

MULTIDIMENSIONALITY THROUGH SELF-EVALUATION: FROM THEORY TO PRACTICE IN THE BRAZILIAN GRADUATE SYSTEM

ANDRÉ BRASIL

DOI: 10.22163/fteval.2022.546

ABSTRACT

Nearly all science and technology research in Brazil is conducted within a national system of graduate education. Since the 1970s, a graduate program assessment has been an integral part of such a system, and it is currently held on a quadrennial basis. The evaluation model is dynamic, evolving from the experiences of evaluators, policymakers, and the scientific community during each four-year cycle. This study analyses policy initiatives from the 2017-2021 evolving effort, focusing on strategies and recommendations to implement multidimensionality and self-evaluation as integral components of Brazilian evaluation. The paper traces how the idea for a multidimensional assessment was introduced in the country and how U-Multirank, an international ranking of higher education institutions (HEI), has come to inspire an evaluation that is not institutional but of graduate programs instead. The study identified some benefits and limitations of the chosen inspiration and analysed how the Brazilian proposal aligned with the U-Multirank principles. Furthermore, the investigation shows there is little concrete difference from the proposed new model to the one Brazil has already in place. Finally, the last section of this study looks into the once pivotal idea to pursue a self-evaluation component, now relegated to a minor role in the model, but that could be raised to a position supporting the design of an actual multidimensional assessment model.

1 INTRODUCTION

The Brazilian science system is relatively young. By 1965, the country was yet to develop a research tradition, and the little science conducted was essentially confined to a few research institutes and a graduate system of only 27 master's and 11 doctoral courses. Considering that Brazil reached a population of 90 million before the end of that decade, the numbers were far from optimal. However, the scenario started to change over the following years as a robust National System of Graduate Education (SNPG) was launched by a series of government initiatives. This system was conceived based on the core idea that science and education should be strongly connected. As a consequence, most of the country's science and technology research is conducted within graduate programs, both at the master's and doctoral levels (Balbachevsky, 2005; Brasil, 2020; CFE, 1965; Martins, 2018).

Since its conception, evaluation has been an integral part of the SNPG. For nearly 50 years, the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) – a public foundation linked to the Ministry of Education – has been in charge of evaluating such a system. The adopted model has evolved over the years, and today it includes both a mandatory accreditation and a quadrennial evaluation of graduate programs (PPG). A grading system on a 1 to 7 scale applies, with grades one and two insufficient for accreditation renewal. Superior grades guarantee not only program continuity but may lead to additional funding, access to a broader set of grants, added institutional prestige, and more (Brasil, 2020; Ferreira and Moreira, 2002).

In 2017, the national evaluation assessed the 2013–2016 performance of 4175 graduate programs, with a total of 6303 doctoral and master's courses. The effort involved nearly 2000 panel members, organised in 49 disciplinary committees, in a large-scale endeavour to combine qualitative and quantitative methods to assess the whole SNPG (CAPES, 2018a).

As the evaluation results were announced, Faljoni-Alario et al. (2018) formulated a report with critical suggestions to improve the evaluation model for the 2017–2020 cycle and beyond. The document resulted from discussions between CAPES – as the agency in charge of the evaluation – and coordinators from the disciplinary committees. The report recognised the accomplishments of the evaluation system and included recommendations regarding: time between evaluations, methods for assessing research outputs (e.g., journal articles, books, technical production), criteria and indicators adopted across disciplines, and more. A series of initiatives followed, including the establishment of thematic working groups to propose changes to various aspects of the evaluation system (CAPES, 2019; Monteiro et al., 2019).

Aiming to promote a collaborative redesign of its evaluation model, CAPES organised international seminars to further working group proposals, thus engaging Brazilian academia and the broader scientific community in the debate. As a result, two core concerns became central for the intended evolution of the evaluation system. The first was the need to design a self-evaluation strategy for the SNPG, as its significant expansion had become an obstacle for a central assessment to capture the complex narratives from thousands of graduate programs (Monteiro et al., 2019; Verhine et al., 2019). The second came from the perception that evaluation promoted an overly homogeneous science system since its one-dimensional approach led graduate programs to become substandard photocopies to the top-performing ones. A multidimensional assessment, capable of recognising and valuing differences, was now required (PNPG Committee, 2018; FOPROP, 2018).

This paper investigates ongoing institutional efforts to implement multidimensionality and self-evaluation as components of the Brazilian national evaluation model. Through the analysis of policy documents, legislation, reports, and assessment guidelines, the study traces the motivation and the path towards a multidimensional evaluation, including an overview of the leading proposal for its implementation. The paper also explores the use of U-Multirank – an international ranking of higher education institutions (HEI) – as a source of inspiration for the proposed model, highlighting both the benefits and drawbacks of such adoption. Finally, the study explores the self-evaluation component and identifies the current recommendations for its adoption underestimate its potential to enable a genuinely multidimensional model.

2 TOWARDS A MULTIDIMENSIONAL EVALUATION

The report produced by Faljoni-Alario et al. (2018) was a significant yet initial analysis of what was needed to evolve the evaluation of graduate programs conducted at the time. Considering the need for a broader perspective, CAPES tasked the special committee in charge of monitoring the National Plan for Graduate Education (PNPG)¹ to supplement the material. For that, the group reached out to over a dozen influential organisations in the country's science system, including the Brazilian Society for the Advancement of Science (SBPC), the Brazilian Academy of Sciences (ABC), and the National Council for Scientific and Technological Development (CNPq) (PNPG Committee, 2018, p. 3).

Many of the submitted contributions conveyed concerns about the role of the current evaluation model in shaping a science system that was too academic, focused on the training of future professors for the country's higher education system. A document prepared by the National Forum of Pro-Rectors for Research and Graduate Education (FOPROP) – one of the leading interlocutors between HEI, science policymakers and funding agencies in Brazil – clearly expressed the collective expectations, stating that graduate programs should not be required to excel in every dimension; they could be excellent according to their vocation or specific mission. Evaluation should be able to recognise value across multiple dimensions (FOPROP, 2018, p. 2).

From the joint effort, the PNPG Committee (2018) prepared a report delineating an evaluation model in five dimensions: Training of human resources; Internationalisation; Scientific production; Innovation and knowledge transfer; Economic and social impact. CAPES' Higher Council unanimously approved the proposal, making multidimensionality a priority to evolve the evaluation model (Audy, 2020).

While Audy (2020) mentions the initial multidimensional proposal was not based on any existing system, the aforementioned FOPROP (2018) document suggested the Times Higher Education Ranking as a

possible inspiration. Even though the specific suggestion did not seem to find space in the following discussions, it might have directed attention towards other rankings, leading to the discovery of U-Multirank (UMR) as a potential reference to build the new evaluation model.²

2.1 U-MULTIRANK: A PROVISIONAL INSPIRATION

The predominant view from many scholars such as van Raan (2005), Calero-Medina et al. (2008), and Gadd et al. (2021) seems to be that rankings are an undeniable part of the higher education landscape with recognised applications despite their evident flaws, biases, and shortcomings. While the objective of this study is not to analyse the value of such rankings, previous research provides relevant arguments to frame their potential as an inspiration to reform a complex national evaluation system such as the Brazilian one.

Starting from the work of Hazelkorn and Gibson (2017), we understand that global rankings often do not count with meaningful, reliable and verifiable indicators and data, especially for international comparisons. As a result, they usually give preferential weight to research outputs, favouring higher education institutions with a focus on the physical, life and medical sciences, and favouring countries where English is the native language. According to Waaijer (2018), that problem is made worse by the fact that most university rankings yield composite scores, often the result of nontransparent raw data, transformation of scores, and weighting. That makes it difficult to analyse the meaningful differences that exist between universities.

In line with the presented perspective, Gadd et al. (2021, p. 16) call for "open and transparent assessment of the relative strengths and weaknesses of the global university rankings to make them more accountable to the higher education communities being assessed". Aligned with that, Moed (2017) mentions current rankings are mostly one-dimensional, and changing that is not a simple task, as even the local, national or international orientation of universities is a dimension often challenging to consider (Calero-Medina et al., 2008).

Considering the presented flaws of global rankings and more, van Vught et al. (2012) introduced U-Multirank as a new approach to ranking in higher education and research. Following a feasibility study conducted by a consortium of universities and research organizations (known as CHERPA), UMR published its first set of results in 2014, aiming to be a "multidimensional", "multilevel", "participative", and "user-driven" ranking.

Fanelli (2016), Hazelkorn and Gibson (2017), and Moed (2017) are among those that recognise UMR to be unlike most international rankings, primarily because of the multidimensional perspective that comes from addressing more than research, as four other essential dimensions of higher education are included: teaching and learning, knowledge transfer, internationalisation, and regional engagement.

UMR's multilevel perspective is about providing information of value for distinct groups of stakeholders. While for some, reports about a particular field may be desirable (e.g., potential students), for others, the

1 See Brasil (2020) for further discussion on the National Plans for Graduate Education.

2 After a preliminary investigation about U-Multirank, a Brazilian delegation visited lead partners of the consortium engaged in its development, in Germany and the Netherlands. Demonstrating the country's commitment to a multidimensional evaluation and U-Multirank influence, the mission included influential representatives of the SNPG, such as the national evaluation director (CAPES), the president emeritus of SBPC, and the president of the PNPG Committee. A month later, UMR representatives also visited CAPES to discuss a potential collaboration (F. Marques, 2019).


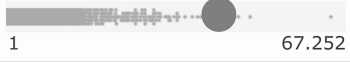
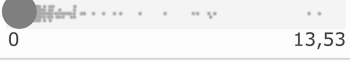




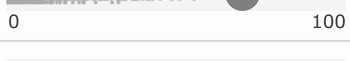
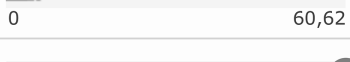

institutional-level ranking results might be the most relevant (e.g., HEI managers). Thus, UMR organises its data and indicators aiming to allow for comparisons at the organisation level, but also at the level of disciplinary or multidisciplinary fields (Federkeil, Kaiser, et al., 2012).

For the participative aspect of U-Multirank, the current methodology is based not only on national datasets and collection of organisational or bibliographic data but also on institutional and student surveys conducted for every new edition (U-Multirank, 2021). Finally, for the user-driven perspective, van Vught et al. (2012, p. 3) state that usual university rankings have the pretension of being guided by a nonexistent theory of the quality of higher education, and thus they present collec-

tions of indicators as a reflection of a definitive quality of the institution. U-Multirank was designed to be interactive so that users could have control over the available indicators. Information is made transparent so that personal rankings can be tailored to suit specific purposes and users' needs.

Figure 1 shows how this transparency materialises in the ranking results presented in the 2021 edition of UMR. The example retrieved from U-Multirank (2021) shows the "research" dimension and the accompanying indicators³ for the University of São Paulo (USP), one of the largest HEI in Latin America.

Figure 1: U-Multirank research dimension for the University of São Paulo (USP)

Research			
	Score	Grade	Other universities
Citation rate	0,77	C	 0,08 2,53
Research publications (absolute numbers)	41.982	A	 1 67.252
Research publications (size-normalised)	0,519	A	 0 13,53
External research income	46,99	B	 0 282,88
Art related output	—		 0 1,99
Top cited publications	6,5%	D	 0 34,2
Interdisciplinary publications	9,6	B	 0 24,6
Post-doc positions	69,29%	A	 0 100
Professional publications	—		 0 60,62
Open Access Publications	28%	A	 0 100

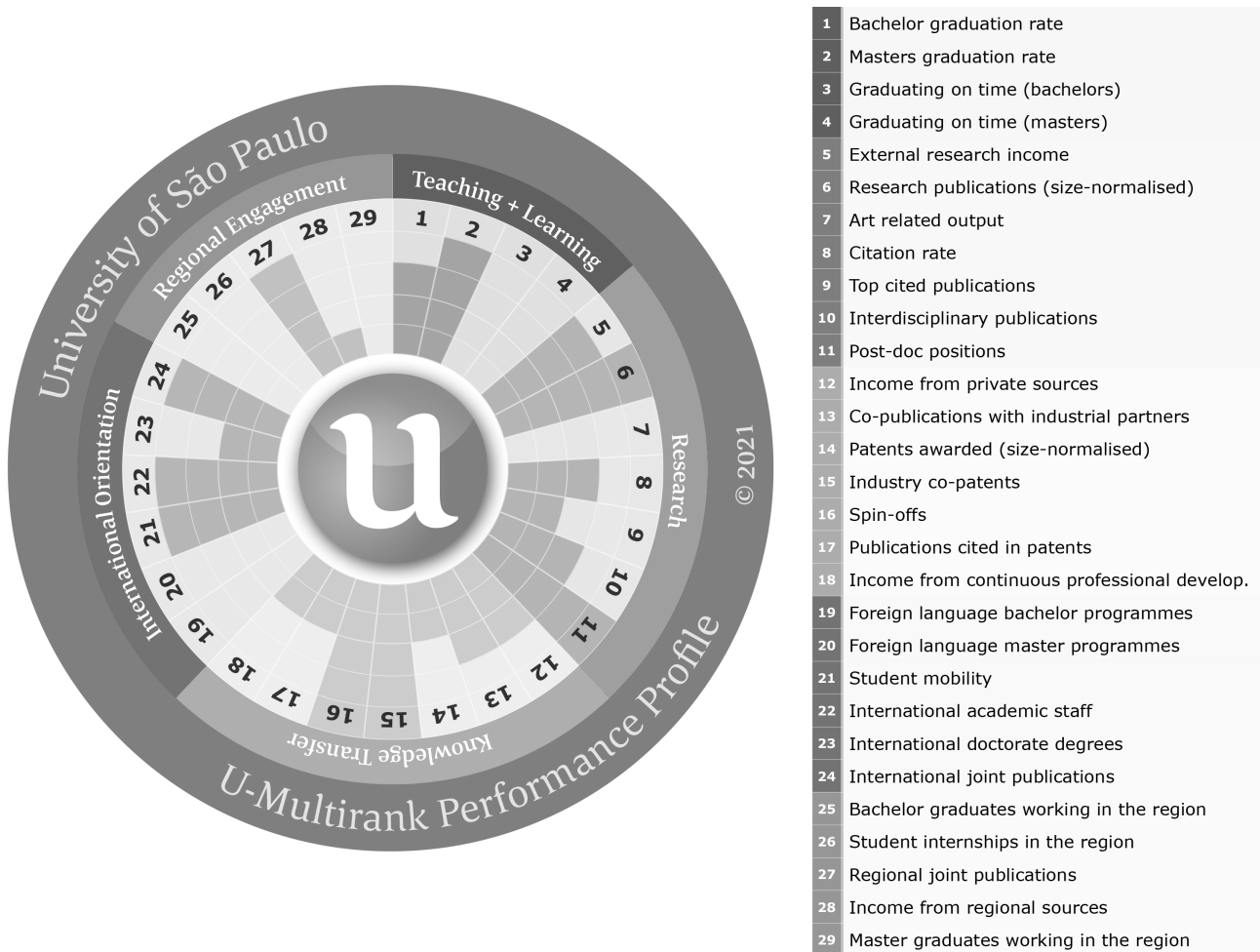
3

In the U-Multirank website, the term "measure" is used to refer to indicators, which will be used in this paper instead, as we consider it to be more appropriate to reflect the type of information included in UMR.

The “research” dimension displayed on Figure 1 includes ten indicators. Calculated scores for the institution are presented against those for all other universities in the database. Data points for USP are shown resized according to the obtained grade in each measure, attributed on a five-level scale: A (very good), B (good), C (average), D (below average), and E (weak). Missing grades are indicated with “–”, usually a result of unavailable data from the institution.

The current version of U-Multirank includes 35 indicators distributed across the five dimensions mentioned. A sunburst chart presents 29 of those indicators in a visual profile for each covered institution. An example is shown in Figure 2, once again with data from USP, according to the ranking’s 2021 edition.

Figure 2: U-Multirank university performance chart for the University of São Paulo (USP), also listing the 29 indicators displayed in the sunburst



The sunburst seen in Figure 2 can be considered an evolution of the visualisation approach previously adopted by the U-Map project on the European Classification of Higher Education Institutions. Federkeil, Kaiser, et al. (2012) recognise that project as an essential inspiration to U-Multirank, not only because of how results could be presented, but also because it proposed comparing institutions in the face of their missions, profiles, and characteristics.

From a comparison perspective, the sunburst approach becomes an effective tool to visually analyse the profiles of different institutions and see the strengths and weaknesses in the areas of interest of each end-user. This powerful visualisation enables UMR’s decision not to produce

oversimplified league tables of the world’s top universities. Dropping the standard tables also makes it more feasible to go beyond the comparison of internationally oriented research universities to cover profiles such as: regionally oriented colleges, music academies, teacher training colleges, and universities of applied sciences. (Federkeil, File, et al., 2012; van Vught et al., 2012)

In the example of the University of São Paulo, eight indicators are empty, indicating the absence of data for the institution (e.g., graduating on time). The remaining bars are filled in five levels, from “E” (1) to “A” (5). Comparing the list on Figure 2 with Figure 1, we notice some research indicators are not displayed in the chart, such as “professional

publications". Those omissions result from the expansion in the number of indicators since UMR's conception, which were not incorporated into the sunburst after its original design. A consequence seen in the case of USP is that some high-performance results, such as "Open Access publications", are not visible in the chart.

According to Moed (2017), ranking developers have made enormous progress over the past decade, in some cases offering informative, user-friendly systems with series of indicators that allow institutions to be ranked accounting for the diversity of their profiles. That seems to be the case of U-Multirank, which is recognised by some scholars as one of those that better meet the community's expectations of fairness and responsibility, despite existing reservations regarding overall ranking shortcomings (Fanelli, 2016; Gadd et al., 2021; Hazelkorn and Gibson, 2017).

2.2 U-MULTIRANK IN BRAZIL

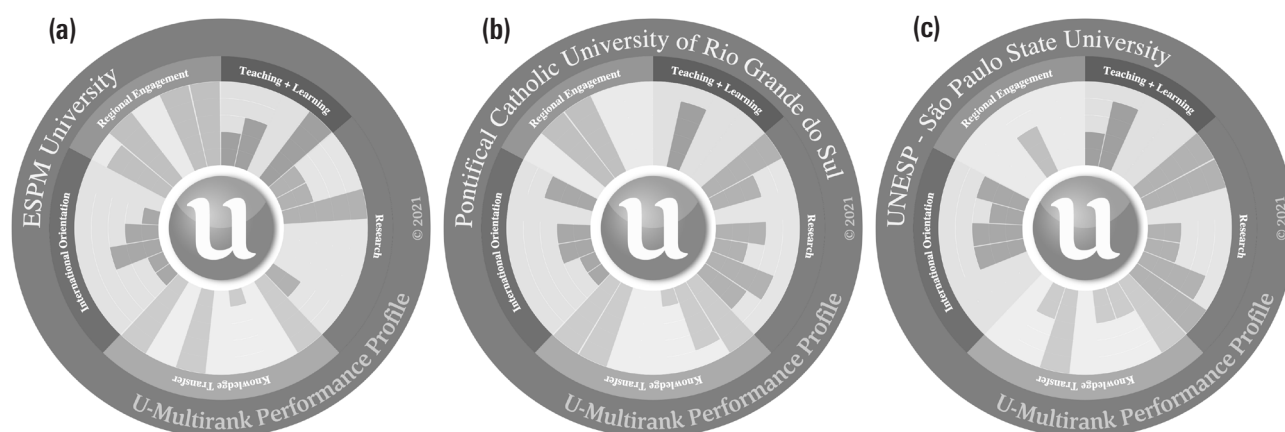
The coverage of Latin American (LA) institutions in U-Multirank is too small to have value within the continent. Of the 1948 HEI currently covered by UMR, only 52 are in LA, 34 in Brazil. The number is far from representative as the Brazilian higher education census reports 2537 HEI active in the country, most of them focused on offering undergraduate degrees (INEP, 2020; U-Multirank, 2021). As detailed in Brasil (2020), a total of 432 of those institutions also offer graduate programs, which may include master's or doctoral courses.

The low representation of Latin America in UMR partly results from the lack of institutional initiatives to register and provide the necessary data for the ranking. The challenge to overcome that problem, however, is made clear by Fanelli (2016, p. 8), who mentions "the quantity and quality of statistics on LA higher education systems vary per country and even per category of institution". The scholar also highlights that only a few Latin American HEI have adequate information about nonresearch indicators available, something evident from the blank indicators in the performance chart of the University of São Paulo (USP), presented in Figure 2.

The lack of complete information from higher education institutions has potentially harmful effects on comparisons, as users may find it challenging to produce their tailor-made lists. For instance, when using U-Multirank's interactive web tool, it is possible to filter HEI based on the subject area, country, and sets of variables associated with the five different dimensions. The resulting list can be sorted alphabetically, based on any particular measure or using "top scores". This system is based on the Olympic medal approach, where the list is ordered according to the number of gold medals won (which would be the "A" scores in UMR), and then by the subsequent levels (U-Multirank, 2021).

Using the described flow to evaluate Brazilian HEI involved in graduate education, in this study, universities were compared as a whole instead of by disciplinary field, and then they were filtered to include those offering master's or doctoral degrees. The selection of indicators was then expanded to include all of the 35 available across the five UMR dimensions. The resulting list of 33 HEI was sorted according to "top scores", and Figure 3 displays the charts for the top three universities.

Figure 3: U-Multirank performance charts for the three top scoring Brazilian HEI offering graduate programs. Interactive version at <https://bit.ly/3fRH30m>



The best-ranked institutions are shown from Figures 3a to 3c, with the Higher School of Advertising and Marketing (ESPM) as the top-performing HEI in the country. That is a surprising result. While ESPM is a traditional institution with undeniable quality from over 70 years of experience, it is nevertheless a specialised HEI, offering nine undergraduate courses and five graduate programs in communication-related areas. Three of the institution's PPG offer only master's courses, all ranked "regular" by CAPES (grade 3). The other two programs count with doc-

torates as well and are ranked as "very good" (grade 5). ESPM may be considered a midsize HEI, and it counts with campuses in four different cities (MEC, 2021; CAPES, 2021d; INEP, 2020).

As a direct comparison, the institution shown in Figure 3c – São Paulo State University (UNESP) – counts with 141 PPG (including 139 master's and 116 doctoral courses), 50 of them ranked as "very good", while 27 are considered of excellence (receiving the top grades 6 or 7). At the undergraduate level, UNESP offers 136 courses, 31 assessed as

“excelent”, and 104 as “very good” by the Brazilian Ministry of Education. Besides that, the institution has 34 campuses across 24 different cities (MEC, 2021; CAPES, 2021d).

This very superficial comparison tells one main story: ESPM and UN-ESP are in two different categories, and they should be compared addressing their differences. From an institution list built without purpose, just selecting every possible measure without filtering for subject area, institution size, legal status and other potential indicators, the results become less significant. A proper list should be built with intentionality, exploring institutional profiles from a combination of desired characteristics (e.g., ESPM excels in marketing, communication, and design, but would not even be listed should the user be interested in health sciences).

While analysing the multidimensionality of rankings, Moed (2017, p. 987) concludes that they only allow “looking into the outside world through a few vertical splits in a fence, one at the time. In this sense, these systems are still one-dimensional”. A consequence of that perspective is that users also become responsible for the proper use of rankings, as they must decide the best way to look through the fence so they can see what is relevant for them. The problem for U-Multirank in Brazil is that the reduced number of institutions, most with data unavailable for many indicators, leads to few and narrow splits in the fence, making it very hard to see any clear picture on the other side.

2.3 THE PROPOSED MODEL

The concept of a multidimensional assessment for the SNPG matured over the course of more than two years, also counting with the lessons from decades of a robust evaluation system. Many actors and organisations are involved in the process, and views of what the system could and should become are not always uniform. Despite that, the PNPG Committee (2020) presented CAPES’ Higher Council with its final proposal for a multidimensional evaluation. The document was unanimously approved and the proposal was given a finality that even contradicted the committee’s original expectations (Audy, 2020).

According to Audy (2020, 27:55), the proposal was intended as one of many contributions for the improvement of evaluation in Brazil, as the committee never had the ambition of being in charge of producing a new model by itself, even considering the multiple contributions from the involved organisations. The actual model would come from the work of CAPES and the academic community shaping those inputs. While that might have once been the intention, the idea of a full-fledged evaluation model could not be avoided from its approval by the top instance of CAPES’ management, which is its Higher Council.

Audy (2020) also mentioned there were few changes from the previous report by the PNPG Committee (2018), but a significant one was the adjustment of the five original dimensions to fit those adopted by U-Multirank, despite minor nomenclature variations. Now, the new model would consider: Personnel training; Research; Innovation and knowledge transfer; Societal impact; Internationalisation. In complement to that, some of the core suggestions from the PNPG Committee (2020) are listed in summary below:

- i) **Grading system** – The result of the evaluation will no longer consist of a single grade for a graduate program. Each of the five dimensions will be graded separately on the already discussed scale from one to seven.

- ii) **Accreditation** – In the existing single-grade model, the minimum grade required to renew a PPG accreditation has been three (considering the 1–7 scale). Requirements are yet to be defined in the new system, but the committee suggests three to remain the lowest possible grade for what it considers core dimensions: “personnel training” and “research”.
- iii) **Indicators** – The proposal includes a series of indicators as suggestions for the assessment of each dimension. While some of them would be new to the Brazilian evaluation, especially those regarding “innovation and knowledge transfer” and “societal impact”, most are well established from previous cycles. A major suggestion is that indicators should be universal to all disciplines, and custom ones would not be allowed.
- iv) **Funding** – The new model should be taken into account in funding strategies for research and graduate education, as it is suitable for diversification. Regardless, it should not be the only guidance in the decision-making process.
- v) **Self-assessment** – An institutional strategic plan should be a fundamental requirement in the evaluation process, serving as a reference for a self-assessment process within the PPG. That should be an essential component for evaluating each dimension of the new model.

While the PNPG Committee (2020) includes other suggestions for the new evaluation model, they are not pivotal to the multidimensional proposal.

3 ASSESSING THE PROPOSAL

The proposal for a multidimensional evaluation of graduate education in Brazil considers that “several recommendations constitute important paradigm shifts and require time for implementation” (PNPG Committee, 2020, p. 27). Because of that, the proposed changes were to be implemented only for the following cycle (2021–2024). However, despite that ambitious statement, one of the findings from this study is that the new model does not change much from the evaluation already in place, and it wastes the opportunity to promote an actual multidimensional assessment.

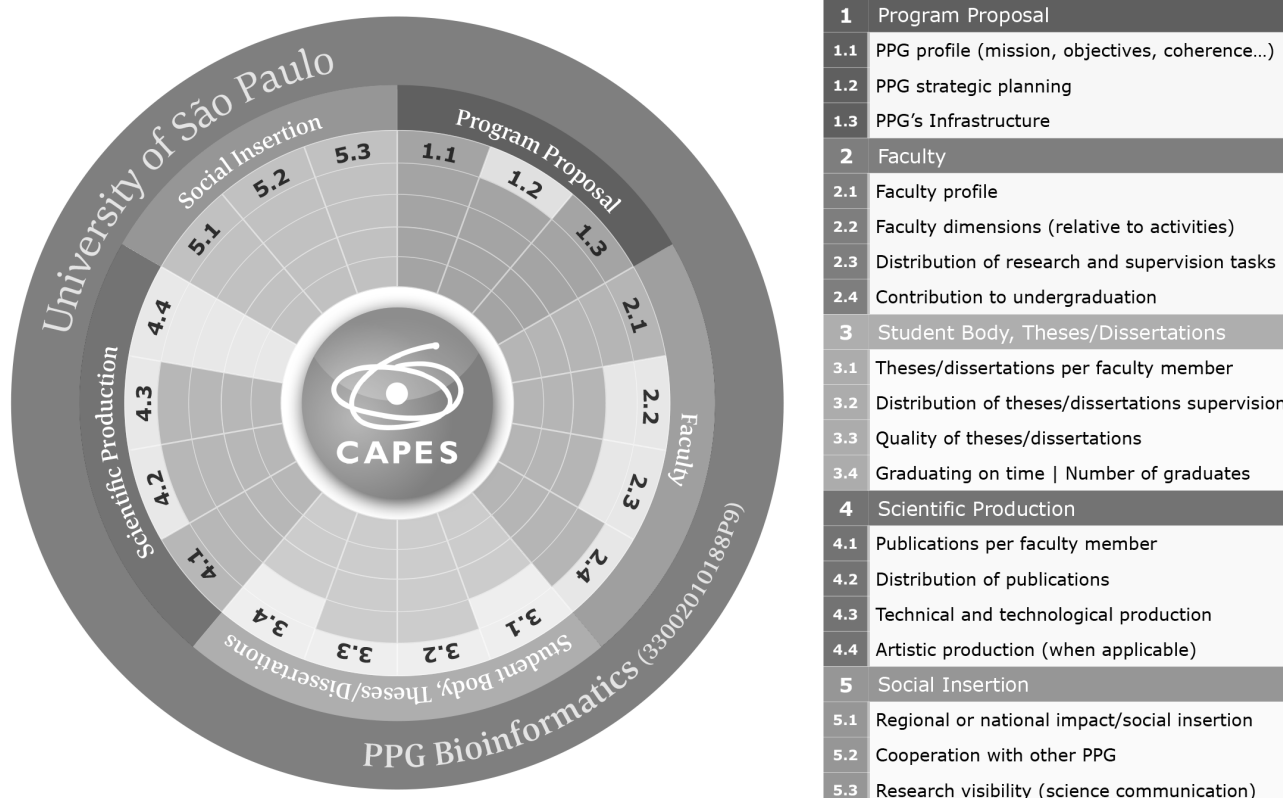
3.1 EVALUATION WAS ALREADY MULTIDIMENSIONAL

One of the essential principles behind Brazilian evaluation is that it should be comparative, so the SNPG can have a transversal equivalence among graduate programs from different disciplines. Thus, a PPG in mathematics is expected to present the same level of quality as one in sociology, provided they have the same grade and respecting inherent characteristics of each area (CAPES, 2010).

To make that possible, CAPES standardised its assessment form in 1998. Each discipline could adapt the proposed indicators to their reality, but they should assess the same set of items: seven in the first version. That number was reduced to five in a subsequent revision (2005–2007), and the form went through additional adjustments with every cycle. The version adopted at the 2017 quadrennial evaluation consisted of two levels: 18 subitems organised into five items. Each subitem is graded on a five-level scale: “insufficient”, “weak”, “regular”, “good”, and “very good”. (CAPES, 2010; Monteiro et al., 2019).

Despite some terminology differences, the similarity to U-Multirank's methodology is quite apparent. To demonstrate that, Figure 4 shows how the UMR sunburst could be applied to visualise the 2017 evaluation results of the graduate program in Bioinformatics from the University of São Paulo (USP).

Figure 4: Results from CAPES' 2017 evaluation of the PPG on Bioinformatics (USP), transposed to U-Multirank's sunburst



The Bioinformatics program shown in Figure 4 was selected at random from PPG graded four in the 2017 evaluation. The grade was chosen because a multidimensional profile would be easier to visualise for a program with an intermediate level of quality than for one that has weaknesses in all dimensions or excels at everything. The PPG is part of the "Biological Sciences I" evaluation area in CAPES' classification system and counts with a master's and a doctorate in genetics. Its evaluation report is publicly available at CAPES (2021b).

As displayed in Figure 4, the PPG had seven subitems evaluated as "very good" (e.g., 1.1), six as "good" (e.g., 1.2), four as "regular" (e.g., 2.2), and one was not applicable (4.4, regarding artistic production). The sunburst shows the distribution of strengths across the five dimensions, and the evaluation report reflects that perception in the aggregation of results, attributing "very good" to dimensions 1 and 5, and "good" to the remaining three. After weighting the five-dimensional results, the PPG received grade four as its final result.

Aggregation and weighting of the assessment items, while transparent through the public regulation of evaluation, have been a major

problem in the process. For instance, Marques et al. (2020) performed a statistical analysis to map the probability for each of the five items to influence grade changes in the PPG assessed in 2017. The conclusion was that "scientific production" and "student body, theses/dissertations" had the most impact to achieve better grades, while "social insertion" was of no relevance across all 49 disciplines in CAPES evaluation. Regulations stated that item should count for at least 10% of the PPG grade, and every committee kept that at the minimum so that programs like the one in Bioinformatics could not benefit from the top performance it displayed in that dimension.

The objective of this study is not to discuss the quality of the assessment form that was used during the quadrennial evaluation of 2017, especially since that has already been revised by Monteiro et al. (2019) for the evaluation planned for 2021. The goal here is merely to show that the evaluation process in Brazil has already been multidimensional, even though the results are not.

3.2 MISSED OPPORTUNITIES

Understanding how the evaluation has been organised around an assessment form reveals that the new multidimensional model changes very little in the process. The only real difference is that the results will come from one step before the usual final grade calculation. Thus, considering the inspiration from U-Multirank and its message against composite scores and nontransparent aggregation, opportunities seem to be wasted in the Brazilian proposal.

As it happens in UMR, end-users should be able to select the indicators (or subitems) that would help them understand the profiles of the graduate programs according to their interests. While this user-driven flexibility could increase the complexity of the evaluation process, it would nevertheless produce richer results. Besides that, the proposed rigidity of indicator selection imposed for all disciplines takes even more multi from the multidimensional.

When Moed (2017, p. 987) analysed five of the most prominent rankings of today, he concluded that “a system should not merely present a series of separate rankings in parallel but rather a dataset and tools to observe patterns in multifaceted data”. Without that, a national evaluation system that should strive to go beyond what university rankings can offer, ends up behind what U-Multirank already does.

While the Brazilian multidimensional evaluation should go beyond what has been proposed, it is crucial to recognise the multiple grade system as an advancement. The unique composite scores that aimed to define the quality of a graduate program were too outdated, and in previous evaluations a PPG with top performance in social insertion would not be valued as much as another with significant scientific production, as the weight applied to the items in the final grade was unbalanced: usually of 10% for the first and 35% for the later (CAPES, 2017). Why should a PPG’s primary mission always be expanding the frontiers of knowledge and never focus on regional or societal impact? A five-grade system is a modest but relevant step to allow graduate programs to find their own identities.

Regardless, the proposed model overlooks another significant opportunity: the lack of attention to self-evaluation. In its 28 pages, the proposal by the PNPG Committee (2020) dedicated only a single paragraph to the issue, while it could be the most powerful instrument in a genuinely multidimensional assessment.

4 SELF-EVALUATION

As stated early in the paper, a self-evaluation strategy has been an integral part of the intended evolution of the Brazilian assessment of research and graduate education. When multidimensionality became a part of that, it was clear that the two initiatives should walk together, but this has not been the case so far.

According to Trevisol and Brasil (2020) there is little literature investigating self-evaluation from the perspective of the SNPG, and the system had almost no experience with those practices. Despite higher education institutions regular development of Institutional Development Plans (PDI), that knowledge was rarely applied in the planning and monitoring of PPG. Thus, while the working group created at CAPES to propose a self-assessment methodology for graduate programs faced a challenge, it also had the opportunity to build something new.

Through Verhine et al. (2019), the working group reported its find-

ings on self-evaluation, proposing strategies for its adoption in the assessment of PPG. The central concept was that each graduate program would implement a custommade process capable of capturing relevant aspects of its mission and objectives, including societal impact, international profile, and distinct scientific decisions. Furthermore, the proposed Brazilian self-evaluation came from a grounded understanding of the SNPG and international inspiration. One of the highlighted countries was the Netherlands, where the group recognised self-evaluation as a unique process, as it is the core of the national assessment and a pivotal instrument leading to the improvement of the country’s research units.

The model currently in use to assess the quality of research in Dutch universities is based on six-year cycles, and it is known as Strategy Evaluation Protocol (SEP). An essential lesson from such a model is in its collaborative design since SEP is a joint effort by the Association of Universities in the Netherlands (VSNU), the Netherlands Organisation for Scientific Research (NWO), and the Royal Netherlands Academy of Arts and Sciences (KNAW). With a focus on three dimensions (viability, research quality, and societal relevance), a self-evaluation report is prepared by research units in light of their mission and strategies. Reports can include appropriate indicators to support the presented narrative, but no uniform measure of success is prescribed. That means each unit can choose the best metrics that serve as evidence of its performance, provided they keep away from indicators such as the Journal Impact Factor (not allowed) and the h-index (strongly discouraged). An external assessment committee is then appointed to analyse the self-evaluation document and, after a site visit, an assessment report is produced (VSNU et al., 2020).

Another example mentioned by Verhine et al. (2019) comes from Finland, where a *benchlearning* system was implemented. In it, developing research units could seek established ones for active interaction and commitment to mutual development (Leite et al., 2020). Such a strategy would be very beneficial in an asymmetric country like Brazil, where distinct levels of scientific maturity are observed across PPG. Because of that, it would directly align with the working group’s proposal for a formative self-evaluation where complementary site visits could serve as the external assessment element but also as a means of support by more experienced PPG to developing ones (Verhine et al., 2019).

While Verhine et al. (2019) brought additional inspiration and presented a strong proposal for self-evaluation, the current implementation follows the path of the multidimensional model, with missed opportunities. For the current assessment cycle (2017–2020), self-evaluation has been relegated to adding two subitems to the new assessment form proposed by Monteiro et al. (2019). Each subitem – “strategic planning” and “self-evaluation” – would have a recommended minimal weight of 10% only, putting the effort in danger of becoming statistically irrelevant to the final result, as Marques et al. (2020) measured to be the case for “social insertion” in the previous assessment form.

Verhine (2020) recognises that the initial implementation of self-evaluation is very modest. The working group coordinator states CAPES’ Higher Education Council (CTC-ES) believed institutions should have time to adapt and build internal assessment infrastructures before the changes significantly impact their grades. The benefit of the initiative right now is in changing the institutional mindset so that a broader self-evaluation model could be implemented in the future. That seems to be a sound decision for the 2021 evaluation. The problem is that the allegedly comprehensive multidimensional model is already planned for 2025, so why is the next level self-assessment not a part of that?

5 CONCLUSION

Leite et al. (2020) see evaluation as a political act that interferes in the life and existence of the evaluated, whether they are people, courses, institutions. CAPES and the evaluation community are aware of that responsibility, so they have strived to evolve assessment strategies continuously, sometimes through minor adjustments to match the dynamic nature of the Brazilian National System of Graduate Education, but other times as fundamental paradigm shifts (Brasil, 2020; CAPES, 2010).

Considering the dimensions of the SNPG and the ever-increasing effort to assess its performance and impact, a central evaluation system was becoming impossible to maintain, and it was progressively unable to capture the complexity of Brazilian science. Therefore, it was time for those paradigm shifts regarding the established one-dimensional perception of quality and the role the evaluated could play in their own evaluation. Hence, a multidimensional approach to quality was required, as was outsourcing part of the assessment task to higher education institutions and graduate programs (PNPG Committee, 2018; Faljoni-Alario et al., 2018; FOPROP, 2018; Verhine et al., 2019).

From the analysis of documents, reports, guidelines, and legislation regulating Brazilian evaluation, this study concludes that the multidimensional model proposed fails to meet expectations. Despite being inspired in part by U-Multirank, the model is based on universal indicators for all disciplines and continuing weighting and aggregation of results (leading to a grade for each of the five suggested dimensions). Furthermore, while improving the items and subitems of the assessment form to reflect the new dimensions, the new model changes very little in the evaluation process. As this study shows, any multidimensionality proposed was already a part of the system, only lost through weighting and aggregation. Those have improved but will still be there, hindering future results as well.

While the multidimensional proposal took a step towards addressing one of the longest standing flaws in the Brazilian evaluation – which was applying a broad benchmarking strategy to a heterogeneous system – it could have benefited from further inspiration. For instance, from U-Multirank, it could have learned to drop the aggregated grading system altogether. Likewise, inspiration from models such as the Dutch Strategy Evaluation Protocol could lead to a revision in the fixed indicator perspective, as each PPG could have the freedom to choose the ones that could better serve as evidence of their accomplishments. However, that would be an impossible strategy unless self-evaluation becomes part of the process, as it was initially intended.

The self-evaluation proposed by Verhine et al. (2019) seems to be sound and well structured, and it deserves more protagonism in the overall assessment process. Self-evaluation should not be limited to a couple of underweighted subitems; it should mean freedom for graduate programs to present unhindered narratives of their accomplishments, based on their missions, and supported by evidence that makes sense, instead of indicators that are easy to compare. It should allow PPG to be different, knowing that the quality of their work would not need to be reshaped to fit formulas incapable of measuring diversity. Finally, it should allow multidimensionality to manifest as more than a set of disconnected resulting grades but as a process that recognises and values dimensions in the core of the evaluation process.

Multidimensionality is necessary for a better evaluation in Brazil, and it is only truly possible through self-evaluation. The strategies comple-

ment each other and must be articulated in such a way as to heighten their potential individual effects.

ACKNOWLEDGEMENTS

The author would like to thank the following scholars for their inspiring comments and suggestions on the first draft of this paper: Frans Kaiser, Ludo Waltman and Robert Tijssen (from U-Multirank), Robert Verhine and Denise Leite (from the Brazilian self-evaluation effort) and Sarah de Rijcke (from CWTS).

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

- Audy, J. L. N.** [Unesp Oficial]. (2020). O novo modelo de avaliação da CAPES [webinar]. Retrieved August 13, 2021, from <https://youtu.be/TyNrLlilnHU>
- Balachevsky, E.** (2005). A pós-graduação no Brasil: Novos desafios para uma política bem-sucedida. In S. Schwartzman & C. Brock (Eds.), *Os desafios da educação no Brasil* (pp. 276–304). Nova Fronteira.
- Brasil, A.** (2020). Building a national system of research and graduate education: How did the university become the house of science in Brazil? *Revista Nupem*, 12(27), 222–253.
- BRASIL.** Ministério da Educação. (2021). Cadastro e-MEC de Instituições e Cursos de Educação Superior. <https://emec.mec.gov.br>
- Calero-Medina, C., López-Illescas, C., Visser, M. S. & Moed, H. F.** (2008). Important factors when interpreting bibliometric rankings of world universities: an example from oncology. *Research Evaluation*, 17(1), 71–81.
- CAPES.** (2010). Plano Nacional de Pós-Graduação - PNPG 2011-2020 (Vol. 1). CAPES.
- CAPES.** (2017). Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal (legislation). *Diário Oficial da União*. Brasília, DF.
- CAPES.** (2018). Avaliação Quadrienal em números (Rep.). Brasília, DF. <https://bit.ly/2Y2rPvO>
- CAPES.** (2019). Aprimoramento do processo de avaliação da pós-graduação (Rep.). CAPES. Brasília, DF. <https://bit.ly/3wDRgDU>
- CAPES.** (2021a). Avaliação Quadrienal. Retrieved June 18, 2021, from <https://bit.ly/35vw16S>
- CAPES.** (2021b). Sistema de Informações Georreferenciadas [Geocapes]. Retrieved August 20, 2021, from <https://geocapes.capes.gov.br/>

Comissão Especial de Acompanhamento do PNPg 2011-2020. (2018). Proposta de aprimoramento do modelo de avaliação da PG (Rep.). CAPES. Brasília, DF. <http://shorturl.at/glvxE>

Comissão Especial de Acompanhamento do PNPg 2011-2020. (2020). Proposta de aprimoramento da avaliação da pós-graduação brasileira para o quadriênio 2021-2024: Modelo Multidimensional (Rep.). CAPES. Brasília, DF. <https://bit.ly/3wyr097>

Conselho Federal de Educação. (1965). Parecer CFE nº 977, de 3 de dezembro de 1965. Estabelece a definição dos cursos de pós-graduação no país (legislation). Brasília, DF.

Faltoni-Alario, A., Junior, C. F. d. S., Brito, E. P. Z., Gontijo, J. A. R., Romero, M. A., Santos, P. J. P. d. & Canuto, S. (2018). Avaliação da pós-graduação: Considerações do CTC-ES (Rep.). Brasília, DF.

Fanelli, A. (2016). U-Multirank and Latin American Universities. *International Higher Education*, (84), 7–9.

Federkeil, G., File, J., Kaiser, F., van Vught, F. & Ziegele, F. (2012). An interactive multidimensional ranking web tool. In F. van Vught & F. Ziegele (Eds.), *Multidimensional ranking: The design and development of u-multirank* (pp. 167–177).

Federkeil, G., Kaiser, F., van Vught, F. & Westerheijden, D. F. (2012). Background and design. In F. van Vught & F. Ziegele (Eds.), *Multidimensional ranking: The design and development of u-multirank* (pp. 85–96). Springer.

Ferreira, M. d. M. & Moreira, R. d. L. (Eds.). (2002). CAPES, 50 anos: Depoimentos ao CPDOC/FGV. FGV.

FOPROP. (2018). Avaliação da pós-graduação: Proposições à Comissão Especial do PNPg. Brasília, DF.

Gadd, E., Holmes, R. & Shearer, J. (2021). Developing a Method for Evaluating Global University Rankings. *Scholarly Assessment Reports*, 3(1), 1–19.

Hazelkorn, E. & Gibson, A. (2017). Global science, national research, and the question of university rankings. *Palgrave Communications*, 3(1), 21.

INEP. (2020). Censo da Educação Superior 2019. Retrieved August 17, 2021, from <https://bit.ly/3ABVsFK>

Leite, D., Verhine, R., Dantas, L. M. V. & Bertolin, J. C. G. (2020). A autoavaliação na Pós-Graduação (PG) como componente do processo avaliativo CAPES. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 25(2), 339–353.

Marques, C. A., Veiga, A. d. C. & Borges, L. M. C. (2020). A avaliação da pós-graduação no Brasil: resultados e determinantes da avaliação da Capes (2013-2016). *Revista Meta: Avaliação*, 12(37), 979–1003.

Marques, F. (2019). Avaliação em cinco dimensões. *Revista FAPESP*, (286), 28–31. Retrieved June 16, 2021, from <https://bit.ly/3iNeIKW>

Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). *Revista Brasileira de Sociologia*, 6(13), 9–26.

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

Monteiro, A. L., Brasil, A., Horta, B. L., Mamiya, E. N., Federizzi, L. C., Verhine, R. E. & Júnior, W. R. d. S. (2019). Ficha de Avaliação (Rep.). Brasília, DF. <https://bit.ly/2PmAf0q>

Trevisol, J. V. & Brasil, A. (2020). A autoavaliação nas políticas da pós-graduação brasileira. *Anais da XIII Reunião Científica da ANPEd-Sul*.

U-Multirank. (2021). U-Multirank: Universities compared your way. Retrieved August 9, 2021, from <https://www.umultirank.org>

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

van Vught, F., Westerheijden, D. F. & Ziegele, F. (2012). Towards a new ranking approach in higher education and research. In F. van Vught & F. Ziegele (Eds.), *Multidimensional ranking: The design and development of u-multirank* (pp. 1–7). Springer.

Verhine, R. E. [Unesp Oficial]. (2020). Autoavaliação dos PPG e planejamento estratégico institucional da pós-graduação [webinar]. Retrieved August 20, 2021, from <https://youtu.be/GNdYu18KRts>

Verhine, R. E., Lima, A. C., Carvalho, D. B. B., Leite, D., Bertolin, J. C., Dantas, L. V., Oliveira, N., Oliveira, R., Avellar, S. & Rosa, D. L. (2019). Autoavaliação de Programas de Pós-Graduação (Rep.). Brasília, DF. <https://bit.ly/3jTSRS8>

VSNU, KNAW & NWO. (2020). Strategy Evaluation Protocol (2021–2027) (Rep.). The Hague. <https://bit.ly/3wAFbzi>

Waaiker, C. J. (2018). Why the public nature of university rankings torpedoes their value for benchmarking: an institutional researcher's opinion. *Proceedings of the 23rd International Conference on Science and Technology Indicators*, 1031–1037.

AUTHOR

ANDRÉ BRASIL

Centre for Science and Technology Studies (CWTS), Leiden University
Willem Einthoven Building, Kolffpad 1
2333 BN - Leiden, The Netherlands
E: a.l.brasil@cwts.leidenuniv.nl
ORCID: 0000-0003-1832-5199

KEYWORDS: policy analysis, evidence-informed policy, research evaluation, multidimensional evaluation, self-evaluation.