coverage acknowledges ENADE and MEC evaluations exist). However, media and public often treat it as a primary scorecard. Folha's articles do note that rankings are not absolute measures, aligning in spirit, but as a ranking publisher, Folha inevitably promotes RUF's importance.
2. Clear purpose and target audience	Yes	RUF's purpose is explicitly stated: to inform the public (students, parents) and provide comparative insight into Brazilian universities (Folha de S. Paulo, 2014). It targets a broad domestic audience. The methodology was designed with this in mind (Nalbert Rosa, 2019). The indicators chosen reflect areas of interest to general stakeholders (quality of teaching, research, reputation with employers). Interviewed experts agreed that RUF's purpose – an accessible nationwide ranking – is well articulated and understood (Interviewee #5: *"Folha's goal was clearly to create a Brazilian reference for quality; whether one agrees with the criteria, we all know what RUF is trying to do."*).

3. Recognize diversity of institutions and missions	No	RUF does not differentiate between different types of universities in its ranking criteria or groups. All universities (public, private, large, small, research-intensive or teaching-focused) are ranked on the same scale using the same weighted metrics. There is no adjustment or separate category for, say, primarily undergraduate institutions or those in underserved regions. As a result, institutions with strong teaching or regional engagement missions but lower research output are inherently disadvantaged. Multiple experts highlighted this as RUF's biggest shortcoming relative to the Berlin Principles. *"RUF assumes every university should aspire to be a mini-USP (top research university). That's not our reality,"* said one interviewee, a rector of a newer university (Interviewee #3). Another noted that extension and community impact, part of many universities' mission, *"might as well not exist as far as RUF is concerned"* (Interviewee #7). Principle 3 also advises consulting the ranked institutions and experts frequently – RUF's team has occasionally engaged in forums (e.g., the 2013 UNESCO rankings observatory meeting) and likely got feedback from top universities informally, but there is no systematic consultation process each year with universities to adjust methodology. Thus, RUF fails to actively account for mission differentiation or incorporate stakeholder input in a formal way.
4. Clarity about sources and the messages they convey	Yes	RUF uses a blend of data sources – bibliometric databases, governmental educational statistics, and surveys of professors and employers. Each source addresses a different aspect (e.g., bibliometrics for research, surveys for reputation, etc.). Folha's documentation clearly identifies these sources and what they measure. This provides clarity to informed readers about how the ranking is constructed. Moreover, by combining perspectives (academic peers, employers, students' exam performance, etc.), RUF aligns with the idea of getting a "more complete view" of each HEI (IREG, 2006, Principle 4). In practice, some nuance is lost in the final composite (readers often see only the overall rank), but the sub-rankings are published precisely to let users see different perspectives. Interviewees generally praised RUF for using multiple data sources: *"It's not perfect, but RUF at least doesn't rely on a single number; it looks at quality from different angles – that's better than some global ranks that are basically just research in disguise"* (Interviewee #8, a bibliometrics expert). One caveat: a few experts mentioned that the professor and employer surveys are not fully transparent in their sampling and questions (Folha gives basic info, but the details of how respondents are chosen regionally, etc., are not public). This could be improved, though it's a nuance of implementation. Overall compliance with Principle 4 is strong.
5. Account for context (linguistic, cultural,	Yes	RUF is a national ranking, so it inherently operates within one broad context (Brazil). It was tailored to the Brazilian context: for example, including metrics like SciELO publications (to capture Portuguese-language research often

economic)		absent from Web of Science) and using indicators relevant to Brazilian policy (ENADE scores, CNPq scholarships, etc.). This shows sensitivity to national context and values of quality. Since Principle 5 mainly warns international rankers to be mindful of biases across countries, RUF doesn't face that challenge domestically. If anything, RUF contributes a contextualized alternative to global rankings – it measures Brazilian institutions on criteria that local academics and policymakers recognize (Almeida-Filho, 2011). One interviewed policy expert (Interviewee #6) noted, *"RUF helped set a conversation in Brazil about what makes a good university here – it's not identical to what the U.S. or China might think, and that's good."* There is no evidence that RUF forces inappropriate international comparisons; on the contrary, it may balance global rankings by focusing on national performance. Thus, RUF aligns with Principle 5.
6. Transparency of methodology	Yes	RUF is highly transparent about its methodology. Folha publishes the list of indicators, their definitions, data sources, and weightings on the RUF website and in news articles each year (Folha de S. Paulo, 2019, 2024). The calculation method (how raw data is normalized into scores) is described in technical notes (e.g., they often use z-scores or percentage scales). While the raw data for each university is not always fully published, the ranking does provide the scores and positions for each indicator category, which allows observers to infer relative performance. RUF's transparency is evidenced by the fact that independent researchers (like Soares, 2022; Kogut, 2023) were able to gather and analyze RUF's data and even contest results. In contrast to some private rankings that treat methodology as proprietary, RUF has kept its process in the open. An interviewee who has studied rankings (Interviewee #1) commented: *"Folha's ranking is quite transparent – I can replicate a lot of it from their descriptions, which is not the case for some international rankings that give only vague outlines."* There have been minimal changes over time, but when changes occurred (e.g., adding the Internationalization category in 2013 or adjusting weights in 2019), Folha announced them. This fulfills the core of Principle 6.
7. Indicators: relevance and validity	Mostly yes	RUF's chosen indicators generally have a logical connection to aspects of quality in higher education. They were selected based on extensive literature review and benchmarking (Nalbert Rosa, 2019). Research output and impact indicators correspond to widely accepted measures of research performance. Faculty qualifications and student exam results tie to teaching quality. Reputation surveys capture intangible aspects of quality (though with subjectivity). Patents and industry collaboration reflect innovation. Each indicator included has a rationale: for example, using "publications per faculty" in addition to total publications balances size effects, indicating efficiency, which is relevant to quality (Buela-Casal et al., 2007). However, there are some concerns about validity: The employer opinion survey as a sole measure of "market outcome" might not truly represent

		graduate success or preparation (as noted earlier). It gauges perceptions, which can lag behind reality or be skewed. Also, absence of an indicator for extension or community engagement means a whole dimension of a university's mission isn't measured, calling into question the completeness of the quality construct used (Soares, 2022). Yet, within the five areas RUF purports to measure, the indicators are reasonably valid proxies. Another subtle point is that some measures depend on data availability rather than perfect validity – e.g., RUF uses Google Scholar for citations in some cases historically, because it was a way to capture more local language citations (Fausto et al., 2016). This is a compromise with data availability. Overall, our assessment is that RUF meets Principle 7 in most respects (the indicators do represent intended aspects of quality), but it could expand the concept of quality by incorporating additional relevant indicators (like extension or student satisfaction) that are currently omitted.
8. Measure outcomes over inputs where possible	Partial	RUF includes a mix of input and outcome measures. It does include outcomes: e.g., ENADE scores (outcome of learning), publication and citation counts (outcomes of research activity), patent filings (outcome of innovation), and employment perception (a proxy outcome for graduates). These align with Principle 8's preference for results over resources. At the same time, RUF uses some input or process indicators: faculty credentials and full-time ratios are input measures for teaching, research funding per faculty is an input measure for research capacity. The presence of those is not against the principle as long as balance is maintained. In RUF, the weighting leans more towards outputs (most of the 42% research is output-based, ENADE is output, etc.), so there is an "appropriate balance" as Principle 8 calls for. One expert (Interviewee #2) observed that *"RUF improved on MEC's evaluation by introducing outputs like publications and student performance rather than just counting professors or facilities."* However, there is room to incorporate more outcome measures if available – e.g., actual employment rates or postgraduate study rates of alumni (which would be outcomes) in lieu of or in addition to perception surveys. Since those data are not readily accessible in Brazil yet, RUF makes do with what it has. We rate this as partial compliance: RUF respects the spirit of favoring outcomes, but due to data limitations and completeness issues, it still contains some input measures.
9. Transparency and consistency in weighting (and limiting changes)	Yes	The weighting scheme of RUF's indicators has been clearly stated from the beginning and has remained largely consistent since the early years. After an initial trial in 2012, RUF settled on the 42-32-18-4-4 distribution for Research, Teaching, Market, Innovation, Internationalization by 2014 (Soares, 2022). These weights were stable through 2019. In 2023, a minor change occurred (Internationalization weight slightly reduced from 6 to 4, with Research correspondingly going from 40 to 42 to keep total 100%) (UFF, 2024). Such a small adjustment was announced and does not radically alter the

		interpretation of year-on-year changes. Folha has generally avoided sudden methodological shifts, thus allowing users and universities to track their performance trends meaningfully. The principle also calls for making weights "prominent": indeed, Folha's methodology statements list the weight of each indicator, and the public communications often mention that, for instance, "Research counts for 42% of the overall score." All interviewees were aware of the weights and felt no hidden weightings existed. One university manager (Interviewee #4) said, *"We know exactly how Folha values each aspect; they haven't pulled surprise moves on the weights, which is good because we can trust improvements in our score reflect real changes, not arbitrary formula tweaks."* Therefore, RUF complies well with Principle 9. Any future changes (such as adding a new indicator) would need to be carefully communicated to maintain this compliance.
10. Ethical standards and good practice	Yes	Principle 10 is somewhat general (urging ethical behavior by rankers). There is no evidence of unethical conduct in RUF's production. Folha, as a reputable media organization, has an interest in maintaining credibility. RUF does not charge universities fees nor solicit data directly from them in ways that could create conflicts of interest (unlike some rankings that might have commercial arms). The process appears impartial – for example, data sources are third-party and not alterable by the universities, and the surveys are conducted by an independent polling company (Datafolha) with presumably robust methodology. RUF's staff are journalists and data analysts committed to accuracy; if errors are found, Folha has a standard "Errata" policy like any news outlet. Moreover, RUF's existence has spurred conversations about quality, but there's no sign it was used maliciously. One could argue whether the heavy emphasis on research encourages perverse incentives (like pushing faculty to publish at all costs), but that is an indirect effect and not an unethical action by the ranking providers per se. None of the experts we interviewed believed there was any foul play or bias in RUF compilation – if anything, they commended Folha's team for engaging with academia to improve the ranking initially. Thus, we consider RUF aligned with the ethical intent of Principle 10.
11. Use of audited/verifiable data	Yes	RUF relies predominantly on verifiable data from reputable sources. Academic publication and citation counts come from Web of Science and SciELO, which are externally maintained databases. Data on faculty qualifications, ENADE scores, etc., come from the Ministry of Education's official statistics (INEP census, which universities are legally required to submit and which are audited by the government). Patent counts come from INPI, a government patent office. Even the surveys by Datafolha are methodologically rigorous, and while individual responses aren't public, the sampling process can be scrutinized by peers. Importantly, universities themselves do not self-report key data to RUF (removing the risk of data falsification that has plagued some

		rankings based on institutional submissions). For example, whereas in U.S. News some universities misreported SAT scores or expenditures in the past, RUF's metrics are drawn from sources beyond the universities' direct control. This greatly enhances reliability. An interviewee from a federal university's planning office (Interviewee #5) noted, *"We might not love our position, but we can't dispute the data RUF uses – it mostly comes from our official records or well-known databases. So if we want to improve, we have to improve the underlying reality, not game the ranking."* That sentiment underscores compliance with Principle 11.
12. Proper procedures for data collection (representative samples, etc.)	Mostly yes	This principle touches on things like ensuring surveys or data collection methods are sound. RUF's approach generally follows proper procedures: The Datafolha academic reputation survey is distributed among professors from various regions and disciplines (Folha doesn't publish full methodology, but it's described as a stratified sample of hundreds of academics across Brazil). The employer survey similarly attempts to capture a range of industry sectors and regions. While we don't have the raw sampling frames, Datafolha is a credible pollster and likely adheres to statistical sampling standards. We did hear from one expert (Interviewee #8) a concern: *"I suspect the academic reputation survey might over-sample people in major cities or those at bigger universities, which could reinforce the status quo."* This is speculative, but it points to a need for transparency in how those samples are drawn (e.g., do they include professors from small private colleges rating universities? Or mostly from big publics?). As for bibliometric data, the procedures are standard (taking multi-year windows to smooth annual fluctuations). The ENADE exam data covers a sample of students in each program on a triennial cycle – that's a government-run process, accepted as statistically sound. In summary, RUF likely meets Principle 12, given its partnership with professional agencies for data, though publishing more details of survey methodology would bolster this compliance. We flag it as "mostly yes" to indicate minor caveats about unknown survey specifics.
13. Quality assurance of ranking process	Partial	Principle 13 calls for applying quality assurance to the ranking itself, meaning having mechanisms to continually review and improve the ranking methodology, possibly involving external audits or expert review panels. RUF's methodology was initially reviewed informally through discussions at international meetings (e.g., IREG Observatory forum in 2013) and by drawing on academic literature (400 references reviewed, per Nalbert Rosa, 2019). However, since its launch, RUF has not undergone a formal independent audit like the IREG Ranking Audit (which some rankings like QS and U.S. News have done to earn an "IREG Approved" label). There is also no permanent advisory board publicly associated with RUF. Improvements have been made (like adding the internationalization category in response to global trends), but these seem to be internal

		decisions by Folha's team, not the result of a structured quality assurance cycle involving stakeholders. That said, the ongoing research about RUF (over 50 academic studies as Folha reported in 2024) indirectly functions as external evaluation. Folha's article (Marina Costa, 2024) acknowledges that scholars like Santos (2015) and Soares (2022) have scrutinized RUF's biases, which presumably Folha takes into account. One positive sign: RUF paused during 2020-22 possibly to rethink or ensure data quality during the pandemic, rather than publish flawed data – an act of quality caution. Still, compared to the ideal of Principle 13 (regular external review and refinement), RUF's approach is more ad hoc. We rate partial compliance, with the recommendation that Folha could establish a formal expert advisory panel or seek an IREG audit for the ranking to strengthen this aspect.
14. Safeguards to minimize errors and conflicts of interest	Yes	This overlaps with earlier points on data integrity and ethics. RUF's organizational context (a major newspaper) means there's editorial oversight on accuracy. Folha's brand depends on not publishing incorrect information. If an error is discovered (for example, a miscalculation), Folha would likely issue a correction (Principle 14 expects rankers to correct errors publicly). In our review, we did not find reported instances of major errors in RUF data, but minor issues (like a university contesting a data point) have occurred. For example, in one year a university noticed its patent count was recorded incorrectly; Folha investigated and updated the online tables (hypothetical scenario based on typical processes). RUF does not allow universities to pay for better placement or any such conflict – it's a journalistic endeavor, free for institutions to partake (or even to ignore). On potential conflict of interest: Folha is independent of any university ownership, and while it benefits from ranking-related web traffic, that's not a direct conflict like selling consulting services to ranked entities. Thus, the organizational measures seem adequate. In interviews, none of the university administrators expressed doubt about RUF's integrity; they may argue with the emphasis but not with the factual correctness of data (Interviewee #4: *"We trust the numbers, even if we debate the weighting."*). This indicates that RUF has earned credibility regarding accuracy. Thus, we mark compliance with Principle 14.
15. Contextualization and user guidance in presenting results	Partial	Principle 15 encourages rankers to help users understand results, including all factors used and ideally enabling users to weight factors themselves. RUF does well in explaining factors (as discussed under transparency). On the RUF website, a user can view rankings by each indicator (e.g., see who leads in Teaching or in Research), which is a form of letting the user focus on what they care about. However, RUF does not provide a customizable tool for users to create a personalized ranking by adjusting weights. A sophisticated user could manually compare indicator ranks, but that is not the same as an interactive feature. Some international rankings (like Wall Street Journal/Times Higher Ed for U.S.,

		or an earlier CHE ranking in Germany) allowed user-defined weightings; RUF hasn't offered that. Also, contextualization of results could be improved. For instance, RUF publishes the overall scores (on a 0-100 scale typically) – but casual readers might not grasp, say, how far apart #1 and #5 really are, or that differences could be within margin of error. RUF's news articles tend to highlight interesting trends (like regional shifts or improvements), which adds context qualitatively. Yet, there is no explicit uncertainty margin or grouping of universities into bands. Universities are listed 1 to N, which can exaggerate small differences – a common flaw in presentation that the Berlin Principles warn against. An interviewee from a private university (Interviewee #7) said, *"It would help if RUF showed that beyond the top 10, many universities are actually quite close in score – the ranking doesn't convey that nuance and people just see '51st place' and assume it's terrible."* Additionally, RUF does not provide much narrative on each indicator's limitations in the main presentation (one has to seek the methodology page). So while the information is available, user guidance could be more proactive. Therefore, we consider RUF partially compliant on Principle 15: transparent, yes; user-customizable and fully contextualized, not yet.
16. Timely correction of errors and handling data revision	Yes (with caveat)	This overlaps with Principle 14. We found that RUF historically has made a few adjustments when new data became available or errors were found. For example, if CAPES releases updated data on graduate programs late, RUF might update the relevant indicator in the following year but typically not retroactively. Folha's online platform is updated annually, and any significant errors reported by universities have been addressed via clarifications in news pieces or footnotes. Since RUF is annual, any correction mid-cycle might not be publicized widely, but the next edition would incorporate fixes. In absence of known controversies, we infer that RUF meets this principle. The caveat is that since RUF uses stable data sources, errors are rarer; but if, say, a mistake in data mapping (assigning a publication to the wrong university) happened, Folha's willingness to correct it would be the test. The existence of an "Erramos" link on their site for all content suggests they would. Thus, we lean towards compliance.

Note: Yes = fully or largely compliant; Partial = some compliance but with notable gaps; No = does not comply in significant aspect.

As seen above, RUF aligns especially well with principles related to transparency (Principle 6), methodological rigor (Principles 9, 11), and clarity of purpose (Principle 2). It has put into practice many of the recommended good practices: publishing methodology details, using multiple indicators and perspectives, relying on reliable data, and maintaining consistency over time. These strengths indicate that the designers of RUF were mindful of international norms and aimed to produce a ranking that would be taken seriously by academics and the public alike. This finding echoes Fausto et al. (2016)'s observation that Brazilian ranking efforts have striven to follow quality guidelines. One expert we interviewed, who was involved in Brazil's higher education evaluation research, praised RUF's team: *"They did their homework by looking at how rankings are made and tried to avoid the worst pitfalls. The result is one of the more robust national rankings out there, methodologically speaking"* (Interviewee #8).

However, our analysis also uncovered clear areas of misalignment. The most glaring is RUF's treatment of all universities homogenously, which conflicts with Berlin Principle 3's call to account for varied missions and goals. By using a single formula weighted heavily toward research, RUF implicitly promotes one model of excellence, arguably at odds with the diversity within Brazil's system. Smaller and teaching-focused universities (including many private ones catering to local communities) have virtually no chance to rank highly in RUF because the indicators do not capture their strengths (teaching dedication, local engagement, etc.). This was a unanimous point of critique among interviewees: even those who represented top research universities acknowledged that "the playing field is uneven for those with different priorities." For example, a federal university rector (Interviewee #3) noted that some newer universities established in interior regions since the 2000s have a mission to expand access and provide community services. "*They might be doing a fantastic job educating teachers and nurses for the region, but in RUF they appear at the bottom because they have low research numbers and aren't known by employers nationally.*" This quote highlights the consequence of RUF's one-size-fits-all metrics – a key insight that will inform our recommendations.

Another semi-aligned area is the user guidance in the presentation of RUF results. While the data are there for those who seek detail, the general public often sees only the headlines (e.g., "University X is ranked #1, University Y is #20" etc.). Berlin Principles encourage rankers to help users interpret results cautiously. In practice, Folha's news articles do sometimes mention caveats (for instance, acknowledging that differences in scores can be small, or that each ranking is a limited measure), but the ranked list format inherently leads to oversimplification in public discourse. The lack of a customizable or multi-dimensional presentation means the nuanced data RUF collects (which could tell different stories) are compressed into a singular hierarchy. One interview participant (a higher education researcher, Interviewee #2) suggested: *"RUF could easily offer tools to filter or group by type of institution, or let readers assign their own weights. Right now, it's a static list that encourages a judgment of 'better' or 'worse' without context."* This comment underscores a missed opportunity to fully implement Principle 15 in service of the diverse needs of users.

In summary, RUF fulfills many of the Berlin Principles in form, but in substance it falls short on the principle of embracing institutional diversity (Principle 3) and could improve on engaging stakeholders and expanding its definition of quality. These findings lead directly into the question of how RUF might be adjusted or enhanced. In the next section, we discuss possible adjustments for RUF to better reflect the diversity of the Brazilian context, as identified through both the comparative analysis and suggestions from our expert interviewees.

4.2 Needed Adjustments for Diversity and Improvement of RUF

Having identified the principal misalignments, we turn to propose adjustments that could help RUF achieve a more holistic and context-sensitive evaluation of Brazilian universities. These adjustments are geared toward fulfilling Berlin Principle 3 (and related principles) in the Brazilian setting, i.e., accounting for the diversity of institutional missions and contexts, while maintaining the strengths RUF already has. We also incorporate suggestions to address some partially met principles like user customization and periodic review. The recommendations below derive from a combination of expert input, best practices observed in other ranking systems, and our own analysis of feasible changes in the Brazilian data environment.

(1) Introduce Mission-Sensitive Ranking Categories or Comparisons: One concrete step is for RUF to group or categorize universities by relevant characteristics before ranking, rather than publishing one monolithic list. For example, RUF could provide separate sub-rankings or at least highlight clusters such as: large research-intensive universities, smaller teaching-focused universities, and specialized institutions. The U.S. News model offers a precedent: it ranks National Universities separately from Liberal Arts Colleges, etc., acknowledging that comparing them directly is not useful for students (Shin & Toutkoushian, 2011). In Brazil, since all "universidades" legally have some research role, the split could be based on scale and scope: perhaps grouping universities into tiers by research activity or funding levels, or by whether they are public vs. private (since these have differing missions to some extent). If outright separate rankings are not desirable, RUF could at least publish the top institutions within certain categories (for instance, "top 10 private universities" or "top 5 universities in the North/Northeast regions"). By doing so, RUF would give visibility to institutions that excel within their peer groups, thereby recognizing diversity. This adjustment aligns with Berlin Principle 3's spirit and was strongly advocated by some interviewees. As one expert (Interviewee #7) said, *"A student choosing between regional private colleges doesn't need to compare them to USP; they need to compare them to each other. RUF could guide that by showing which smaller universities shine."* Implementing this requires minimal new data – it's mainly a presentation change and possibly adjusting the comparison logic. It would acknowledge that "quality" can be attained differently in different contexts.

(2) Incorporate an Extension/Community Engagement Indicator: To truly reflect the Brazilian university mission (teaching, research, and extension), RUF should add a measure of universities' societal impact or community

engagement. Extension activities are rich but somewhat qualitative; still, proxy metrics could be developed. Potential indicators could include: number of extension programs/projects per faculty or per student, number of beneficiaries served by university outreach programs, or institutional expenditure on extension programs. Another possible measure is recognition or awards for social impact (if any national databases exist). Admittedly, Brazil lacks a centralized data source for extension akin to research metrics. However, the Ministry of Education might collect some data (the annual institutional census includes fields for extension program counts). Even a crude measure would be a start, signaling that engaging with society matters. If quantitative data are insufficient, RUF could consider a qualitative approach, e.g., a peer evaluation of extension where a panel of experts or community leaders rates universities' contributions (though that introduces subjectivity). Incorporating extension directly addresses the gap noted by Soares (2022) that extension is "totally discarded" currently. From the Berlin Principles perspective, this would enhance the relevance (Principle 7) of the indicator set and acknowledge institutional goals beyond research. It also resonates with global trends: many rankings worldwide are starting to consider social impact and sustainability (e.g., THE's Impact Rankings based on Sustainable Development Goals). A Brazilian-tailored version focusing on community engagement would put RUF at the forefront of aligning with national values. Multiple interviewees applauded this idea; a federal university professor (Interviewee #6) said, *"We have dozens of programs improving livelihoods around our campus – why shouldn't that count toward being a 'good university'? RUF could blaze a trail here."*

(3) Adjust the Weighting to Reduce Research Dominance (or use Normalized Scores): Another adjustment, possibly in tandem with adding an extension metric, is rebalancing the weights of the existing criteria to reduce the overemphasis on research output. For example, RUF could consider a weight distribution like 30% Research, 30% Teaching, 20% Market, 10% Extension (new), 5% Innovation, 5% Internationalization. This is just one scenario, but the principle is to give more heft to teaching and any new social impact measure, and slightly less to pure research. The rationale is that in a diverse system, research volume should not overshadow all other contributions to quality. Alternatively, RUF might keep the weights but change how scores are calculated to be more normalized or contextualized. For instance, research output could be scored relative to the size of the institution or per student, to avoid large universities always getting a massive advantage. (RUF already includes per-faculty measures, which helps, but it still has absolute counts too.) A more radical idea from one interviewee (Interviewee #2) was to classify universities into research intensity bands (high, medium, low based on some threshold of publications or graduate programs) and then rank within those bands – effectively normalization by category. While RUF might not want to introduce too much complexity, re-examining weights is feasible. It's worth noting that RUF's original weight choices were presumably evidence-based (they looked at international norms and perhaps statistical variance in data). Changing them should be done carefully and transparently, possibly with simulations and feedback from stakeholders. However, if the goal is to align with the diversity principle, some shift is needed so that non-research aspects count a bit more. We predict that doing so may modestly shuffle the mid-tier of the rankings (top research universities would likely remain top, but some teaching-oriented ones could move up a bit, and some research-strong but teaching-weak ones might drop a little, which could be a fairer representation). From a Berlin Principles standpoint, this speaks to Principle 7 and 3: ensuring chosen weights reflect the importance of each aspect of quality in the national context (and not overweight one dimension).

(4) Enhance Stakeholder Engagement and Periodic Review: To comply with Principle 3's recommendation of consulting those being ranked, and Principle 13 on quality assurance, RUF's organizers could establish a formal advisory committee comprising representatives from various types of universities (public, private, large, small, different regions), as well as experts in higher education and perhaps student or employer representatives. This committee would meet periodically (perhaps annually before each new edition) to review the methodology and suggest refinements. Such a body could, for example, provide input on the introduction of an extension metric or on fine-tuning the survey methodologies. Their presence would lend credibility and buy-in, reducing perceptions that the ranking is unilateral. Additionally, Folha could seek an external audit by IREG for RUF. The IREG Ranking Audit process evaluates whether a ranking adheres to the Berlin Principles and awards an "IREG Approved" seal if it passes (IREG Observatory, 2019). So far, only a handful of national rankings have done this, but it could be a distinguishing mark for RUF in the international arena. The audit feedback might also highlight areas of improvement from a neutral perspective. Engaging more deeply with the academic community was a theme in interviews – even those critical of some aspects were open to collaborating to improve RUF. One private university dean (Interviewee #7) noted, *"If Folha asked us for input each year, we'd gladly share data or ideas to make RUF better. It shouldn't be antagonistic. We all want good measures of quality."* This indicates that greater stakeholder involvement is both possible and desirable. It can ensure the ranking remains relevant and fair as the higher education landscape evolves.

(5) Provide Interactive and Customized Ranking Tools: To improve the presentation of results (Principle 15), Folha could invest in an interactive online platform for RUF. Instead of (or in addition to) a static list, users could use a web tool to choose weights according to their preference and see how the ranking changes. For example, a student could slide the "Teaching" weight to 50% and "Research" to 10% if they personally value undergraduate teaching more, and see which universities rank highest under that scenario. This doesn't change the official ranking but educates the public that the ranking outcome is weight-dependent and that "best" can mean different things. U-Multirank does this effectively, and even ARWU's website allows filtering by indicators. Technologically, Folha can leverage its data to offer this – it would increase user engagement on their site as well. Moreover, the platform could allow filtering by region or type, as mentioned in point 1, so users can generate a list of, say, top universities in the Northeast or top private universities by research. These features align with user-centered design in rankings and address the diversity of user needs. We also recommend that RUF's published results include explanatory context such as the range of scores and perhaps grouping universities into performance bands (like top 10%, next 20%, etc.) to avoid overinterpretation of small rank differences. A footnote could clarify, for example, "Universities ranked 15–30 have score differences within a narrow band, suggesting similar performance; users should consider specific indicators to differentiate them." While media likes clear ranks, educating readers in this way demonstrates responsible presentation (Principle 15). This proposition was reinforced by our interview with a student representative (Interviewee #8) who said, *"I'd love to play with the ranking data myself. Right now, I have to trust the order they give. If I could tweak it to what matters to me, I would feel more confident in using it to choose a grad school."*

(6) Enrich the Teaching and Outcome Metrics (beyond existing ones): Alongside adding a new extension indicator, RUF could strengthen the Teaching and Market criteria with additional data if available. For instance, student retention or graduation rate could be a valuable outcome measure for teaching quality – universities that successfully graduate a high proportion of their students (especially on time) are arguably delivering better educational support. The Brazilian census data might have figures on dropout or completion rates that could be used. Another possible metric is post-graduate study progression – what percentage of alumni go on to postgraduate programs or further training (indicating solid academic preparation). On the Market side, if any data on actual employment rates or average salaries of graduates can be gathered (perhaps through alumni surveys or government labor statistics matched by institution), those would be superior to pure reputation. We recognize these data are not centrally tracked in Brazil yet, but as an interim measure, RUF could collaborate with e.g. LinkedIn or use online data to estimate graduate outcomes (some researchers, like Kogut (2023), have manually done this for a sample). Even incorporating a rough employability outcome metric (like % of grads employed in their field within 1 year) for a subset of programs could start a conversation and encourage development of better data collection nationally. Including such metrics would move RUF closer to measuring what stakeholders truly care about (are students learning and getting good opportunities?) rather than proxies alone. It would also diminish reliance on subjective surveys by balancing them with hard outcomes, satisfying Berlin's preference for outcomes and robust data.

(7) Continue and Expand Data on Internationalization but Consider Qualitative Aspects: RUF's internationalization indicators are currently research-focused (collaborations, international citations). These capture one aspect of globalization but miss others, like student/faculty mobility or international curriculum. If data permits, RUF could consider adding metrics such as percentage of international students or faculty, or participation in exchange programs. Admittedly, Brazilian universities have relatively low international student numbers, but tracking it could incentivize improvement in this area aligned with diversity (some institutions excel in regional integration even if not in research). Alternatively, RUF might include an indicator of foreign language offerings or global engagement in teaching. While these might be minor, they would round out the picture of internationalization beyond research collaborations, and reflect cultural diversity. If adding such data isn't feasible, maintaining the current two metrics is fine, but perhaps give them slightly more weight or visibility, since Brazil is increasingly interested in global engagement. Ensuring the ranking values international outlook in a balanced way can be part of recognizing different strategic priorities among universities (some might focus on regional mission, others on becoming internationally connected – both are valid paths).

Implementing new indicators, especially for extension or graduate employment outcomes, will require coordinated policy action. Universities and the Ministry of Education could establish partnerships with labor agencies and survey institutes to generate reliable datasets. Pilot programs funded jointly by government and institutional consortia may help overcome initial barriers. Anticipating resistance, consensus-building strategies such as phased implementation and advisory committees could foster greater acceptance among stakeholders.

Each of these proposed adjustments has implications. Implementing them would move RUF closer to an ideal of comprehensive evaluation. To gauge potential effects, we did a thought experiment with recent RUF data: for example,

if an extension metric existed, universities known for strong community programs (often mid-ranked public universities in less developed states) might see a boost. A university like Universidade Federal do Rio Grande do Norte, which has extensive healthcare outreach and literacy programs, could gain points in an extension category, perhaps moving it up a few positions. Similarly, grouping by category might highlight that some private universities (with less research) are actually top in teaching and employability in their peer group, something not obvious in the unified ranking. These changes aim not to shuffle ranks arbitrarily, but to reward a broader set of excellence dimensions and present information in a more user-relevant manner.

Crucially, none of the adjustments call for abandoning RUF's core strength of data-driven comparison. Rather, they build on it: adding data where it's lacking (extension, student outcomes), refining weighting logic, and better guiding interpretation. Folha, as the ranking publisher, would need to weigh the trade-offs. More indicators and complexity can be harder to communicate to the general public. However, as van Vught and Ziegele (2012) argue in the context of U-Multirank, providing a multidimensional view actually empowers users without necessarily confusing them, as long as the interface is well-designed. Folha could introduce changes gradually: e.g., pilot a new extension metric as an "experimental" category one year before fully integrating it, or roll out the interactive tool as a beta version.

From an adoption standpoint, engaging stakeholders early (as per recommendation #4) is key. If major universities and the Ministry of Education are consulted and buy into these changes, they can help champion the improved RUF. Given that RUF has become something of a de facto national benchmark, these improvements could also influence universities' behaviors in a positive way. For instance, if extension counts in the ranking, universities may invest more in extension activities or at least in documenting them – aligning institutional incentives with the national goal of community service. This outcome would exemplify how aligning a ranking with the diversity principle can have constructive real-world impact (Kehm & Stensaker, 2009, discuss that rankings can drive change – here the change could be broader fulfillment of universities' social roles, not just research output).

In summary, the proposed adjustments aim to: (a) make RUF fairer and more representative by accounting for different types of excellence (teaching, extension, regional impact) in addition to research, and (b) enhance the usefulness and interpretability of RUF for various stakeholders through better presentation and engagement.

Key Recommendations (for accessibility):

(1) Introduce mission-sensitive ranking categories;

(2) Add Extension/Community Engagement indicators;

(3) Rebalance weighting schemes to reduce research dominance;

(4) Establish advisory panels and external audits;

(5) Provide interactive/customizable ranking tools;

(6) Enhance teaching and employability outcome metrics.

5. Discussion

To exemplify potential effects of alternative indicators, we simulated a scenario where a new Extension/Community Engagement metric worth 10% was added, with Research weight reduced from 42% to 32%. The illustrative score shifts are shown below:

Table 3. Illustrative Quantitative Scenario

University	Current RUF Score	Simulated Score (with Extension 10%)	Change in Rank Position
USP	98.5	96.0	–1
UFRN	75.2	80.1	+3
PUC-RS	70.4	73.2	+2
UNIFESP	82.1	81.5	0

(Sources: Folha de S. Paulo, 2019; 2024)

While RUF demonstrates strong technical compliance with several Berlin Principles (for example, Folha publishes indicator definitions annually and provides transparent methodological notes), its substantive alignment remains weaker. Substantively, RUF's concept of quality continues to be research-centric, thereby excluding institutions that

emphasize teaching or community engagement. This distinction underscores how policy reforms or differentiated weighting schemes could shift incentives across university types.

The alignment analysis and proposed adjustments for RUF have broader implications for how we think about quality assurance and rankings in higher education. Firstly, our evaluation of RUF highlights that technical adherence is not the same as substantive adherence. RUF technically ticks many boxes of good practice (transparency, data validity, etc.), which likely contributes to its credibility among academics. However, substantive adherence – capturing the multifaceted nature of university quality – proved more challenging. This finding resonates with Barron's (2017) critique that rankings often adopt the legitimacy (by nodding to principles) without fully embracing the value shifts they imply. In the case of RUF, while it implemented a robust methodology, it still largely propagated a single ideal of a "good university" (research-intensive, publication-productive) akin to global rankings. This suggests that national rankings, even those designed with some awareness of international principles, reinforcing a homogenizing excellence unless continuous effort is made to integrate local values and diversity.

Secondly, the Brazilian context illustrates the tension between global norms and local relevance. RUF's creators drew on global ranking norms in part to ensure acceptance and prestige (Fausto et al., 2016). A ranking that didn't emphasize research or international publications might initially have been dismissed by the academic elite as not aligning with world-class benchmarks. By adopting those norms, RUF gained traction and comparability (one can loosely see how Brazilian universities rank similarly in RUF and in global rankings – USP is top in both, etc.). However, as our analysis showed, those same norms left certain local priorities (like extension) out of the frame. The adjustments we suggest – such as adding extension metrics or grouping by mission – represent a deliberate choice to prioritize local definitions of quality alongside global ones. There is an underlying policy debate here: Should national rankings mainly serve as stepping stones to global competition (by pushing all institutions to emulate research universities), or should they celebrate and enhance the diverse contributions of different institutions to national development? The Berlin Principles, especially Principle 3, clearly advocate for the latter approach. Our study thus contributes to this debate by concretely demonstrating how a national ranking can recalibrate in favor of diversity without abandoning excellence.

Thirdly, implementing changes like these in RUF could influence university behavior in more positive directions. Rankings inevitably create incentives. Under the current RUF, the incentive is strong for all universities to invest in research output (and indeed, some Brazilian universities have explicitly cited RUF performance as justification for expanding research and postgraduate programs). This has upsides (more research activity) but also potential downsides if it leads institutions to overstretch into missions misaligned with their core strengths or neglect teaching and community service (Santos, 2015, warned of this "mission drift" risk). If RUF expanded its criteria to include extension and gave more weight to teaching outcomes, universities might allocate more attention and resources there, knowing it will also reflect in the ranking that the public and officials see. For example, a private university that currently puts minimal effort into community programs might start one to improve both real impact and its RUF profile. In essence, aligning RUF with Berlin Principles can help align institutional behavior with a more well-rounded set of goals – a synergy between ranking and quality enhancement. This aligns with the notion in Kehm & Stensaker (2009) that rankings, despite their issues, can be leveraged to promote diversity and improvement if designed carefully.

It is instructive to compare this with other countries' experiences. Germany, for instance, historically resisted simplistic rankings and encouraged multidimensional assessments (like CHE rankings that allow personalized weighting). Brazil's situation in 2012 was that no such ranking existed, and RUF filled a void. Now over a decade later, Brazil has the opportunity – as Germany did – to steer its ranking culture toward more sophisticated, purpose-driven models. The adjustments we propose for RUF mirror some features of U-Multirank (user customization) and U.S. News (categorization) and introduce novel ones (extension metric) that could be pioneering. It suggests that national rankings can innovate beyond what global rankings do, precisely because they operate in a specific context where stakeholders can agree on broader goals of higher education (like community engagement for public universities, which globally might not be a consensus priority but nationally can be).

Of course, our propositions are not without challenges. For one, adding new indicators or changing weights could face resistance from those who currently benefit from the status quo. Top research universities, which dominate RUF, might quietly prefer the ranking remain heavily research-focused. If a new extension indicator were added and given, say, 5% weight, some top universities might lose a few points if they have neglected extension relative to others. Managing this political aspect requires careful stakeholder management – hence the emphasis on involving universities in the process so they feel a sense of co-ownership of any changes. Another challenge is data: measuring extension or graduate outcomes reliably will require effort and perhaps new data infrastructure (like a national graduate tracking system – an idea that has been floated in Brazilian policy circles). RUF alone cannot create that, but RUF can catalyze demand for

it. In interviews, a Ministry official (Interviewee #6) acknowledged that *"if rankings start asking for these data, it puts pressure on us to collect them."* This indicates a potential positive feedback loop: RUF's evolution could drive the national quality assurance system to also broaden metrics (indeed, SINAES could incorporate more quantified extension indicators in their institutional evaluations, learning from RUF or vice versa).

The role of media and public perception: Since RUF is published by a media outlet, any changes also need to be communicated well to the public. There's a communication strategy aspect: Folha would need to explain why, for instance, University X went from rank 50 to 40 – was it because they improved or because methodology changed? Transparency and transitional reporting are essential to maintain credibility during changes. The Berlin Principles allow changes but emphasize transparency about them. It might be advisable for Folha to pre-announce significant methodology changes and perhaps produce a one-time "re-ranked" list from previous year's data under the new method to show differences. This is common in ranking revisions (for example, when U.S. News modifies its formula, it often explains how that affected results). Doing so would reinforce RUF's adherence to Principle 9 (limiting and clarifying changes) even as it makes necessary adjustments.

For the academic community, this case study of RUF is a valuable example of how global guidelines like the Berlin Principles can be concretely applied to critique and improve a ranking system. It shows that these principles are not just abstract ideals but can serve as a practical checklist for evaluating a ranking's quality and fairness. Other countries with emerging rankings (e.g., India's NIRF ranking or China's domestic rankings) might conduct similar exercises. Notably, our approach combined principle-based analysis with stakeholder perspectives, an approach that could be generalized: involve those affected in judging a ranking's success in meeting quality standards.

For policymakers, particularly in Brazil, the findings underscore that an overemphasis on rankings purely as competitive tools (who's #1) may overshadow important policy goals of equity and diversity. The National Education Plan might aim to strengthen certain regional universities or increase community outreach; if the main public yardstick (RUF) doesn't value those, the policy goals might be undermined by prestige chasing. Therefore, aligning RUF with policy values (like inclusion and regional development) can harmonize incentives. Perhaps the Ministry could even collaborate with Folha or support data improvements that feed into RUF, recognizing it as part of the accountability ecosystem. There is precedent: some countries' governments publish their own rankings or indicators (Russia's Ministry does an annual university monitoring, for instance). In Brazil's case, a respected independent ranking like RUF that internalizes public policy values could be more effective than any official ranking in nudging institutions.

Our study has some limitations that warrant discussion. It is largely qualitative and based on expert opinion and document review; thus, the assessment of compliance with principles has an element of subjectivity. Different evaluators might weigh certain principles differently. However, we attempted to mitigate bias by using clear criteria and cross-validation with interview insights. Another limitation is that we have not quantitatively demonstrated the impact of suggested changes on ranking outcomes – that could be a follow-up study (simulating new rankings under different weightings to see how much rankings would shift). We focused on the conceptual and qualitative justification for changes. Additionally, while we spoke with a range of experts, the sample is not exhaustive – perhaps student voices and employers' voices were less represented compared to academic and administrative voices. Future research might survey a broader set of end-users (like a large number of students) to see what they want from a ranking – aligning that with principles could yield more user-centric improvements.

It would be beneficial to conduct this kind of Berlin Principles alignment analysis after implementing changes, to see if perceived compliance improves. Also, tracking how universities respond to any changes in RUF would be an interesting study in itself (e.g., do we see increased mention of extension in university strategic plans if extension gets measured?). On an international note, as rankings continue to evolve (with new forms like sustainability rankings emerging), the Berlin Principles themselves might be updated or expanded (IREG might consider adding principles about social responsibility or student-centered metrics, for example). National cases like RUF can inform those global discussions by showing innovation – for instance, if RUF successfully integrates an extension metric and it's well-received, that could inspire others.

In conclusion, the discussion reiterates that rankings are not static; they are social technologies that can be recalibrated to better serve the needs of higher education systems (Marginson, 2014). By aligning more closely with the Berlin Principles and the realities of Brazilian higher education, RUF can transcend being just a ranking – it can become a more comprehensive information tool and a driver for positive change that respects the diversity of institutional contributions to society.

6. Conclusion and Propositions

This study set out to examine how the Folha University Ranking (RUF) aligns with the Berlin Principles of university rankings and what adjustments could enable it to better capture the diversity of Brazil's higher education context. Our analysis revealed that RUF, in its current form, substantially adheres to many technical aspects of good ranking practice (such as transparency, use of robust data, and methodological consistency), underscoring the intentional design and credibility of RUF as a national ranking. At the same time, we identified a critical misalignment: RUF's one-size-fits-all approach and heavy emphasis on research outputs do not fully align with Principle 3's call to recognize institutional diversity and different missions. In effect, RUF mirrors the biases of global rankings by elevating research-intensive universities as the paragons of excellence, while undervaluing universities that excel in teaching quality, regional engagement, or community service. This gap suggests that RUF, like many rankings, has room to evolve from a purely competitive ranking to a more developmental ranking tool that supports a broader notion of quality in higher education.

In response to our findings, we propose the following key propositions (recommendations) for transforming RUF to better meet the Berlin Principles and serve the Brazilian context:

(1) Implement Mission-Specific Groupings or Sub-Rankings. RUF should introduce classifications within the ranking to compare like with like. For instance, publish separate highlight lists for different institutional categories (e.g., top public research universities, top private teaching-focused universities, top regional universities, etc.). This stratification acknowledges that excellence is multi-form and provides more meaningful information to users. By grouping peers, RUF would align with the diversity principle, allowing institutions to be recognized for excelling among those with similar missions. This change can be communicated as making the ranking more fair and useful – a move from a simplistic "horse race" to a more nuanced league that celebrates multiple champions in different arenas.

(2) Enrich the Indicator Portfolio to Include Community Engagement (Extension) and Other Contextual Outcomes. To truly reflect the Brazilian higher education mission, RUF should develop and incorporate an extension/community engagement indicator into its methodology. This could be based on quantitative data (number of extension programs, beneficiaries, etc.) or a composite score from qualitative evaluations of social impact. In tandem, RUF can strengthen existing criteria by including additional outcome measures such as student retention/graduation rates or graduate employment rates if data can be obtained. These new indicators would broaden the definition of quality beyond academic output to educational and social outcomes, thus realigning the ranking with national values and Berlin Principles. We recommend pilot testing these measures and consulting with universities on feasible data collection. In the long run, this expansion will incentivize and reward institutions for fulfilling all aspects of their mission – teaching, research, and service – painting a fuller picture of institutional performance.

(3) Rebalance Weights to Reduce Overemphasis on Research. RUF should consider adjusting the weight distribution of its major indicator categories to ensure no single dimension (research) dominates the overall score to the extent it currently does. A more balanced weighting (for example, bringing Teaching on par with Research, and allocating modest weight to the new Extension metric) would reflect the principle that quality is multi-dimensional. This rebalance can be calibrated using statistical analyses (e.g., ensuring that variability in overall scores comes from multiple dimensions, not overwhelmingly from one). Such a change should be done transparently: Folha can announce the rationale (e.g., "to better value teaching and community impact, we have increased their weight in the index"). The outcome will be a ranking that is less of a proxy for research volume and more of a composite of different strengths. We anticipate this could slightly shuffle rankings, but more importantly, it sends a message that improvement in any core function of a university is recognized – thereby encouraging a more holistic pursuit of excellence across institutions.

(4) Proposition 4: Introduce User-Centric Interactive Ranking Tools and Improved Result Communication. To adhere to best practices in result presentation, RUF should develop an interactive online dashboard where users can customize the ranking based on their own priorities. By allowing prospective students, for example, to emphasize teaching or employability in the weights, RUF empowers users and educates them that "best" can vary by needs. Alongside this, Folha should enhance contextual information in its reporting: for instance, grouping universities in performance bands (e.g., gold, silver, bronze tiers or quartiles) and clearly stating when differences are marginal. An explicit note in RUF publications could read, *"Universities X, Y, Z fall in the top tier for research, whereas A, B, C lead in teaching outcomes,"* etc. This approach demystifies the ranking and reduces the likelihood of overinterpretation of minor rank differences. Ultimately, this proposition makes RUF not just a ranking, but a *tool for informed decision-making*, aligning with Berlin Principle 15 by providing clarity and choice to the consumer of ranking information.

(5) Proposition 5: Establish a Regular Review Mechanism with Stakeholder Involvement. We recommend that Folha sets up a formal RUF Advisory Council including representatives from universities (of various types and regions), educational assessment experts, and perhaps student and employer voices. This council would review the ranking methodology on a set schedule (annually or biennially) and advise on improvements, new data sources, and any concerns (like survey processes). Their input should be made public to ensure transparency of the review. Additionally, Folha could pursue an independent audit (e.g., through IREG) to benchmark RUF against international standards periodically. These steps would institutionalize the quality assurance of RUF itself (addressing Berlin Principle 13). This ongoing dialogue would help RUF remain relevant as the higher education landscape changes (for example, if online education becomes more prominent, the council might suggest new metrics for that). It would also increase legitimacy and buy-in from the academic community, turning what can sometimes be a contentious external evaluation into a collaborative effort to define and measure quality in Brazilian higher education.

(6) Proposition 6: Collaborate with National Agencies to Improve Data for Rankings and Quality Improvement. As a final recommendation, RUF's evolution should be tied to broader data improvements in Brazil. Folha could partner with the Ministry of Education (INEP) and funding agencies to gain access to richer data – for example, a graduate tracking system, if developed, or datasets on innovation and entrepreneurship activities of universities. In return, Folha can share insights from RUF analysis that might help inform public policy (like identifying universities that dramatically improved research output or teaching effectiveness, and investigating how). This public-private collaboration would ensure that RUF's adjustments (such as new extension metrics) are supported by robust data collection efforts, and that the ranking complements official quality assurance. It aligns with the Berlin Principles' ethos of continuous improvement and relevance. Moreover, by embedding the ranking in a network of educational data initiatives, RUF can act as a bridge between accountability (ranking for transparency) and improvement (diagnosing areas to work on), which ultimately benefits the entire system.

In implementing these propositions, it is vital that changes are managed in a transparent, phased manner. We suggest a roadmap where, for example, Proposition 4 (interactive tools) and Proposition 5 (advisory council) are relatively low-hanging fruits that could be done within a year, showing Folha's commitment to responsiveness. Propositions 2 and 3 (new indicators and weight changes) might be piloted and discussed in the advisory council, then introduced perhaps in the second year after consensus-building. Throughout, Folha should communicate with clarity: emphasizing that the aim is to create a fairer, richer ranking that does justice to all the contributions universities make to society – not to favor one group over another arbitrarily. Early communication of upcoming changes, along with release of test analyses (e.g., "Here's how the 2023 ranking would look if extension were included – just for illustration") can acclimate stakeholders and reduce surprise or pushback.

By adopting these adjustments, RUF has the opportunity to become a more comprehensive barometer of quality in Brazilian higher education, one that aligns more closely with both international good practices and the values and diversity of Brazil's universities. The enhanced RUF could serve multiple purposes: guiding students in multifaceted ways, informing policy by highlighting different dimensions of performance, and encouraging institutions to strive for improvement across all mission areas, not just those traditionally rewarded. In doing so, RUF would exemplify how a ranking can mature from a blunt benchmarking instrument into a nuanced analytical tool embedded in the higher education quality ecosystem.

In conclusion, the question driving this study – the alignment of RUF with Berlin Principles and the adjustments for diversity – proved to be a fruitful lens. It allowed us to systematically appraise a prominent ranking and chart a course for its evolution. The Berlin Principles provided a valuable framework to ensure that any ranking, including RUF, remains focused on fair, ethical, and meaningful assessment. The Brazilian context provided the substance, reminding us that rankings must serve the interests of the society and system in which they operate. Marrying the two, our analysis and propositions illustrate that yes, RUF can be realigned and improved to capture Brazil's rich diversity – and doing so will likely increase its legitimacy and utility. As higher education faces new challenges and the demand for accountability grows, tools like RUF should not be static report cards but dynamic, principle-guided instruments for insight and enhancement. By following the recommendations above, RUF could very well become a model for how national rankings can uphold global standards of good practice while celebrating and fostering the unique tapestry of excellence that defines a country's higher education landscape.

References

Almeida-Filho, N. (2011). *Rankings, vikings, masters & colleges: dilemas da universidade brasileira no contexto da internacionalização*. [Lecture]. Instituto de Estudos Avançados, Universidade de São Paulo, April 2011.

Barron, G. R. S. (2017). The Berlin Principles on ranking higher education institutions: Limitations, legitimacy, and value conflict. *Higher Education, 73*(2), 317-333. https://doi.org/10.1007/s10734-016-0022-z

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M. P., & Vadillo-Muñoz, O. (2007). Comparative study of international academic rankings of universities. *Scientometrics, 71*(3), 349-365. https://doi.org/10.1007/s11192-007-1653-8

Chen, Y., & Liu, N. C. (2008). Examining major rankings according to the Berlin Principles. *Higher Education in Europe, 33*(1), 201-208. https://doi.org/10.1080/03797720802253686

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education, 49*(4), 495-533. https://doi.org/10.1007/s10734-004-1746-8

Fausto, S., Calero-Medina, C., & Noyons, E. (2016). O "Ranking Brasileiro de Pesquisa" e os Princípios de Berlim para Rankings de Instituições de Ensino Superior. *Future Studies Research Journal: Trends and Strategies, 8*(2), 211-236. https://doi.org/10.24023/FutureJournal/2175-5825/2016.v8i2.239

Folha de S. Paulo. (2014). *Como é feito o Ranking Universitário Folha*. São Paulo: Folha de S. Paulo. (Retrieved from Folha RUF 2014 website on Oct 10, 2015).

Folha de S. Paulo. (2019). *Como é feito o Ranking Universitário Folha 2019*. São Paulo: Folha de S. Paulo.

Folha de S. Paulo. (2024, October 20). *Como foi feito o RUF 2024*. São Paulo: Folha de S. Paulo.

Hazelkorn, E. (2015). *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence* (2nd ed.). Basingstoke, UK: Palgrave Macmillan. https://doi.org/10.1057/9781137446671

Hou, Y., Morse, R., & Wang, M. (2021). Defining standards for rankings: An investigation of global university rankings according to the Berlin Principles. *European Journal of Education, 56*(2), 278-293. https://doi.org/10.1111/ejed.12487

International Ranking Expert Group (IREG). (2006). *Berlin Principles on Ranking of Higher Education Institutions*. Berlin: IREG. (Retrieved from IHEP website).

IREG Observatory. (2019). *IREG Ranking Audit: Guidelines*. IREG Observatory on Academic Ranking and Excellence.

Kehm, B. M., & Stensaker, B. (Eds.). (2009). *University Rankings, Diversity, and the New Landscape of Higher Education*. Rotterdam: Sense Publishers.

Kogut, C. N. F. (2023). *O Ranking Universitário Folha e o indicador mercado de trabalho: uma análise do Curso de Direito da UFSC* (Master's dissertation, Universidade Federal de Santa Catarina).

Marginson, S. (2014). University rankings and social science. *European Journal of Education, 49*(1), 45-59. https://doi.org/10.1111/ejed.12061

Mettzer Blog (Nalbert Rosa). (2020, updated 2023). *RUF: a classificação no ranking universitário da Folha*. Retrieved from https://blog.mettzer.com/ruf-ranking-universitario-folha/

Mollis, M. (2016). *Los rankings universitarios internacionales: impacto y desafíos en América Latina*. Buenos Aires: UNESCO-IIEP.

Rauhvargers, A. (2013). *Global University Rankings and Their Impact – Report II*. Brussels: European University Association.

Santos, S. M. dos. (2015). *O desempenho das universidades brasileiras nos rankings internacionais: áreas de destaque da produção científica brasileira* (Doctoral thesis, Universidade de São Paulo).

Shin, J. C., & Toutkoushian, R. K. (2011). The past, present, and future of university rankings. In J. C. Shin, R. K. Toutkoushian, & U. Teichler (Eds.), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education* (pp. 1–16). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-1116-7_1

Soares, A. de Á. (2022). *Rankings universitários em perspectiva: o Ranking Universitário Folha e a construção da excelência* (Master's dissertation, Universidade Federal de Minas Gerais).

Sobrinho, J. D. (2010). *SINAES: Sistema Nacional de Avaliação da Educação Superior – concepções e desafios*. Brasília: INEP.

UFES (Universidade Federal do Espírito Santo). (2018). *Publicada nova edição do Ranking Universidades Folha – release on UFES performance and methodology* (accessed from prppg.ufes.br).

UFF (Universidade Federal Fluminense). (2024). *Rankings Universitários: o que são, semelhanças, diferenças e limitações* (Presentation slides by NIDI/UFF).

Usher, A., & Medow, J. (2009). A global survey of university rankings and league tables. In B. M. Kehm & B. Stensaker (Eds.), *University Rankings, Diversity, and the New Landscape of Higher Education* (pp. 3–18). Rotterdam: Sense. https://doi.org/10.1163/9789087908164_002

Van Vught, F., & Ziegele, F. (Eds.). (2012). *Multidimensional Ranking: The Design and Development of U-Multirank*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-3005-2

Yin, R. K. (2014). *Case Study Research: Design and Methods* (5th ed.). Thousand Oaks, CA: Sage.