Advancing the evaluation of graduate education: towards a multidimensional model in Brazil

Brasil Varandas Pinto, A.L.

12 Conclusions & Reflection

> *It is essential to retain the recognition and derived respect that CAPES has earned with the national and international community . . . . For this, it is necessary to continuously improve its performance, particularly by updating the evaluation system.*
>
> — **PNPG Committee (2020)**

Throughout this dissertation, a comprehensive exploration of multiple facets of the Brazilian national evaluation system has been carried out, contributing to a deeper understanding of this intricate landscape. The chapters offer a nuanced perspective, beginning with a broad overview of the origins and development of the Brazilian science system and its evaluation practises, and further by conducting international comparisons, such as contrasting the Brazilian evaluation with the Dutch Strategy Evaluation Protocol (SEP). Subsequent chapters delve into specific dimensions of evaluation, including scholarly output classification, and the imperative for evaluation to account for regionally significant research not adequately covered by international databases.

The core objective underlying this discourse revolves around shedding light on the complexities inherent in the evaluation of graduate education, advocating for a more balanced and comprehensive system. To accomplish this, a combination of qualitative methods, such as document analysis and participant observation, and quantitative methods, primarily within the domain of scientometrics, have been employed. Multiple sources of evidence and critical analysis have been incorporated, encompassing relevant scholarly literature, public letters, interviews, and a vast array of primary sources, such as legislation and policy documents.

Throughout the investigation, several crucial concerns have been highlighted. For example, the movement to impose indicators like the Journal Impact Factor to replace the autonomy previously enjoyed by peer review committees has been brought to attention. Imperfections have also been discovered, including the need to reform the disciplinary classification system adopted in the country. However, notable achievements have been acknowledged, such as the role of evaluation in fostering diversity in scientific output, as technical and technological productions gain recognition as viable alternatives to traditional paper-based output. Moreover, potential avenues for further development have been elucidated, particularly with regard to the exploration of a multidimensional evaluation model. For that, a critical assessment of previous initiatives is presented, along with a proposal to prioritise self-evaluation as the primary approach toward achieving a truly multidimensional evaluation framework. In summary, the unique context of Brazil and the significance of locally relevant research remain central considerations, emphasising the need for a balanced and inclusive evaluation system as a consistent thematic thread.

This concluding chapter provides an opportunity to integrate reflections, insights, and recommendations into a framework that aligns with the trajectory of the Brazilian National System of Graduate Education (SNPG) and its evaluation. The foremost ambition is to synthesise these findings and shape them into actionable advice to progress toward a multidimensional evaluation model. The proposed shift aims to accommodate the sprawling complexity and diversity of the academic landscape of the country, with the multidimensional approach being a consensus echoed in numerous analyses conducted over the past few years (Barbosa, 2020; PNPG Committee, 2020; Faljoni-Alario et al., 2018).

12.1 Self-reflection from the Leiden Manifesto

The Leiden Manifesto for research metrics has been referenced numerous times throughout this project, with distinct purposes (Hicks et al., 2015). In the Introduction, it plays two roles: it emerges as a catalyst for a series of opportunities that culminated with this work, but it also inspired a moment of thoughtful debate among the evaluation leadership and representatives of all graduate programs (PPG) in Brazil. The discussions eventually spurred a large-scale movement to rethink the Brazilian evaluation system, as detailed in Chapter 3.

However, the CAPES evaluation leadership conducted a self-reflective exercise to structure those initial dialogues. Their chosen task was to critically analyse current assessment practises in light of the principles of the Leiden Manifesto. The outcome of this introspection, to the best of my knowledge, was never published. Instead, the exercise was primarily confined to slides used to engage with the evaluation community. Having been an integral part of these conversations, I retrieved the results from my archives so that they could become a part of the recorded memory of the Brazilian evaluation, as laid out in Table 12.1.

For each principle of the manifesto, Table 12.1 presents three elements derived from the reflection process led by CAPES in 2015. The first includes observations on ongoing activities that are part of the CAPES evaluation and were considered to align with the respective principles. Additionally, the table references the type of information used to meet each principle, followed by a performance score self-attributed to the various evaluation processes conducted. These ratings are rather revealing of CAPES' perception of its activities. By assigning a score of six or seven for alignment with the manifesto's principles, the agency reveals self-confidence, indicating that its performance exceeds the average in all aspects. However, it also acknowledges that its performance merely exceeds sufficiency, with ample room for improvement across all aspects.

With regard to the type of information available, the adherence to the principles of the Leiden Manifesto is partially dependent on qualitative data, which are featured for 70% of the principles. However, these data are never used in isolation; they always complement quantitative data. This widespread reliance on a quantitative approach, supporting all principles of the manifesto, demonstrates that the self-assessment performed aligns with discussions throughout this dissertation, particularly in Chapter 4, where the effect of high-stakes evaluation in Brazil on the type of information that can be used was discussed. When an evaluation has significant implications for funding, prestige, and the very existence of graduate programs, dependence on quantitative analyses is accentuated, as it makes the evaluator's job easier when justifying results.

Turning to the included observations, the centrality of the assessment form in the evaluation process becomes evident. With the many mentions of its role in the responsible evaluation process, the self-reflection process appears to confirm that a core principle of Brazilian evaluation is the quest for a dynamic equilibrium between data, peer review, and criteria, as discussed in Chapter 3.

Table 12.1.: Compilation of the self-reflective exercise conducted by CAPES on the ocasion of the publication of the Leiden Manifesto.

#	Principles of the Manifesto	CAPES reflections on principle alignment	Information	Score
1.	Quantitative evaluation should support qualitative, expert assessment.	Data are available on the Sucupira Platform. Several items on the Assessment Form that guide expert committees are based on calculations and indicators.	Quanti	6
2.	Measure performance against the research missions of the institution, group or researcher.	Data are available on the Sucupira Platform. Different parameters are used for each area, listed in the publicly available Area Documents. The evaluation does not follow a single model.	Quanti/Quali	7
3.	Protect excellence in locally relevant research.	Data are available on the Sucupira platform. Each area starts the evaluation through the analysis of area-specific indicators, resulting in its own Qualis.	Quanti	6
4.	Keep data collection and analytical processes open, transparent and simple.	The Sucupira Platform aims to make graduate education data available for public consultation.	Quanti/Quali	6
5.	Allow those evaluated to verify data and analysis.	The Sucupira Platform enables the auditing of the data at any time. The mid-term seminars promoted by CAPES with PPG directors also contribute to the audit process.	Quanti/Quali	7
6.	Account for variation by field in publication and citation practices.	Data are available on the Sucupira Platform. Evaluation areas use different indicators and instruments, depending on the type of production: Qualis Journals, Books, Artistic and Technical/Technological.	Quanti/Quali	6
7.	Base assessment of individual researchers on a qualitative judgement of their portfolio.	Qualitative information on faculty members regarding background and experience is available on the Sucupira Platform and is used in the Assessment Form.	Quanti/Quali	7
8.	Avoid misplaced concreteness and false precision.	Several indicators are used to determine a PPG grade, as described in the Assessment Form.	Quanti/Quali	7
9.	Recognize the systemic effects of assessment and indicators.	Several indicators are used to determine a PPG grade, as described in the Assessment Form.	Quanti/Quali	7
10.	Scrutinize indicators regularly and update them.	The assessment in each area is benchmarked, with metrics calculated from an analysis of the area as a whole. The process and resulting metrics are outlined in the evaluation report for each area.	Quanti	7

In the first instance, there is considerable reliance on collecting, processing, and treating research and graduate education data in the country. Such data are made available to expert committees who, benefiting from a system structured around 49 evaluation areas (see Chapter 6), can orchestrate a peer review process that contextualises available quantitative and qualitative information. Additionally, committees rely on the evaluation form to guide the application of criteria and indicators, ensuring not only the consideration of a diverse range of necessary dimensions, but also facilitating the process of replication, adoption by different evaluators, and comparability across disciplines or evaluation areas.

12.2 Findings, or how would evaluation fare today?

This dissertation concludes with a synthesis of findings and forward-looking strategies that align with the evolution of the Brazilian National System of Graduate Education (SNPG) and its evaluation. The body of work, which features articles published across various journals and conferences over a four-year span, unfolds a detailed narrative divided into four parts.

Part I served as a backdrop, introducing the SNPG and its evaluation system as a central focus. Inspired by the theory of path dependence and its implications for public policy reform, Chapter 2 offered a deep exploration of the Brazilian science system, tracing its origins and growth influenced by pivotal policy decisions. An analysis of the reasons behind these decisions, their ties to broader public policy frameworks, and their impact on the top-down structure of the Brazilian science system was presented. The focus shifted in Chapter 3 as it traversed the historical landscape of the evaluation of research and graduate education in Brazil. From its humble beginnings as a tool for funding allocation, it has grown into a multifaceted construct integral to the national science landscape. Despite its transformative role in promoting transparency, fairness, and academic quality, the system is on the brink of further evolution, which requires sophisticated reform strategies to be implemented to guarantee its future.

Part II continued the exploration of path dependency, using the approach as a lens to compare the Brazilian evaluation system with some of its global counterparts. Chapter 4 presented a side-by-side comparison of research evaluation systems in Brazil and the Netherlands. This comparison highlighted unique

attributes, challenges, and outcomes, resulting from different historical, geographical, and policy contexts. While the performance-based Brazilian system advances the nation's scientific pursuits, it also fosters research homogeneity, potentially inhibiting innovation. Conversely, the Dutch system emphasises research quality and societal relevance, fostering diverse research trajectories. Lessons from both systems are valuable, yet their very different backgrounds caution against a universal strategy. Chapter 5 harnessed a plethora of resources and data sources to dissect the design ethos of the SNPG centred on graduate education, revealing its above-average influence on academic publishing in Brazil, offering fresh perspectives on the contrasts between the SNPG and other science systems.

Part III delved into the complexities and multifaceted nature of the Brazilian evaluation system. Chapter 6 analysed the Brazilian classification of research and graduate education, illustrating how evaluations unfold within 49 areas shaped by disciplinary and managerial perspectives, leading to consequential results that are intrinsically influenced by these dynamic landscapes. Chapter 7 dived into the Brazilian Qualis system, a critical tool for assessing the quality of scholarly publications. While this system balances quantitative and qualitative methodologies, the case is made for enhancing it further towards a more comprehensive evaluation framework.

Chapters 8 and 9 offered two different, yet interconnected, perspectives on the Brazilian evaluation system. Chapter 8 highlighted the importance of regional databases and publications in the local language, advocating comprehensive assessments to capture the complexity of Brazilian science. In contrast, Chapter 9 discussed the financial constraints faced by nations like Brazil and proposed alternative publication paths aligned with the country's economic reality, while also relying on its long-standing experience in diamond open access practises. Finally, Chapter 10 reviewed efforts to diversify Brazilian research production, focussing on an evaluation effort to induce PPG to invest more of their time and efforts in outputs of a technical and technological nature. The results have shown that evaluation indeed has a significant steering power that should be further exploited to enhance the societal impact of Brazilian science.

The final Part IV confronted some of the challenges facing the evaluation system due to the exponential growth of the SNPG. In this sense, Chapter 11 emphasised the risk of overreliance on quantitative indicators and advocated

a multidimensional approach to accommodate the complexities of Brazil's academic landscape. A failed attempt by CAPES to implement a multidimensional model inspired by uMultirank underscores the dangers of simplistic solutions to complex evaluation problems. As a solution, came a recommendation of a shift towards self-assessment, emphasising the need for further involvement of the country's higher education institutions (HEI) in the much needed exercise of their institutional autonomy.

Now, the concluding analysis of this dissertation refers back to the guiding principles of the Leiden Manifesto. As a framework for the preliminary self-reflection exercise shown in Table 12.1, an inpiration to discuss the advancement of evaluation in Brazil, and a catalyst for this research project, the manifesto can now help distill the key findings of the investigation in line with its guiding principles. The synthesis aims not only to encapsulate the essence of the scholarly journey but also to offer insights into the underlying research and to outline a strategic direction for the future evaluation dynamics of the SNPG.

1. Quantitative evaluation should support qualitative, expert assessment

When the Brazilian evaluation system started to develop in the 1970s, a preliminary attempt was made to construct a system grounded in quantitative data. However, that quickly changed to a more adequate system that relied on the expertise of specialist committees (Chapter 3). The genesis and growth of these committees were examined in detail in Chapter 6, which concludes that the arrangement of the evaluation system in the evaluation areas of CAPES represents a significant achievement. This is due to the ability of expert panels in these areas to carry out their evaluations based on quantitative and qualitative data. Preserving the diversity of these areas is vital, yet a reform in the current composition of 49 areas was noted to be necessary. This need arises from the evolutionary process over the years, often driven by administrative and practical considerations, which have created disciplinary distortions. These require adjustments for better international equivalence.

The worth of evaluation committees is also made evident in Chapter 7, which refers to the classification of Brazilian scientific output in periodicals. This is also discussed in Chapter 4, where the role of the committees in handling bibliometric and scientometric data in the evaluation process is explored. However, although the crux of Brazilian evaluation lies in expert assessment supported by available data, recent developments have put part of established practises

at risk. For example, Chapters 3, 7 and 8 highlight that committees have lost some autonomy in interpreting available data and exercising their judgment. For example, in the Qualis classification of journals, committees receive recommendations based on sets of indicators and can make adjustments if necessary. However, the latest rules restrict the number of permissible modifications and their extent. Consequently, the evaluation system is moving opposite to the critical principle of the Leiden Manifesto, which recommends that indicators should not substitute informed judgment and peer review should be strengthened instead of giving decision making to numbers (Hicks et al., 2015).

2. Measure performance against the research missions of the institution, group, or researcher

Hicks et al. (2015) underscores the importance of stating the objectives of a program at the start of an evaluation process and ensuring that the indicators adopted to assess performance are related to these objectives. Regrettably, in the context of Brazilian evaluation, this principle is poorly addressed. As underscored in Chapters 3 and 6, the active evaluation model in Brazil operates on a comparative basis within each evaluation area. Consequently, there is a propensity to employ indicators and criteria with minimal variation within each area, ensuring that the same ruler is adopted to measure all graduate programs. An exception is depicted in Chapter 2, which outlines the establishment of the Brazilian National System of Graduate Education (SNPG). The exception pertains to the professional modality of PPG, which was implemented in Brazil in 1998 to stimulate more applied research.

Chapter 3 explains how a distinctive assessment form was devised and adopted to evaluate professional programs. In addition to that, Chapter 10 describes how adjustments were made to the evaluation process to expand recognition of technical and technological production to complement the value given to more traditional research outcomes, usually bibliographic production. Such initiatives enable evaluators to account for the applied dimension of professional programs compared to more academically orientated programs. However, consideration of specific missions and objectives is often overlooked.

In that sense, Chapter 4 contrasts the Brazilian evaluation model with the Dutch model, demonstrating how the latter accommodates the missions and objectives of the programs due to its structure centred on self-evaluation. This allows each assessed unit to select the indicators best aligned with the narrative

of the conducted evaluation. However, this model is only feasible due to the nature of assessment in the Netherlands, where benchmarking is an optional component and the impact of outcomes on resource distribution is negligible.

Although acknowledging that the impact of Brazilian evaluation needs to be reconsidered to allow more significant consideration of program objectives and more custom-made indicator application, Chapter 11 proposes broadening the initial self-evaluation experiences recently undertaken in Brazil. This would afford greater freedom in the selection of indicators by the PPG themselves, highlighting the quality of the work conducted.

A critical conclusion drawn from this research is that a Brazilian evaluation that does not incorporate flexibility and customisation will continue to promote homogenisation, and the SNPG will be unable to expand with the diversity of approaches and objectives necessary to meet Brazil's research needs.

3. Protect excellence in locally relevant research

The Leiden Manifesto cautions against the pressure to prioritise the English language in scientific output, warning that it could precipitate negative implications for conducting locally relevant research (Hicks et al., 2015). Chapter 8 devotes significant attention to this issue, exploring it from two distinct perspectives. The first concerns language, demonstrating that almost half of the Brazilian scientific production, as quantified by data collected by CAPES, is carried out in Portuguese. The second facet of the analysis reveals that most of this locally produced research is not captured by predominant databases used in the assessment processes, such as the Web of Science (WoS).

Conversely, a considerable amount of local language output, addressing locally relevant topics often in the social sciences and humanities (SSH), is indexed in regional databases like SciELO, Latindex, and RedALyC. In this regard, the Qualis journal classification, introduced in Chapter 3 and extensively discussed in Chapter 7, has been an outstanding solution to reward excellence in non-English publications and locally relevant research. Despite its recognised imperfections, Qualis is a direct application of the manifesto's first principle, allowing committees to move from quantitative indicators to their quality judgment of all journals utilised by Brazilian graduate programs for publication.

A significant volume of journals for which indicators such as the Journal Impact Factor and CiteScore are unavailable can be qualitatively assessed, often

revealing high-quality levels recognised by expert committees across various evaluation areas. Furthermore, as demonstrated in Chapter 9, a large portion of these high-quality journals operate on the diamond open access model, crucial for a country facing economic and social challenges, such as Brazil.

Regrettably, the recent trend in the evaluation of scientific output in Brazil has succumbed to pressures to valorise output in English, indexed by databases such as WoS and Scopus. This trend is predicated on the argument that Brazil needs to promote internationalisation. However, the observed pathway contradicts the approaches adopted by developed countries like the Netherlands, which abolished the use of indicators such as the Journal Impact Factor in its Strategy Evaluation Protocol (Chapter 4). Unfortunately, Brazil appears to be charting the opposite course, thereby risking relinquishing a significant achievement: its ability to valorise locally relevant output in Portuguese.

4. Keep data collection and analytical processes open, transparent, and simple

A central strength of the Brazilian evaluation approach lies in the quality of its data collection. This subject has been contemplated in various chapters of this dissertation, receiving substantial attention in Appendix A. The appendix's principal goal is to introduce an R package developed to potentialise the use of data provided by one of CAPES' many systems: GeoCapes (CAPES, 2021e). Despite this explicit goal, the paper produced to accompany the package's launch delves into the history, motivation, and development of many of the agency's Current Research Information Systems (CRIS). The most recent is the Sucupira Platform (CAPES, 2021c), launched in 2014 to incorporate essential functionalities for evaluation. Graduate programs can use the platform to report data on publications, infrastructure, faculty, student body, degrees awarded, theses and dissertations produced, and more.

Regarding the analytical processes, CAPES and the representatives of the evaluation areas collectively prepare such processes, together with the criteria and indicators to be adopted in each evaluation cycle. Area documents, as described in Chapters 3 and 6, are published with a process description to be adopted. Notably, each area specifies how it will assess the items included in the evaluation form. Should any postgraduate program feel disadvantaged in the evaluative

process relative to the disclosed criteria and indicators, a request for reconsideration is possible, with evaluators obligated to uphold the commitments made in published documents.

5. Allow those evaluated to verify data and analysis

As explored in Appendix A, data collection was sometimes perceived as a "black box" before the Sucupira Platform. Graduate programs (PPG) provided their information, and data gaps became apparent during the evaluation process, often requiring reliance on institutional reputation to fill these gaps. With the advent of the Sucupira Platform, CAPES' evaluation data transparency reached an apex. For example, data collected from PPG are almost immediately made available for public access on the platform itself, being subsequently published as consolidated spreadsheets in a complementary open data system (CAPES, 2021a). This process allows faculty and students to verify the information entered at any time, and the widespread availability of data facilitates an auditing system across the academic community.

The investments made by CAPES to collect, process, and publish evaluation data are unparalleled. For example, in the 2017 evaluation, the transparency process evolved to such an extent that the spreadsheets with microdata used in the evaluation process were pre-disclosed, accompanied by the Qualis classification results, which directly impacted the programs' scoring and final results.

However, as discussed throughout the dissertation, especially in Chapter 3 and the Introduction, a unique crisis hit the evaluation in the last cycle, and the high level of transparency observed in the previous cycle was not replicated. In the most recent evaluation conducted in 2022, the microdata made available to the evaluation committees were not disclosed, and the Qualis classification was only made public after the results of the evaluation process were published. The decline in transparency in the Brazilian evaluation has an undeniable impact on the Brazilian National System of Graduate Education (SNPG), and the evaluation system moves, yet again, in the opposite direction of a principle of the Leiden Manifesto.

However, it is relevant to mention that the loss of transparency has been partially a result of judicial interference in the evaluation process – situation described at length in Chapters 1, 3 and 4 – and the advent of the General Data Protection Law (LGPD), the Brazilian version of the General Data Protection

Regulation (GDPR). Although not as strict as its European counterpart, LGPD has imposed some significant limitations on the transparency of the evaluation data available in the country. For example, before LGPD, microdata available on faculty and the student body of graduate programs included information on gender, making it possible to perform relevant analyses on the gender balance of faculty members, scholarship distribution, PhD success rate, time to graduation, and more. After LGPD, the information had to be removed, even from previously available datasets, making studies such as the one carried out in Brasil (2021d) no longer possible at the same level of precision[1].

6. Account for variation by field in publication and citation practises

Again, the Qualis classification is the primary instrument for ensuring the proper treatment of disciplinary variations in publication and citation practises.

From a citation perspective, Qualis journal classification plays a central role in accounting for variations in the 49 evaluation areas. The procedure starts with a quantitative indicator analysis followed by peer review within each area, taking into account disciplinary practises. However, as delineated in Chapter 7, the process is not without flaws. For instance, the model continues to classify scientific production using journals as a proxy when technological advancements and database enhancements already enable at least part of the task to be performed at the article level. An additional issue lies in handling classifications by evaluation areas that occasionally encompass diverse citation practises in their subareas. For example, as discussed in Chapter 6, one of the existing evaluation areas includes architecture and design – two disciplines with markedly distinct citation practices – which invariably complicates the work of the committees in effectuating the internal classification of their journals. Moreover, the classification model is heavily predicated on the use of indicators such as the Journal Impact Factor, CiteScore, and H-Index – all metrics with limited capacity to appropriately manage citation variations across different disciplines, thus burdening the expert committees with the entire responsibility of judgment.

Regarding the differences in publication practises among disciplines, Qualis emerges again as the solution, but this time in its various iterations. As explored in Chapters 3 and 7, the original Qualis was developed for journal eval-

[1] Algoritms to infer gender from names in Brazil have demonstrated an accuracy of around 85%, which is now a limitation to gender studies against the previously available information for every individual.

uation. However, variations have been implemented over time to assess books, events, artistic production and, more recently, the technical and technological production discussed in Chapter 10.

Evaluation areas have the autonomy to adopt Qualis variations that capture the diversity of their publication practises. Areas such as History and Education classify books as one of the pillars of their scientific production evaluation. Computer science has often classified events to value conference proceedings. This diversity is precious to the Brazilian evaluation process, presenting another reason to protect the evaluation areas. As described in Chapter 6, a proposal from PNPG Committee (2020) suggests a significant reduction in the number of areas active today. While the need for restructuring has become evident from this research, a mere decrease in the number of areas represents a regression in one of the main achievements of the Brazilian evaluation model.

However, as valuable as evaluation areas can be in addressing disciplinary variations, the model must also be perfected within areas with better methods and indicators. For instance, Hicks et al. (2015) mentions how a group of historians working for a psychology department were unfairly evaluated because their books were not considered in the assessment prioritising WoS-indexed papers. This is a recurrent situation in the Brazilian system, as evaluation areas such as environmental sciences, public health, and biotechnology are inherently multi- and interdisciplinary. The system even includes an area named interdisciplinary, accounting for more than 300 graduate programs in the country. So, while the geography area can be equipped to consider the disciplinary variations within its realm, the same may only be possible for some areas with significant advancement in the adopted methods.

7. Base assessment of individual researchers on a qualitative judgment of their portfolio

A frequently invoked argument by CAPES to justify certain practises in its collective model is predicated on the fact that the evaluation carried out by the agency is not individual-focused, but rather group-focused. As discussed in Chapters 2 and 4, the unit of evaluation in Brazil is the graduate program. Thus, any potential injustices that may arise in the process are somewhat diluted.

For example, the fact that Qualis classification is conducted based on journals can undoubtedly penalise high-quality articles published in periodicals consid-

ered of lower quality. For example, low-quality articles occasionally benefit from being published in high-quality journals, while high-quality articles may be penalised in the reversed situation. According to the rationale of the Brazilian evaluation system, the impact of such unfairness is diluted when evaluating a group consisting of several PPG researchers that potentially publish hundreds of articles over a four-year evaluation cycle. Therefore, the discussion of this manifesto principle would not necessarily apply to this group evaluation setting.

Nevertheless, the concept of portfolio judgment applies to CAPES and illustrates a positive movement towards a fairer evaluation in Brazil. The notion of portfolios, as opposed to exhaustive evaluation of long lists of publications, was introduced to CAPES through the accreditation process of graduate programs (PPG), as described by Brasil (2019). The implementation of such an approach was not straightforward, as the academic community in Brazil needed to become more accustomed to constructing portfolios. Therefore, proposals for PPG accreditation were sometimes accompanied by portfolios containing only high-impact articles, rather than those that best demonstrated the group's potential for conducting the proposed research. The ramifications of this challenge were evident in professional programs, as the submitted portfolios were predominantly composed of academic production, often failing to demonstrate the researchers' capabilities for conducting the desired applied research they were proposing to pursue.

In time, the process evolved and the academic community began to adapt. Soon, graduate programs started creating portfolios during data collection for their periodic evaluation, identifying outstanding products among the different types of scholarly output reported. Indeed, in Chapter 10, a document analysis on the evaluation of technical and technological production reveals that 19 areas chose to employ only qualitative methods in their analysis, focussing exclusively on portfolio assessment. The remaining 30 evaluation areas informed that they would adopt a mixed-method approach, complementing a quantitative analysis of the PPG output with a qualitative approach to their portfolios.

Although portfolio adoption marks a significant development in the Brazilian evaluation system, Hicks et al. (2015) spotlights the preference to adopt an approach that considers a more comprehensive understanding of an individual's expertise, experience, activities, and influence, even when comparing large numbers of researchers. Although this advice can be partially achieved

by applying the assessment form for the PPG evaluation, as it includes many dimensions to be considered, there is room for improvement by fostering the self-assessment practises advocated in Chapter 11.

8. Avoid misplaced concreteness and false precision

Viewed from the lens of the Leiden Manifesto principle, CAPES has developed an evaluation system that genuinely avoids the trap of false precision. Instead of utilising numeric indicators in its evaluation processes, the primary inputs for assessment, represented mainly through various types of Qualis, generally employ percentiles. For instance, the Qualis journals adhere to an eight-tier scale, each encompassing approximately an octile of the distribution. These classifications, in turn, feed into evaluating the items and subitems on the assessment form, along with qualitative elements and scientometric indicators related to degrees awarded, supervisions, and more. Each of these aspects is appraised on a five-tier scale, ranging from insufficient to very good.

The final result of the assessment is then reflected on a scale of 1 to 7, with the highest grades (6 and 7) representing excellence and 3 to 5 depicting a range from regular to very good. PPGs that achieve any of these grades have their accreditation renewed for the subsequent four-year cycle, but grades 1 and 2 lead to program closure. As can be noticed, these results may be considered an instance of misplaced concreteness since the difference between a grade of 1 or 2 is immaterial; in both instances, the punishment is closure, with no distinction derived from the separate grades.

Returning to the comparison made in Chapter 4 between the evaluation models of Brazil and the Netherlands, we see that the current Dutch model does not assign grades to evaluated units. However, this is a recent achievement of a predominantly formative evaluation model (VSNU et al., 2020). Yet, the Dutch evaluation protocols active until 2015 scored programs on a five-tier scale, which were condensed into four in the 2015-2021 protocol. This could serve as an example for the Brazilian evaluation, which might adopt the four categories once used in the Netherlands: world-leading/excellent (1); very good (2); good (3); unsatisfactory (4).

Moreover, as discussed in Chapter 11, it would be desirable if this scoring could be applied independently to each dimension of the evaluation. For example, in the Dutch case, the four-tier scale was applied to research quality, societal

relevance, and viability. This approach would allow for a more comprehensive and nuanced understanding of the multidimensional nature of academic and scientific performance.

9. Recognize the systemic effects of assessment and indicators

Hicks et al. (2015) state that indicators prompt systemic transformations through the incentives they introduce and their repercussions should be anticipated. The Brazilian case, as illustrated in Chapter 10, provides robust evidence that this anticipation is recognised and actively exploited. CAPES acknowledges its evaluation's role as a catalyst shaping graduate programs' behaviour. Therefore, by assigning a higher value to technical and technological production, the agency stimulates the diversification of research output, reaching broader audiences, and consequently amplifying the impact of Brazilian science. As evidenced by this study, the initiative exhibits promising signs of success. Although a more extended period is needed to measure the true impact of this policy, an increase in targeted productions is already discernible.

Despite acknowledging its influence, the evaluation has been slow in transitioning to the aspired multidimensional model, despite knowing that, in not doing so, it continues to foster the undesired homogenisation of graduate education in Brazil (Chapter 11). Furthermore, CAPES has demonstrated a keen awareness of its evaluative power through its misguided attempt to promote the internationalisation of Brazilian science (Chapters 7 and 8). Although the purpose is admirable, the evaluation has tried to accomplish the goal by prioritising the adoption of indicators derived from international databases and attenuating the value assigned to non-English publications indexed in regional databases only.

From this perspective, it is apparent that the inductive power of evaluation is not only recognised but has been strategically harnessed by CAPES. However, like most powerful policy instruments, it can be a catalyst for good and evil; it can be wielded to foster a scientific system capable of focussing on both quality and regional relevance, but it also bears the potential risk of misapplication, thereby stifling the development of such a system.

10. Scrutinize indicators regularly and update them

Throughout this dissertation, a recurring theme has been CAPES' continuous quest to evolve its evaluation model. While acknowledging the influence of path

dependence and avoiding excessively disruptive changes, a prevailing sentiment persists that each evaluation cycle should supersede the previous one. More than rethinking indicators, the overall approach of the system is consistently scrutinised.

As part of the scrutiny process, indicators and criteria are often re-examined. Sometimes, this leads to advancements, but unexpected consequences are occasionally experienced. However, the system and its leaders are usually open to embracing positive changes and recognising unsuccessful attempts. Consequently, the process refines itself through further adjustments to optimise the evaluative process within known constraints of human and financial resources or time restrictions.

However, it is essential to note that the constant review process was one of the triggers that led the Federal Public Prosecutor's Office (MPF) to bring a civil action against CAPES, resulting in an injunction that temporarily halted the evaluation process (Justiça Federal, 2021). The legal action accused the agency of neglecting the legal security and expectations of those evaluated by how it dealt with the establishment, disclosure, and transparency of the evaluation parameters.

The list of demands that integrated the injunction required CAPES to provide a complete list of "evaluation criteria", "types of production and strata", and the "cutoff scores" being used for evaluation, broken down by evaluation area. The key to the legal security approach was in the cutoff scores, as the expectation was that an evaluation should not be comparative, as is the case in the Brazilian system. Therefore, PPG should be able to know beforehand the final grade they would be awarded when reaching predetermined cutoff scores for each indicator, whether those are the number of publications, degrees awarded, the average time to graduation, and more.

Regrettably, despite the ongoing advocacy from numerous actors and organisations within the Brazilian scientific community to preserve and continuously evolve the CAPES evaluation model, the system now faces a considerable hurdle (Chapter 3). There is an imminent risk that it may devolve into a mere temporally static checklist exercise, thus reducing Brazilian science to a relentless pursuit of predefined metrics. As a result, the essence of its comparative approach is under threat, necessitating a robust defence against such a simplistic reduction and an affirmation of the system's dynamic and evolving nature.

12.3 Recommendations

The findings of this dissertation have brought to light several critical areas of concern and consideration in the landscape of the evaluation of Brazilian research and graduate education. As a sequel to the rigorous analysis, this concluding section attempts to ascribe pertinent propositions that can be deployed to enhance the current framework. These recommendations encapsulate the essence of the preceding chapters and are meticulously tailored to resonate with Brazil's distinctive socio-cultural contours. However, readers are advised to draw upon the detailed discussions in the previous chapters to gain a more detailed understanding of each recommendation and its contextual relevance.

1. The establishment and evolution of the science system, graduate education, and evaluation model in Brazil are not arbitrary; they are the products of formative historical paths. Their understanding is paramount for any assessment or debate about the value, role, or needs of evolution.
2. An overarching sense of the structural and historical intricacies that have shaped the current landscape of the SNPG and its evaluation system should be the foundation of any reformative attempt. Only through understanding that some apparent peculiarities serve as structural pillars for the whole system can any devised changes be able to foster advancements that address the system's inherent strengths and idiosyncrasies.
3. Recognising the inherent uniqueness of the Brazilian scenario is paramount, particularly when contrasted against globally accepted alternatives. Although foreign science and evaluation experiences may offer a framework for advancement, it would be inappropriate to duplicate them entirely onto the country's context. The pursuit of constructive adaptation requires understanding these international instances as sources of inspiration. However, any direct implementation could compromise the nuances and complexities that the Brazilian system embodies.
4. It is vital to preserve a Brazilian evaluation system that relies on a wide range of data but is based on the work of expert committees organised in evaluation areas. Although there is a clear need to restructure and streamline the areas, this must be backed by a comprehensive study engaging the academic community. Furthermore, this initiative should aim to balance international orientations and appreciation of the national context.

5. Brazil's pioneering role in institutionalising mechanisms to measure the quality of diverse scientific outputs should be recognised and strengthened. This includes the classification of articles, books, conference proceedings, and artistic, technical, and technological products. Promoting this diversity is essential, as it allows the valuation of different research profiles across different modalities and types of graduate programs.
6. In balancing the use of data, peer review, and the adoption of criteria comparable between programs and disciplines, the autonomy and power of peer review must be maintained. Quantitative indicators can provide reference points and support evaluators so that they are not biased or misled in their analyses. However, quantitative indicators should never dominate the evaluation process.
7. While many developed countries seek solutions to avoid reductionist indicators and value diversity in their scientific production, Brazil is moving in the opposite direction. Most countries that continue to employ indicators such as the Journal Impact Factor – against recommendations from the San Francisco Declaration on Research Assessment (DORA) (ASCB, 2012) – do so because of need rather than choice. Brazil does not suffer from the same lack of options, as the Qualis system can allow qualitative analysis of committees to supersede any quantitative imposition. Thus, embracing indicators the developed world tries to abandon does not pave the way to internationalisation. Brazil must recognise that it is ahead rather than behind on this issue.
8. Brazil has made commendable strides with its Current Research Information Systems (CRISs) and Science & Technology databases. However, the potential of these systems should be harnessed to incorporate open databases such as OpenAlex into the evaluation process. Furthermore, the same motivation should lead CAPES to help advance and connect regional databases such as SciELO, Latindex, and RedALyC.
9. Brazil's experience with open access (OA) is another area that deserves attention. The country should leverage its two-decade experience with Diamond OA to elevate national scientific quality. In this sense, evaluation has the ethical responsibility to support these long-established publishing practises, which deserve worldwide promotion, as they are still in their infancy in many developed nations (Ancion et al., 2022).

10. While CAPES negotiates the country's first transformative agreements for OA publishing, the Brazilian socioeconomic landscape and its well-established diamond OA model should be the foundation for bargaining with publishers. Awareness of the country's publication dynamics, investments made to access scientific publishing behind paywalls, and the high costs of Article Processing Charges (APCs) should also inform the negotiations. Nominal exchange rates should never guide these since actual costs go beyond the mere dollar value of the investment made – metrics such as the Purchasing Power Parity (PPP) should be adopted instead. In any case, evaluation should address any injustices derived from the financial dimensions of producing science, open or otherwise.

11. The power of the Brazilian evaluation system to stimulate change, positive or negative, is substantial. Despite continuous improvements, the current model still promotes homogeneity and limits the space for diversity and innovation. A multidimensional evaluation system is essential to reverse this picture, as it allows the valuation of different research profiles. In this way, the SNPG has the necessary space to address various pressing societal and developmental challenges.

12. CAPES must acknowledge that its current evaluation system is nearing exhaustion. With the expansion of the SNPG, it is no longer possible to continue to expand the centralised and top-down evaluation model practised, especially in the search for a multidimensional alternative. Therefore, it is essential to work in partnership with higher education institution (HEI) and National Forum of Pro-Rectors for Research and Graduate Education (FOPROP) to develop alternative self-assessment strategies that encourage diversity in Brazilian science and can employ the institutions themselves as more active partners in the evaluation process.

13. CAPES should also enhance the formative element of its evaluation. For example, suppose that the Bemelmans-Videc et al. (2010) approach to policy instruments is considered. In that case, the direct results of the evaluation efforts conducted by the agency have been primarily based on "carrots" – such as additional funding and the gain in reputation from high grades – and "sticks" – linked to funding cuts or the discontinuation of a PPG. However, approaches to educate and guide graduate programs about the importance of particular behaviours need to be developed. Admittedly, one of the evaluation results is the production of individual reports for

the PPG, and an occasional site visit may be recommended. However, a change of focus – where evaluation is held not to punish or to reward, but to educate and guide – is more than welcome. This reorientation will undoubtedly better equip Brazilian science to tackle pressing societal problems.

Although these recommendations provide a strategic compass, they must be considered in light of the complexity and diversity of the Brazilian graduate system. A one-size-fits-all approach is likely to fail to deliver equitable results. Hence, it is crucial to rethink evaluation in a flexible and adaptive way, allowing for variation and customisation in line with the specificities of individual graduate programs, in harmony with broader institutional and disciplinary contexts.

12.4 Limitations

Although this study provides valuable insights into the Brazilian graduate evaluation system, its interpretations and recommendations should be considered in light of certain limitations. First, as delineated in the Introduction, there is a significant gap in academic and non-academic publications addressing graduate evaluation in Brazil. This deficiency, in part, emanates from an inherent shortfall within CAPES in publishing detailed accounts of its proceedings, whether due to time constraints or the inherent opacity often associated with governmental agencies.

Thus, the attempt to construct a comprehensive understanding of the conception, development, and current status of the SNPG and its evaluation was primarily contingent on the documentary analysis of accessible official documents, directives, and legislation. This methodology, although comprehensive, was restricted by the extant material and interpretation thereof. Such an approach might only partially represent the practical execution of the evaluation system and the experiences of those directly involved. This gap was somewhat bridged through the availability of historical interviews and the reliance on academic literature authored by active participants in the evaluation process.

Furthermore, as a scholar, maintaining a neutral position throughout the study was paramount. Despite having experience within the Brazilian evaluation of graduate evaluation, that meant leaving out any anecdotal accounts that might

enrich the narrative, yet would lack more substantial corroborating sources beyond potential testimonial evidence. Additionally, during my decade-long tenure with CAPES, I have exerted varying degrees of influence on the evaluation system. As such, I have been a witness, advocate and adversary to many integral parts of the system today. Despite taking pride in some outcomes and harbouring regrets about others, I strived to maintain scientific rigour in every analysis carried out throughout the research process. The intention behind this recollection is not to vent, but to acknowledge the inherent difficulties in analysing a system to which one has partially contributed, for better or worse.

From a practical standpoint, this study faced some difficulties due to the already-discussed crisis that has affected Brazil in recent years. Challenges included a one-year delay in the 2021 quadrennial evaluation, which hindered the field research planned to help mature and further contextualise some of the research results found. Additionally, delays in the planned publication of evaluation data and results – in some cases extending beyond 18 months – together with the 2020 enactment of the Brazilian General Data Protection Law (LGPD), created obstacles to the adequate achievement of the necessary data for various chapters of this dissertation.

Leaving the realm of practicalities, while the study proposes overarching recommendations for enhancing the evaluation system, it is relevant to recognise that it does not delve into the intricate operationalisation of these suggestions. This aspect might limit the applicability of the findings in steering specific policy actions and reforms. Further investigation is warranted to expound on how these proposals could be incorporated within the unique context of the Brazilian graduate system, given the country's diverse institutional structures, academic cultures, and socioeconomic realities.

12.5 Future work

This dissertation has attempted to elucidate the multifaceted and challenging nature of the Brazilian evaluation of research and graduate education, traversing various dimensions to grasp the complexities involved in the system's development and current standing. Rather than delving into a specific aspect of

the evaluation process, the ambition was to understand the system as a whole, including the interconnectedness of its various elements.

Taking into account the comprehensive view of the Brazilian evaluation system achieved through this work, there is a significant potential for deepening the knowledge presented here. Additional research can enrich our understanding of how the system works or how to improve it. For example, further exploration of the Qualis classification of journals could consider alternative indicators and databases to provide quantitative evidence for the evaluation process. Similarly, qualitative procedures capable of fairly valuing the substantial article production in Brazil could be investigated in a way that considers both its scientific contribution and its societal impact, whether regional or global.

The comparative analysis performed is another example of an investigation that can pave the way for future research. This could be possible by extending the comparison to other international models or exploring how different countries have operationalised multidimensional evaluation and self-assessment strategies in their contexts. Such exploration could lead to a more nuanced understanding of various research evaluation approaches, thus identifying successful strategies that could be inspirational for the Brazilian context.

Despite the diversity of research avenues explored, a selection of evaluation processes was made to represent the main facets of the system. However, additional themes remain to be explored. For instance, this study has mentioned certain scientific production classification processes only in passing. The Qualis classification of books, among others, deserves further investigation for a better understanding and potential enhancement of the strategies adopted by distinct evaluation areas to value such an essential research product in many disciplines, which is only sometimes duly appreciated in evaluation processes.

From a methodological standpoint, the complexity of the Brazilian evaluation system also opens avenues for research focused on diversifying the subjects of study and deepening what has already been covered. Qualitative research methods such as interviews and focus group discussions could provide deeper insights into the intricate nature of the system, helping identify potential barriers and facilitators to the implementation of proposed reforms.

From a reform-minded perspective, this research aims to catalyse change, and it is clear that the active involvement of key stakeholders is crucial to achieving

this goal. This could include action research projects, where scholars can collaboratively work with institutional leaders and policymakers to design, carry out and critically evaluate pilot reforms influenced by the findings of this study.

An example of this approach, in conjunction with a prominent Brazilian university, is already underway to target the significant role of self-assessment in higher education institutions. This ongoing research intends to elaborate on the frameworks through which self-assessment can be effectively implemented within Brazilian universities, contemplating the intricacies of cultural and institutional factors while serving the multidimensional approach to evaluation. The exploratory phase has the potential to yield valuable insights into fostering institutional autonomy while preserving the integrity and efficacy of the evaluation process.

Another partnership, already forged, aims to apply responsible evaluation perspectives to the internationalisation efforts of another leading Brazilian university. These perspectives are based on a sound understanding of quantitative evidence and its limitations to emphasise the crucial considerations of diversity, inclusivity, and regional relevance as pillars of the evaluation process. The principal objective here is to demonstrate that a successful internationalisation initiative should not be exclusively quantified by co-authored papers or jointly awarded degrees but by a series of institutional transformations that may defy straightforward measurement but merit understanding and fostering.

Pursuing a more comprehensive, equitable, and effective evaluation system should not be considered a finite endeavour but a continuous journey of discovery and refinement. This dissertation represents but a single stage of that journey, a stepping stone towards the ultimate goal: an evaluation system that faithfully mirrors the scope, depth, and richness of the Brazilian National System of Graduate Education, thus enhancing its potential to produce positive societal impact.

References

American Society for Cell Biology. (2012). *San Francisco Declaration on Research Assessment (DORA)*. ASCB. Retrieved April 22, 2023, from https://sfdora.org/READ/

Ancion, Z., Borrell-Damián, L., Mounier, P., Rooryck, J., & Saenen, B. (2022). *Action Plan for Diamond Open Access*. Science Europe, cOAlition S, OPERAS, and the French National Research Agency (ANR). Retrieved from https://bit.ly/3Ob3FID

Barbosa, G. d. R. (2020). *Avaliação multidimensional de programas de pós-graduação*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3hmQuq2

Bemelmans-Videc, M.-L., Rist, R. C., & Vedung, E. (Eds.). (2010). *Carrots, Sticks & Sermons: Policy instruments and their evaluation* (5th ed.). Transaction Publishers.

Brasil, A. (2019). Accreditation of graduate courses in Brazil: analysing the evaluation of the first proposals of professional doctorates in the country. In *17th Conference of the International Society for Scientometrics and Informetrics (ISSI)*, Rome, Italy.

Brasil, A. (2021d). Defeating the gender gap in Brazilian science: Funding strategies to reduce historical asymmetries [Abstract], In *Funded and Unfunded Science Symposium*, Prague, Czech Republic: Czech Academy of Sciences.

CAPES. (2021a). Avaliação da Pós-Graduação Stricto Sensu. Dados Abertos CAPES. Retrieved from https://dadosabertos.capes.gov.br

CAPES. (2021c). Coleta de dados dos programas de pós-graduação. Plataforma Sucupira. Brasília, DF. Retrieved from http://sucupira.capes.gov.br/

CAPES. (2021e). Sistema de Informações Georreferenciadas. Geocapes. Brasília, DF. Retrieved August 20, 2021, from https://geocapes.capes.gov.br/

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*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3wyR097

Faljoni-Alario, A., da Silva Jr, C. F., Brito, E. P. Z., Gontijo, J. A. R., Romero, M. A., dos Santos, P. J. P., & Canuto, S. R. A. (2018). *Avaliação da pós-graduação: Considerações do CTC-ES*. Brasília, DF.

Hicks, D., Wouters, P. F., Waltman, L., de Rijcke, S., & Ràfols, I. (2015). The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Justiça Federal. (2021). *Liminar que suspende a Avaliação Quadrienal de 2021*. Poder Judiciário. Retrieved from https://bit.ly/3tzrmki

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