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

Brasil Varandas Pinto, A.L.

André Brasil

Advancing the evaluation of graduate education

Towards a multidimensional model in Brazil

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

André Brasil

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

Keywords: National Evaluation Systems; Research Evaluation; Graduate Education; Science Policy; Scientometrics; Scholarly Communication; Open Science

Cover design and layout: André Brasil
Cover photo by: iStock.com/hept27
Printed by: ProefschriftMaken
ISBN: 978-94-6469-599-1

This work was carried out at the Centre for Science and Technology Studies (CWTS), Faculty of Social and Behavioural Sciences at Leiden University, The Netherlands. The research was supported by funding from the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES Grant No. 88887.300590/2018-00) and the Leiden University Fund (LUF #4509).

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

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 24 oktober 2023
klokke 10.00 uur

door

André Luiz Brasil Varandas Pinto

geboren te Belo Horizonte, Brazilië
in 1976

Promotor

Prof.dr. S. de Rijcke

Copromotor

Dr. E.C.M. Noijons

Promotiecommissie

Prof.dr. P.F. Wouters (decaan/voorzitter)

Prof.dr. L.R. Waltman

Dr. T. van Leeuwen

Prof.dr. M. Jacob (Lund University, Sweden)

Prof.dr. R.E. Verhine (Federal University of Bahia, Brazil)

Dr. R. Woolley (INGENIO (CSIC-UPV), Spain)

Acknowledgements

Pursuing a doctorate has been a long-standing personal ambition, which I often feared was delayed too long. However, the experience I have gained over the years, particularly during my decade as a policy officer at the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES), has undoubtedly enriched this academic pursuit. At CAPES, I assumed numerous leadership roles within the country's evaluation system and contributed to significant discussions about the trajectories of research and graduate education in Brazil. Although those experiences alone merit immense gratitude, it is worth noting that without the financial support and leave granted by CAPES, this PhD journey would not have been possible.

The initial foundation for my PhD was laid as I represented CAPES in the 21st International Conference on Science, Technology, and Innovation Indicators (STI 2016). My mission was twofold: to foster connections with the authors of the Leiden Manifesto, a document deeply resonant within Brazilian science policy, and to engage the broader research evaluation community. Of these interactions, my dialogue with Ed Noijons stands out. His deep understanding of the Brazilian research landscape and his enthusiasm for deeper exploration made me realise that he would be an ideal guide for my doctoral journey.

And so it was. Upon completing my ongoing master's degree, I approached Ed with a PhD proposal aimed at enhancing the evaluation of research and graduate education in Brazil. Ed opened the doors of CWTS to me. With his mentorship, unwavering support, and trust in the unique perspective my policy experience offered, I found the motivation to continue moving forward.

Sarah de Rijcke's inclusion in my supervisory team brought a new dimension to my research. As my promotor, she played a pivotal role in broadening my horizons, guiding me into the realm of qualitative research techniques, and thus diversifying my foundational background in practise, policy, and econometrics. This deepened understanding, further cultivated during my studies at the Netherlands Graduate Research School of Science, Technology, and Modern Culture (WTMC), elevated my research to a depth and breadth previously unimagined.

Next, I express my deep appreciation to Ludo Waltman. While he did not serve in an official supervisory role, his consistent guidance was invaluable throughout my doctoral journey. His sharp insight helped me navigate numerous challenges, allowing me to refine my methods and more effectively convey the intended messages. Collaborating with him on various projects has been a highlight of my time at CWTS.

Indeed, while the doctoral journey was intrinsically rewarding, my time at CWTS has been enriched by my involvement in various projects. One pivotal experience was collaborating with Rodrigo Costas, Tjitske Holtrop, and Leonie van Drooge in the working group that assisted our board of directors in formulating the centre's self-assessment. The insights and knowledge derived from this venture significantly shaped multiple chapters of my dissertation.

Furthermore, being a part of the CWTS Open Science Research Hub allowed me to actively contribute to the centre's Open Science policies and engage in various activities alongside many esteemed colleagues. In particular, I thank the hub's leader, Thed van Leeuwen, for the multifaceted experiences we have shared in Open Science, our coauthored works, seminars presented, and above all, the unwavering friendship and support he has provided throughout.

I extend heartfelt gratitude to Ingeborg Meijer for her invaluable friendship and leadership in some of the most rewarding research projects I have participated in, particularly New HoRRIzon and SUPER MoRRI. Collaborating with her and colleagues like Inge, Anestis, Ralf, and Richard has solidified my sense of belonging in the academic community. Our shared aspiration to ensure that research has a tangible impact on society has been particularly meaningful to me.

I also wish to express my gratitude to my colleagues at CWTS-BV, the commercial branch of the centre. It has been my conviction that applying academic insights to real-world challenges not only manifests research in practise, but also informs and inspires future academic endeavours. A special note of thanks goes to Mark for adeptly steering the BV activities and to my colleagues for their continual camaraderie. Collaborations on different projects with peers

like Clara, Vincent, Nees, and Wolfgang have been among some of the best moments of the journey.

Although I would love to continue acknowledging all the people who contributed to the success of this academic endeavour, it is time to dive into the heart of this dissertation. But before we move on, I must extend heartfelt thanks to Esther Feer. With her amazing patience and warmth, she helped my family and me learn not only a bit of the Dutch language but also much of the country's culture. More than that, she has become one of our most treasured friends in our new homeland. Thank you, Esther, for your kindness and for helping us maintain our sanity through challenging times.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABC	Brazilian Academy of Sciences
ABE	Brazilian Education Association
ANPEd	National Association of Research and Graduate Studies On Education
APCs	Article Processing Charges
API	Application Programming Interface
BNDE	National Bank for Economic Development
CAPES	Brazilian Agency for Support and Evaluation of Graduate Education
CDI	Commission for Industrial Development
CFE	Federal Education Council
CNE	National Education Council
CNPG	National Council of Graduate Education
CNPq	Brazilian National Council for Scientific and Technological Development
CONFAP	National Council of State Funding Agencies
CRAN	Comprehensive R Archive Network
CRIS	Current Research Information System
CSV	Comma-separated Values
CTC	Technical & Scientific Council
CWTS	Centre for Science and Technology Studies
DOAJ	Directory of Open Access Journals

DOI	Digital Object Identifier
ESPM	Higher School of Advertising and Marketing
FINEP	Funding Authority for Studies and Projects
Fiocruz	Oswaldo Cruz Foundation
FOPROP	National Forum of Pro-Rectors for Research and Graduate Education
GDPR	General Data Protection Regulation
GTC	Technical Advisory Group
HBO	Applied Sciences Universities (*hoger beroepsonderwijs*)
HEI	Higher Education Institution
IBGE	Brazilian Institute of Geography and Statistics
INEP	National Institute of Educational Studies and Research
ISCED	International Standard Classification of Education
KNAW	Royal Netherlands Academy of Arts and Sciences
Latindex	Regional Cooperative Online Information System for Scholarly Journals From Latin America, the Caribbean, Spain and Portugal
LGPD	General Data Protection Law
MEC	Ministry of Education and Culture
MNCS	Mean Normalised Citation Score
NSF	National Science Foundation
NWO	Netherlands Organisation for Scientific Research
OECD	Organisation for Economic Co-operation and Development
PDI	Institutional Development Plan
PNPG	National Plan for Graduate Education
PPG	Graduate Program
PPP	Purchasing Power Parity

RedALyC	Network of Scientific Journals of Latin America and the Caribbean
RoR	Research On Research
SBPC	Brazilian Society for the Advancement of Science
SciELO	Scientific Electronic Library Online
SEP	Strategy Evaluation Protocol
SNPG	Brazilian National System of Graduate Education
SoSP	Science of Science Policy
UMR	U-Multirank
UnB	University of Brasília
UNESP	São Paulo State University
UNL	Universities of the Netherlands
USP	University of São Paulo
VSNU	Association of Universities in the Netherlands
WO	Research Universities (*wetenschappelijk onderwijs*)
WoS	Web of Science

1 Introduction

> *Without evaluation, there is no quality at the graduate education level, which is where scientific research is conducted in Brazil.*
>
> — **Renato Janine Ribeiro**

Evaluation plays a crucial role in ensuring the quality and advancement of the Brazilian science system. It provides a framework for institutions, policymakers, and stakeholders to thoroughly assess the effectiveness, impact, and overall performance of the country's graduate education[1] system, which serves as the bedrock for research in Brazil. The evaluation system also exerts a tangible influence on the allocation of funding, institutional reputation, and continued accreditation of graduate programs. Consequently, evaluation practises wield considerable power in shaping the course of research and graduate education in the country (Balbachevsky and Schwartzman, 2010; R. J. Ribeiro, 2022b).

The duty of conducting the national evaluation system was entrusted to the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) in the 1970s. As an agency dedicated to fostering excellence in research and graduate education, CAPES is committed to continually improving its evaluation practises, propelling the progression of academic standards (Verhine and Dantas, 2009). It is within this ever-evolving landscape that the genesis of this dissertation can be found, and its starting point was triggered by the publication of the Leiden Manifesto for research metrics (Hicks et al., 2015).

[1] In this dissertation, "graduate education" refers to studies following a bachelor's degree, specifically master's and doctoral programs. In many parts of the world, the term "postgraduate" is often used interchangeably with "graduate".

Upon the publication of the Leiden Manifesto, its comprehensive principles, emphasising transparency, diversity, and contextuality in evaluation practises, captivated the attention of CAPES' president. Consequently, the publication was widely shared with the directors of more than 4000 graduate programs nationwide. Recognising the need to align the CAPES evaluation system with the principles outlined in the manifesto, the agency also embarked on a self-reflection exercise, with subsequent working groups being established to examine various dimensions of the existing framework, which encompassed subjects such as journal classifications and societal impact.

CAPES has often been mindful of contributions made by prominent institutions engaged in research evaluation, indicator development, scientometrics, and science policy. While the Centre for Science and Technology Studies (CWTS) naturally fell within this scope, the publication of the manifesto further reinforced the agency's attention towards the diverse developments emanating from Leiden. Consequently, CAPES decided to designate a representative to attend the 21st International Conference on Science, Technology, and Innovation Indicators (STI 2016), intending to establish connections with the authors of the manifesto and other researchers whose work served as valuable evidence for our policy decision-making. Given my active participation as a policy officer involved in national evaluation efforts, the responsibility of representing the agency in this capacity was assigned to me. This unique opportunity, coupled with the insights gained from examining Brazilian evaluation through the lens of the Leiden Manifesto, served as a driving force that propelled me on the research journey that culminated in the development of this dissertation.

1.1 Motivation and Problem Statement

According to Farmer (2010), policymakers rarely reach out to academic experts before formulating or promoting policies. When they do, the purpose is often to find justification for positions they already hold, rather than looking for objective analyses. Confirmation bias can also be an issue, as the sheer volume of information available makes it inevitable that they rely on trusted sources, often aligned with their ideas (Andrews, 2017). Therefore, time can be the greatest challenge in digesting evidence. Legislators, politicians, and policymakers rarely have the time or support personnel necessary to synthesise data to inform decisions (Gaieck et al., 2020). This is notably true if the time frame in policymaking is considered. Results often need to materialise faster than the time required to carry out any semblance of proper research (Colglazier, 2016).

Then comes the capacity issue. Policymakers should be encouraged to be more science-literate, being able to appreciate not only the role of evidence in their decision-making process, but also to develop methods to separate high- and low-quality sources. The lack of preparation to consume science can be partially solved with the presence of properly qualified personnel, but understanding evidence is an essential quality for policymakers (Cairney and Oliver, 2017). These limitations do not mean that science is not privileged in policymaking. Montana and Wilsdon (2021) state the opposite, pronouncing science to be often prominent but not as central to the discussion as researchers might think. The importance of science is often not self-evident, and influential scientists are those who leave the sidelines and make themselves heard (Gaieck et al., 2020).

In light of these arguments, CAPES could be seen as a deviation from the norm of public institutions. Its leadership, including the agency's president and a substantial portion of its directors, has been routinely comprised of renowned scientists; exactly those leaving the sidelines to get involved in managerial processes. These leaders are equipped to interpret and use scientific evidence to enhance their work. Moreover, they are supported by a team of highly qualified public servants, most of whom boast higher degrees, often up to the doctoral level, thus solidifying the bridge between policy and academia.

In such an environment, trust in science is generally high. However, there is a clear limitation: there seems to be a continuous shortage of time to properly base decisions in research. Reflecting on my own experience and that of my colleagues and leaders, public policy management often hijacks calendars. This most likely leads to our entrapment into an unending cycle of urgencies that pose a critical challenge to translating research into policy action, and vice versa. In the case of CAPES, time constraints also affect the production of knowledge from the evaluation process, as leaders and their teams rarely have time to publish detailed accounts of their plans, motivations, and developments.

The absence of accessible primary sources profoundly influences academic research on evaluation in Brazil. Without access to the intricate workings behind the proverbial curtain, researchers must rely on fragmented, superficial, or even obsolete accounts. As a consequence, part of the research on Brazilian evaluation may be contaminated with inaccuracies, raising the barrier seen by Farmer (2010) between policymakers and academic experts who are not directly involved in evaluation. Furthermore, the extensive literature consulted for this work shows the remarkable scarcity of products in English on Brazilian evaluation.[2]. The language barrier prevents international researchers from developing an interest in the Brazilian system, thus limiting the scope of collaboration.

The underlying rationale of this research project can be succinctly summarised into two key points. These serve as the foundation of this scholarly endeavour and also establish a cohesive narrative that connects the intricacies of the research process with the overarching academic aspirations.

i) This research aims to contribute to a more comprehensive and extensive literature on the Brazilian national evaluation system. This includes an in-depth exploration of its historical trajectory, contemporary challenges, and current status, effectively offering a holistic picture of the evaluation landscape within Brazil. Furthermore, the project also prioritises international accessibility; this research aims to establish itself as a key reference on Brazilian evaluation practises for a global audience. Although the findings will be predominantly distributed in English, an effort will also be made to facilitate access for the Brazilian community by producing multilingual versions of selected research results.

ii) The literature indicates that influence on policy will rarely emerge from outputs such as journal publications, as these are designed to get scientists' attention. Influence comes from policy briefs, reports, participation, advice, and more (Donnelly et al., 2018; Gluckman, 2014). Thus, the ambition of this doctoral research extends beyond the confines of this dissertation to the creation of practical recommendations for evaluation managers within Brazil. This requires providing additional output, such as policy briefs and establishing an accessible online platform featuring interactive dashboards that encapsulate data on Brazilian science, international collaboration, and other relevant statistics. The ultimate objective is to create an information-rich interface that supports informed decision making and could contribute to improving the evaluation process.[3]

[2] See Appendix B (Additional Resources) to access an interactive dashboard designed to explore the literature investigated for this dissertation, including information on document type and language.
[3] Information on the online platform and the policy-brief output are also available in Appendix B.

1.2 Conceptual framework

Two discernible challenges emerged as the project matured from the defined motivations and the preliminary research phase. The first concerns the scope of the project. Instead of focussing on the dissection of a specific component of the evaluation process, this work aspired to provide a diagnostic view of the entire evaluation. The primary rationale for this decision is the observation that many specialised studies tend to ignore that the Brazilian evaluation system may be conceptualised as a “short blanket” predicament, a situation where addressing one issue uncovers another. Numerous studies have fallen into this trap, suggesting solutions that yield significant impacts but inadvertently neglect or exacerbate other problems. In attempting to reform or refine the system, it is crucial to adopt the holistic approach to account for the complex interconnectedness of various facets of the system. Any proposed modification must take into account not only the anticipated benefits, but also potential adverse repercussions that could arise elsewhere within the system.

The other challenge encountered during the first year of research was the quest for a suitable conceptual framework to guide the investigation, as the existing ones seemed incongruous with the unique characteristics of the research object. As explored in Part I and Part II of the dissertation, research exists within the realm of graduate education in Brazil, unlike most global scenarios where graduate education is a component of a broader research environment.

Given the focus on the evaluation of graduate education, the core disciplines essential for this investigation would inevitably be at the intersection of higher education studies and research evaluation. However, understanding the Brazilian evaluation system requires a distinct tilt toward the latter. As further explored in Chapter 3, at a certain point, the assessment of scientific production represented up to 80% of the total evaluation result in Brazil. The weight of graduate education in the country’s science is so significant that Renato Janine Ribeiro, president of the Brazilian Society for the Advancement of Science (SBPC) at the time of this publication, recently stated that "without evaluation, there is no quality at the graduate education level, which is where scientific research is conducted in Brazil" (J. Marques, 2021).

Therefore, this study predominantly anchors itself in the discourse of research evaluation while only tangentially addressing the quality assurance elements characteristic of higher education research. Drawing on the experience of this researcher within the Brazilian evaluation landscape, the lessons of the latter would resonate more with the evaluation of higher education conducted by the National Institute of Educational Studies and Research (INEP) than with that overseen by CAPES, which is the central focus of this research. Given the unique characteristics of the SNPG, emphasising research evaluation becomes logical and imperative.

However, the study often crosses disciplinary boundaries, integrating insights from a myriad of literature within the Research on Research (RoR) domain. For example, scientometrics is prominently showcased in several chapters, not to advance its methodologies, but to harness its state-of-the-art tools to augment our understanding of evaluation practises in Brazil. The history of science also emerges recurrently throughout the dissertation, providing a nuanced account of contextual factors, contingencies, and complexities that shape the evaluation process in the country. By delving into the past, valuable insights are obtained to effectively navigate the intricate landscape of the evaluation of research and graduate education today, while also helping to face the challenges encountered to advance the evaluation process.

Although academic frameworks and interdisciplinary perspectives are instrumental in shaping the narrative and analytical thrust of this work, they alone cannot fully encapsulate the multifaceted nature of the research object. Recognising this limitation, the study extends beyond theoretical paradigms, immersing itself deeply in the empirical realm. This approach resonates with the ideas of Montana and Wilsdon (2021) about bridging disciplinary divides to allow the consolidation of findings and the professionalisation of scholarship in the field of evidence and policy. Thus, grounded empirical perspectives, anchored in tangible data, real-world observations, and first-hand experiences, become paramount in this research.

Given that both the research subject and the delineation of this study's objectives are profoundly influenced by my professional interests and experiences, a shift towards an empirical methodology was anticipated. As articulated by Colglazier (2016), researchers ought to actively engage with policymakers to understand their specific needs, thereby ensuring that their voices resonate when offering evidence or expert advice. As a policy officer, having the privilege to momentarily step back from the day-to-day intricacies of evaluation to under-take independent research during my doctoral pursuit, I find myself uniquely poised to traverse both realms. This positions me to bridge the gap between the practical aspects of policymaking – with its empirical understanding of inherent challenges – and an immersive experience in rigorous academic research.

From this unique vantage point, achieving synergy between my professional insights and the academic rigour essential for a research journey has emerged as a central objective of this study. When executed effectively, this synergy can facilitate a bidirectional learning process. Consequently, this dissertation can not only represent my scholarly evolution but also serve as a conduit for a necessary knowledge transfer from the Global South to the North, thus infusing the academic discourse with fresh and often uncharted perspectives.

Given the challenges encountered and the consequent structural decisions made, the foundational framework of this research project was established near the end of the first year of work. The resulting proposal was presented at a seminar organised by CWTS in early 2020, as shown in Figure 1.1, and can be summarised as follows:

1. Main input – This initial stage entails meticulously collecting various data sources and literature about the Brazilian National System of Graduate Education (SNPG) and its evaluation model;
2. Understanding – The subsequent phase involves developing an in-depth understanding of the SNPG's structure, evolution, and alignment with Brazilian science and evaluation policies;
3. Comparison – The project then moves towards an international comparison analysis with the aim of understanding how the Brazilian system differs and whether inspiration can be extracted from other systems.
4. Lessons learnt – From the analysis of Brazilian and international experiences, the project takes a deeper look into the many facets of Brazilian evaluation in the search for understanding and evolution.
5. A multidimensional model – The project's focus shifts towards an idea for evolution of the evaluation model in Brazil, aimed at further aligning the SNPG with the country's goals within a multidimensional framework.
6. The applied approach – The project faces the challenge of ensuring its relevance and impact on the SNPG, through the production of diverse research outputs which are actionable by influential decision makers.

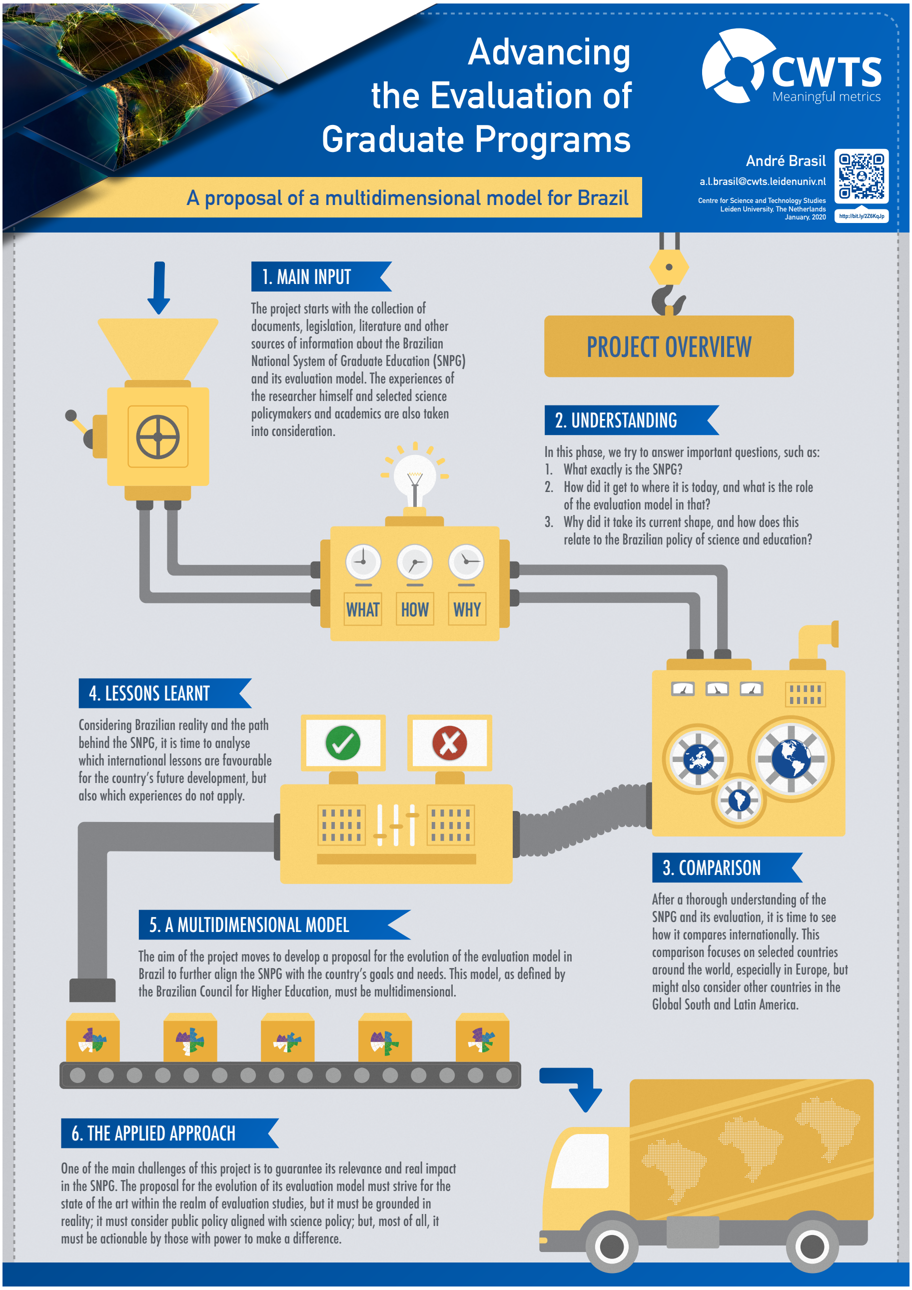


Figure 1.1.: The PhD project that evolved from the initial research proposal, as presented in a seminar organised by CWTS in January 2020.

Based on the proposed stages, the architectural design of the project and the essential concepts used within its structure will be described in more detail in the forthcoming section on the dissertation structure. However, before this discussion, it is necessary to recognise the implications of a crisis that has been quite damaging to Brazilian science and the country's higher education. The events that have occurred over the past few years have undoubtedly had an impact on all aspects of the system, including evaluation practises and the landscape of graduate education.

1.3 Enter a challenging period

As my project evolved to reflect the first year of research and engagement with the international community, a significant crisis engulfed the Brazilian science system. The advent of such a crisis was, unfortunatelly, somewhat anticipated. A tell-tale sign came from CAPES high management at the time Jair Bolsonaro was elected president of Brazil. Since my doctoral project was set to begin during the same week as his inauguration, I was advised not to delay my departure to the Netherlands. The rationale was that the new leadership could potentially suspend my official leave and cancel my research grant, a less feasible disruption if I were already overseas. This advice perhaps reflected a sense of uncertainty about what would happen.

As anticipated, concerns turned into reality when a crisis began to unfold in the initial months of Bolsonaro's administration. This government demonstrated a dismissive stance towards the importance of science and explicitly undermined the value of higher education. To illustrate, soon after the minister of education stated that the university should be reserved for the intellectual elite of the country (Passarelli, 2019), the Ministry drastically cut university budgets by around 30%. At first, the initiative targeted higher education institutions (HEI) that the new government considered bedlam promoters, such as the federal universities of Braslia, Bahia and Fluminense. The justification was that the HEI did not deserve the funding they were receiving, as they were more interested in creating trouble than pursuing academic excellence (Agostini, 2019). Interestingly, the Leiden Ranking lists these three universities in the top 20 most productive Brazilian institutions and in the top 15 if ranked by the percentage of publications among the 10% most cited (CWTS, 2022).

The cuts incited widespread protests from the academic community, particularly since some institutions were specifically targeted in what was considered a political attack (Dias, 2019). Eventually, the budget cuts spread to all federal universities and research institutes. The humanities suffered the worst of these cuts (P. Martins, 2019), inciting further protests nationally and internationally, with academics around the world publishing manifestos on the grave implications of these measures (Furlaneto, 2019).

Brazilian researchers responded to these setbacks by showcasing their ongoing studies on social media, showing that they persisted despite the lack of appreciation of the government (P. Martins, 2019). However, these demonstrations were met with disdain. Bolsonaro even dismissed protesters who took to the streets as "imbeciles", supposedly manipulated by what he considered the left-wing core of Brazilian universities (Dias, 2019). Soon after, a subsequent minister of education publicly stated his opposition to using taxpayer money to fund sociologists, anthropologists, or philosophers. The official position of the government was that only the professions they considered useful, such as medicine or engineering, should be funded (Rezende, 2020).

In line with the attack on higher education in Brazil, Bolsonaro's allies in parliament proposed changes to the Brazilian Constitution so that free public universities would start charging tuition to finance their operations. One of the key arguments of the proposal was that universities from 20 of 28 countries in a 2018 study by the Organisation for Economic Co-operation and Development (OECD) charged tuition (Peternelli Jr, 2019). This mentality of "what works for them should also work for us" reflects a lack of understanding of the higher education system and the socioeconomic reality of the country, as education serves as a critical pathway to upward social mobility in Brazil (Giolo et al., 2020). Though one might advocate for a tuition-based system to tackle Brazil's asymmetries, certain governmental factions seem to be driven by ill-informed or misguided motives. For example, yet another minister of education under Bolsonaro's rule, endorsing the notion that universities only benefit society if restricted to a few, erroneously asserted that Brazil should emulate Germany, where supposedly only a select few attend university.(M. Ribeiro, 2021).

The lack of knowledge also extended to the science system. As part of the movement to end free higher education in the country, Bolsonaro claimed that very few universities conducted research in Brazil and that most of them were private institutions. The academic community quickly counterbalanced this misinformation, pointing out that more than 95% of Brazilian scientific production comes from public universities (Clarivate Analytics, 2019; Moura, 2019). Subsequent discussions revealed that more than 400 Brazilian higher education institutions were engaged in research and that 30 of the most productive universities were public (SBPC and ABC, 2020a).

However, the crisis continued, also affecting research and graduate education. In the case of CAPES, the agency was subjected to budget cuts (Rossini, 2021); high turnover of its high-level management, often followed by protests from the academic community about leadership profiles (SBPC, 2021a; SBPC, 2021b); legal interference that delayed the national evaluation and restricts the transparency of the process and its results (Davidovich and Ribeiro, 2021; Justiça Federal, 2021); and more. Together with the impact of the COVID-19 pandemic, the crisis had immense effects on evaluation, as discussed in more detail in select chapters of this dissertation.

1.4 Dissertation structure and key concepts

In essence, this study's primary goal is not to dissect the crisis that hit science and higher education in recent years. Given the complexity of the situation, dedicated studies will be needed to fully understand it. However, the general lack of knowledge observed on the reality of research and graduate education – as well as of international models often cited as solutions for Brazilian problems – has reinforced the importance of parts of this study related to understanding and comparison. The approach used to organise the dissertation and the research behind it, all in face of the crisis, are presented ahead.

Part I: The what, the why and the how

The economic arena served as the genesis of the path dependence concept, with economic historians trying to understand how less-than-optimal technological solutions seem to linger, not giving way to advancement (Torfing, 2009). An example that reached a paradigmatic status has been described by David (1985), who analysed the development and continued usage of the QWERTY keyboard layout. Introduced in the 19th century, the design prevented jams in mechanical typewriters by placing frequently used letters further apart. The layout prevailed despite the fact that more efficient mechanical alternatives were developed, such as the 1930s DSK keyboard. Likewise, even the advent of computer keyboards has been unable to replace the entrenched QWERTY standard with a more efficient design.

According to David (1985, p. 333), this phenomenon is not a consequence of consumers being "prisoners of custom, conspiracy, or state control. But while they are, as we now say, perfectly free to choose, their behaviour, nevertheless, is held fast in the grip of events long forgotten and shaped by circumstances in which neither they nor their interests figured". While this narrative encompasses all quintessential elements of path dependence, it represents something that maybe should have changed, but it hasn't because of the path behind it.

However, the Common Law legal system, which originated in England in the Middle Ages and spread to other countries, is an example of positive path dependency. As described in Hathaway (2003), this system is based on the previous courts' decisions and can evolve the legal principles over time. The system depends on the path and leads to predictability and coherence of the legal results, with the courts bound by their own decisions and those of the superior courts. This helps to preserve the rule of law, providing a reliable framework for individuals and companies. In this case, path dependency increases social order and the stability of economic activity.

Taking into account the negative and positive dimensions of path dependence, historical sociologists have argued that its understanding has important implications for social research. Path dependence characterises historical sequences in which random events trigger chain events or institutional patterns showing deterministic properties (Mahoney, 2000). Furthermore, the construct of path dependence in the field of policy discourse highlights the extent to which existing policy choices are dominated or influenced by the institutional pathways that result from past decisions. This makes the prospect of public policy reform contingent on institutionalised legacy, shaping our interpretation of problems and objectives, delimiting the spectrum of feasible options, and influencing the cost-benefit analysis of policy changes (Torfing, 2009).

In the spirit of the path dependency approach, the first part of this dissertation aims to establish a contextual framework for understanding the Brazilian National System of Graduate Education (SNPG) and the national evaluation that comes with it. The two chapters included trace their historical evolution to the present state, providing an understanding of why they developed the way they did and how decisions and external influences along the way helped shape what they have become.

The papers included in this first part display the two dominant types of path dependence sequences described by Mahoney (2000). First, chains of causally connected events can be seen, each of them being a direct or indirect reaction to the antecedents. Like in the case of the QWERTY layout, policies around the SNPG and its evaluation were designed to address particular issues, and it is important to understand what they were and if they remain relevant. Second, reactive sequences may lead to the formation and long-term reproduction of a given institutional pattern, also known as self-reinforcing sequences.

Considering the perspective of policy effectiveness and how the Brazilian system can evolve to address current and fluctuating contexts and conditions through inherent adaptability (Bali et al., 2018), Part I is all about understanding the path behind the SNPG and its evaluation, so that policymakers understand the intricacies of the system's development and then can answer the question: "is there a reason for us to keep doing what we are doing?"

Chapter 2 – Building a national system of research and graduate education

This chapter provides a comprehensive review of the Brazilian science system, a unique product of robust public policy initiatives that have bridged research and education for decades. Emphasising the dominant role of the Brazilian National System of Graduate Education (SNPG) within Brazilian science, this study delves into the genesis and evolution of the system, which was shaped by decisive policy decisions over the years. Furthermore, it critically evaluates the justifications behind these decisions, tracing their connections to broader public policy frameworks in Brazil. The study also underscores the top-down structure of the Brazilian science system, marked by influential government agencies that promote change through regulation, assessment, and funding. Using primary sources such as original legislation and policy documents, together with interviews and academic literature, the chapter also delineates the present state of the system, namely its size, organisation, geographical distribution, and demographic characteristics of its faculty and student body. The information obtained is valuable not only for Brazilian academics and policymakers, but also for the international community, offering a unique model for science and high-level education from both its successes and failures.

Chapter 3 – The dynamics of a national evaluation system

This chapter presents a detailed exploration of the Brazilian evaluation system of research and graduate education over the past five decades, detailing its historical evolution, socio-political influences, and current challenges. Central to this investigation is an understanding of the system's transformation from a rudimentary funding distribution instrument to a quality assurance mechanism, and then to its present multifaceted construct, deeply embedded within the national science landscape. The study also underscores the transformative power of the evaluation system in promoting transparency, fairness, and academic quality, which remain critical to allocating funding and conferring legitimacy on graduate programs. However, the chapter acknowledges the need for evaluation to evolve, and it analyses systematic improvements that have been promoted over time, often triggered by a combination of external pressures and internal critique. Looking ahead, the chapter advocates for a multipronged approach for evaluation reform, incorporating self-assessment strategies, stimulating institutional authonomy, and promoting an evaluation model capable of capturing and valuing the diversity in science vital for addressing the country's challenges and asymmetries.

Part II: The Brazilian evaluation system in perspective

Political scientists often use the concept of path dependency to explain why deliberate policy reforms could fail, especially when institutional inertia affects the outcome (Torfing, 2009). However, the concept also explains how reaction sequences and institutional patterns can promote coherent policy strategies and facilitate the evolution of complex systems over time (Mahoney, 2000). An illustration of this dynamic was seen in Part I with the SNPG and its integral evaluation system. Throughout their history, they have been regarded as stable state initiatives rather than fleeting government ones, allowing their continued development to become sources of pride for Brazilian society (Balbachevsky and Schwartzman, 2010; Nader and Davidovich, 2021).

The Brazilian evaluation system owes part of its success and role in advancing the country's science to having developed concurrently with the growth of the

SNPG. This is not a task without challenges, particularly when it comes to blending different aspects of a complex national evaluation with the reality of the research system. However, the interconnectedness of development pathways usually influences how countries deal with their national evaluation systems, leading to a natural tendency to rely on their own established protocols rather than trying to implement models from elsewhere (Ochsner et al., 2020).

However, evaluations have become of substantial importance for national governments, often steering universities and research systems. This influence may, on occasion, encourage them to draw on international models that have demonstrated success (Capano et al., 2016). This is apparent in Brazil, where there are ongoing suggestions to refine the evaluation system based on international experiences. Although countries aiming for similar results with their evaluation procedures can benefit from mutual experiences, it is essential to carefully consider potential lessons, given the uniqueness of science systems and the different paths behind them (Ochsner et al., 2020).

The second part of this dissertation puts the Brazilian science system and its evaluation in an international perspective. First, the invesigation looks at how Brazilian evaluation compares to the long-standing Dutch evaluation system, a frequently cited source of inspiration, to identify any lessons that could be learnt (Barbosa, 2020; Verhine et al., 2019). Subsequently, the scope is expanded to look at the relationship between research and graduate education in Brazil, partly with the objective of confirm its uniqueness and the need for its evaluation to reflect this singularity. Comparative studies like these can enhance our comprehension of challenges across different systems, with focused comparisons particularly attuned to the specifics of individual countries while respecting their historical trajectories. Therefore, any potential differences among systems should be taken into account before trying to replicate strategies that are successful elsewhere.

Chapter 4 – Research evaluation in Brazil and the Netherlands

This chapter presents a comprehensive comparative analysis of research evaluation systems in Brazil and the Netherlands. The investigation explores the differences and similarities between the two systems, focussing on the unique characteristics, challenges, and impacts of their design on the adopted evaluation models. The study underscores the fundamental incompatibility of these systems due to their unique historical, geographical, and policy underpinnings.

On the one hand, the Brazilian system is a high-stakes, performance-based model. Although this system has played a significant part in the advancement of the country's science system, its structure tends to influence researchers' behaviours, inadvertently promoting research homogenisation and potentially limiting innovative exploration. On the other hand, the Dutch system, following the Strategy Evaluation Protocol, adopts a formative, low-stakes approach that emphasises research quality and societal relevance. This model empowers higher education institutions to conduct evaluations and use the results as they see fit, thus fostering diverse and differentiated research. Although both systems have many qualities that can inform and inspire the evaluation strategies of the other, in the context of path dependency, it becomes evident that no "copy–paste" approach can be taken to act on such inspirations.

Chapter 5 – The impact of graduate education on academic publishing

The first part of this dissertation relied on a detailed analysis of legislation, policy documents, and associated literature to gain a complete understanding of the Brazilian National System of Graduate Education (SNPG). Investigating the system from its conception to its present configuration indicates that it was purposefully designed around graduate education, but bibliometric evidence of that reality is often elusive. As such, this chapter combines data from the Web of Science with publication lists gathered from Brazilian Current Research Information Systems (CRIS) to determine the degree of graduate education's influence on academic publishing in Brazil. Moreover, by incorporating additional publication data and international sources on PhD graduates, it was also possible to compare international rates of publications per degree awarded. Although still preliminary, this investigation seeks to explore methods to better understand the differences between the SNPG and other science systems.

Part III: Strenghts and weaknesses in Brazilian evaluation

The Brazilian evaluation system, as discussed in Part I and Part II of this dissertation, is quite complex. It employs a multifaceted methodology, harmonising various evaluative instruments and strategies, and the challenges to bring all that together in a single dissertation are significant. Many of the chapters in this resulting text have the potential to develop into full dissertations themselves, attesting to the scope and complexity of the subjects. However, as mentioned earlier in this Introduction, a holistic diagnostic view of the entire evaluation process has been recognised as a vital gap that justifies this effort for a more ambitious exploration.

Nonetheless, the limitations of individual doctoral research would make it impossible to cover every facet and dimension of the Brazilian evaluation system. Given its complexity, the third part of the dissertation represents the choices made during the research process. That means some representative aspects of the evaluation system under analysis were chosen for a detailed examination, each exemplifying different approaches. One such topic is the classification of scholarly publications through the Qualis system adopted by CAPES. As detailed in Chapter 3, the evaluation of scientific contributions carries significant weight in the final evaluation results, so the system is explored in depth in Chapter 7.

The object of the five chapters included in Part III will be discussed in detail below, but they all draw upon the principle of path dependence. In that sense, the effort behind the previous chapters is replicated after zooming in on each key facet of the evaluation, once again aiming to understand not only their current status, but also their origins, narrative, evolution, missteps, and triumphs. This approach is designed to contextualise reality, allowing the reader to recognise the value of the system in place while also being able to see the need for change and relevant obstacles in tune with current conditions.

Chapter 6 – Rethinking a national classification of research and graduate education

This chapter presents an analysis of the Brazilian classification of research and graduate education. In Brazil, evaluations are organised comparatively within areas, meaning that the performance of graduate programs that integrate each area is assessed in their specific context. The evaluation results derive from the dynamic panorama of the areas, with criteria and indicators adjusted accordingly. To conduct this examination, the dynamics of the design, expansion, and organisation of evaluation areas are explored. Then, an international comparison is conducted, considering other classification systems of research and education, specifically the OECD Fields of Research and Development (FORD) and the UNESCO International Standard Classification of Education (ISCED). Finally, the study explores whether a review of the classification may be needed and introduces scientometric approaches that could support the areas in their potential reorganisation.

Chapter 7 – Between Bibliometrics and Peer Review

This chapter focusses on the evolution and challenges of the Brazilian Qualis system, which employs journals as a means of assessing the quality of scholarly work spanning diverse disciplines. The study starts with an analytical review of the historical trajectory of the system, highlighting its implications as an integral part of the evaluation process of research and graduate education in Brazil. The Qualis system is fundamentally hybrid, balancing quantitative and qualitative methodologies by using a suite of indicators as a reference, but extending the analysis through a peer-review mechanism. Despite not being free of imperfections, Qualis has shown significant progress in its nearly 25 years of existence and continues to evolve. While some advocate for its dissolution in favour of international metrics, the chapter recognises the system's vital role in fostering a balanced and fair academic environment in Brazil and warns against overreliance on metrics, which may not accurately reflect the challenges faced by Brazilian researchers or the significance of locally-relevant research. Instead, the chapter proposes enhancing the Qualis system towards a more inclusive, equitable, and comprehensive evaluation framework.

Chapter 8 – Beyond the Web of Science

The Brazilian evaluation system of research and graduate education is under review. For scientific publishing, policymakers and evaluators consider a more standardised assessment, trading some current qualitative aspects for expanded use of journal-based indicators from international databases such as the Web of Science. This chapter aims to provide evidence for a better-informed discussion, analysing the complete data set of Brazilian articles published from 2013-2018 regarding: i) coverage by the WoS and regionally relevant databases (e.g., Latindex, SciELO); ii) incidence of the local language in the country's publications, and the impact it has on coverage by regional and international databases; iii) disciplinary variations and disparities in thematic coverage across languages and databases. The results show that half of the Brazilian article output is not found in the WoS, and the normalised distribution of indexed publications across disciplines is hugely unbalanced. Publications not in WoS are predominantly in Portuguese, with a significant share indexed by regional databases, often addressing topics not covered in WoS. The main conclusion is that Brazilian science goes beyond WoS, and evaluators should strive for a sound and comprehensive assessment to capture its complexity, rather than trading it for restraining, short-sighted simplicity.

Chapter 9 – The unseen costs of article processing charges

This chapter takes a slight detour from the core dimensions of Brazilian evaluation, given that elements of open science are projected to be integrated only in upcoming evaluation cycles. However, numerous factors related to openness are already relevant following a specific type of scholarly output identified in Chapter 8: scientific production not indexed by international databases, often in the local language, addressing issues of local relevance. As discussed, Qualis has the potential to capture quality in outputs indexed by regional databases such as SciELO, RedALyC, and Latindex, which are important platforms that feature mostly diamond open access journals, where papers are free to read without the authors having to pay for Article Processing Charges (APCs). Therefore, this chapter initially explores the relationship between APC investment and impact at the national level, uncovering that Brazil appears to invest little and achieve a below average impact. Building on this revelation, the study introduces an economic context, advocating the use of indices such as the Purchasing Power Parity (PPP) exchange rate for international comparisons. The review also scrutinises the vast array of publications deemed by Qualis to be valuable, but that often get excluded from consideration when using major databases for comparison. As the results of this chapter show, countries like Brazil are not able to make the same level of financial investment seen in the Global North, thus developing alternative publication paths that are more coherent with their economic reality, but also capable of resulting in high-quality work, even in local language. Evaluation should be able to capture and value such outputs.

Chapter 10 – A national evaluation push towards increased societal impact

The final chapter of this part of the dissertation explores one of the advances proposed to improve Brazilian evaluation that had the opportunity to be partially implemented in the most recent cycle. For decades, evaluations have tried to improve the impact of Brazilian science, often deemed excessively academic, far from society in its ivory tower. Initiatives such as the establishment of professional graduate programs, addressed in Chapter 2, have been rolled out over the years in an attempt to rectify the situation. Other initiatives focused on the diversification of valued research outputs, leading to the creation of distinct Qualis for books, events, and artistic production, as mentioned in Chapter 7. A recent initiative also saw the advent of a Qualis for technical and technological production. This chapter examines the development and preliminary implementation of this new type of Qualis. The primary objective is not only to identify how products such as patents, cultivars, and policy briefs can be valued, but also to show how evaluation has the potential to steer a country's science system to generate more diverse and impactful research outputs.

Part IV: Towards a multidimensional model

According to Barata (2020), the evaluation of the Brazilian National System of Graduate Education is at a crossroads in terms of the next steps of its natural evolutionary process. Over the years, the system has managed to improve significantly, as evidenced in the first parts of this dissertation. However, the exponential growth of the SNPG poses challenges to the capacity of the existing evaluation model, and further evolution may only be possible with an extensive redesign of the process.

A recurring issue with the ongoing evaluation is the growing reliance on quantitative indicators, which represent a significant risk to the evolution of the SNPG. An overly quantitative evaluation tends to foster a homogeneous science system that stifles creativity and innovation, as graduate programs are led to focus predominantly on the pursuit of metrics. This trend is, in part, the result of the high-stakes evaluation conducted in Brazil, as discussed in Chapter 4. The implication is that an evaluation which impacts from accreditation to funding ultimately tends to rely on quantitative measures as a more defensible means of justification for its outcomes.

From the evident fragilities of a quantitative-centric approach, the conversation must pivot towards alternatives that can accommodate the sprawling complexity and diversity of Brazil's academic landscape. The critical shift is to approach evaluation from a multidimensional perspective, so that the inherent complexities of the system can be more effectively managed. This recommendation seems to be almost a consensus in multiple analyses of the Brazilian evaluation, as seen in Barbosa (2020), PNPG Committee (2020), and Verhine et al. (2019).

In the final part of this dissertation, the goal is to bring together the various studies conducted into actionable advice toward a multidimensional model. The opening chapter of Part IV examines an attempt made by CAPES to implement multidimensionality through the inspiration of uMultirank.

Chapter 11 – Multidimensionality through self-evaluation

Although many academics question the value of university rankings, some experts have recognised U-Multirank as one of the few that gets closer to meeting community expectations of fairness and responsibility (Gadd et al., 2021; Hazelkorn and Gibson, 2017). Despite this, when CAPES leadership explored the idea of adopting this model as a replacement for the current evaluation system, many within the agency were concerned. This initiative was in line with the concerns already mentioned about the search for primarily quantitative or international evaluation solutions disconnected from the path behind the Brazilian system. With growing concerns on the value of rankings, adopting one to replace five decades of evolution in the country's evaluation is not adequate. This chapter explores the abandoned initiative to adopt the uMultirank model as a multidimensional strategy in Brazil. Highlighting the dangers of seeking an easy solution to evaluation problems, the research suggests self-assessment as an alternative strategy for multidimensionality and alleviating the burden on Brazilian evaluation. As discussed throughout this dissertation, the current system is no longer capable of dealing with the size of the SNPG unless it can count on HEI as partners, exercising their much needed institutional autonomy.

Chapter 12 – Conclusions & Reflection

A concluding chapter then sets out to compile the learnings from this academic journey, offering recommendations for advancement that match the trajectory of the SNPG and its evaluation. To tackle this task, the chapter begins with an analysis conducted by the CAPES evaluation directorate on how much the Brazilian system was aligned with the principles of the Leiden manifesto. This analysis was part of the agency's efforts to support the discussions about the future of evaluation held during the 2015 seminars with graduate program directors, as outlined at the beginning of this Introduction. In the conclusion, the original analysis will be presented and each of the principles of the manifesto will be revisited in light of the research conducted for this dissertation.

Appendices

As depicted in Figure 1.1, this doctoral research project has an applied approach. This facet is crucial in that it seeks to ensure its relevance and impact on evaluation, facilitating the transmission of research findings to policymakers involved in the process. It is within this pragmatic context that two appendices are included, providing an inventory of deliverables intrinsically tied to the essence of this dissertation.

Appendix A – Expanding GeoCapes usefulness for research

Throughout the trajectory of this doctoral project, significant effort was invested in understanding Brazilian databases on research and graduate education. In many instances, lengthy and intricate algorithms had to be created for data cleansing and to enrich data sets like the one offered by CAPES' Georeferenced Information System (GeoCapes). This system offers around 25 years of aggregated data on graduate programs, faculty, student body, funding, and more. Rather than sharing the algorithms designed to deal with GeoCapes limitations, it seemed beneficial to the research community to devise an R package instead, integrating algorithms as functions for producing data output and visualisations. The packa is bilingual, allowing Brazilian researchers' to use data in their native language and in English, thus aiding in international publications. Additionally, this package targets the international research community, offering English access to data typically unaccessible to non-Portuguese speakers. Consequently, a paper was produced to support the package, also in English and Portuguese versions, the first being included in the Appendix A.

Appendix B – Additional Resources

Several chapters in this dissertation include recommendations for the evolution of evaluation instruments and for the Brazilian system as a whole. However, this Introduction has noted that the consumption of such results may not always be straightforward. Factors such as time availability, the need for specialised knowledge to comprehend certain outputs or linguistic barriers can present obstacles. To this end, a set of policy briefs has been produced in Portuguese, tailored for submission to CAPES to support the understanding and implementation of this dissertation's findings. In addition, digital resources are included, exemplified by a dedicated website comprising a plethora of interactive dashboards with information about Brazilian science that can serve as evidence for Brazilian evaluation policy. Another noteworthy output of this research comes from the extensive literature review that was conducted, which could not be fully incorporated into the chapters that make up the dissertation. Consequently, an interactive dashboard was developed to allow the exploration of such references in a user-friendly tool that can support researchers interested in the topic of graduate evaluation, in Brazil and beyond.

1.5 Methods and data

Given the detailed nature of the research object previously outlined, it becomes clear why this study necessitates the adoption of a multi- and interdisciplinary approach. The methodological framework, in turn, must incorporate both qualitative and quantitative methods. Only through a mixed-method approach can we gain an encompassing insight into the complex dynamics of research practises and the broader scientific landscape in Brazil.

In this vein, the dissertation is anchored in two primary analytical approaches. First, documents such as policies, laws, evaluation tools, and other procedural materials are analysed, often with a focus on constructive criticism. The primary aim here is to allow for a comprehensive examination of the SNPG, its evaluation, and its primary instruments, originating from an understanding of its developmental process, its motivations, and the political and administrative decisions that shaped the research object.

Secondly, a variety of empirical data sources are consulted, especially those concerning publications and graduate education information drawn from Brazilian Science & Technology databases as well as from international providers, bibliographic and otherwise. Echoing Carl Sagan's sentiment that "you have to know the past to understand the present," the data analysis not only yields evidence supporting the understanding derived from the document review but also sheds light on the current state of the research object. With a grasp of the Brazilian evaluation framework today, through the synthesis of both analytical approaches – informed by current regulations and empirical evidence – we can project crucial perspectives on the potential trajectories of the country's evaluation. By merging a reflexive stance with continued support from the literature, this study then introduced propositions for some necessary adjustments for a better future.

The reflexivity cited is consistently threaded throughout the dissertation, illuminating the interactions between the insights I acquired as a participant in the Brazilian evaluation system and the empirical data. Therefore, this methodology incorporates a participant-observer perspective, recognising the inherent value of my experiences and observations. These are deemed valuable additions to the understanding of the dataset, enhancing the depth and context of the analyses. Yet, the looming threat of potential biases from such an approach was constantly acknowledged. I was consistently diligent in ensuring that my interpretations remained tied to both experiential and empirical evidence. Maintaining a stark demarcation between my existing professional knowledge and what I learnt during this doctoral journey proved to be a nuanced endeavour at times. Reflecting on the results, I acknowledge that I haven't always achieved the perfect balance. However, in such instances, I consistently leaned towards giving greater weight to the evidence, ensuring the scientific rigour expected of this work was maintained.

Regarding the literature and other documents mentioned, a total of 1462 research items were consulted in the construction of this dissertation. Among them are 635 journal articles, 154 books, 127 book chapters, and 117 pieces of legislation – including laws, decrees, and ordinances – along with numerous other documents, such as official reports, policy briefs, and informative bulletins. All this material was collected with great care given to the quality of metadata for each document, and they were organised using the ReadCube Papers application.

Clearly, it was not feasible to cite all the academic literature and other documents consulted throughout the dissertation. Thus, a selection of the most pertinent documents had to be made. This process largely prioritised primary sources whenever available, especially when it comes to legislation and official documents. At the end of each chapter, a dedicated list of references is provided, while the end of the book features a complete reference list of cited works, as well as an author index, listing the pages were they are cited. Appendix B provides details on an interactive online dashboard created from the complete set of references, cited or not. Users can explore the list of works with access to the links available for each record. One can also select the various dissertation chapters to filter the cited works. The dashboard also allows multiple filtering options by author, type of publication, and language. It is noteworthy that about a third of the works used for this research are in Portuguese. One of the objectives of this study was to bring this extensive collection of documents to an international arena, at a level of detail which I believe to be unprecedented.

Appendix B also includes details on a dedicated dashboard for the collected legislation. The most relevant documents that have influenced the construction of Brazilian graduate education and its evaluation are presented on a timeline where each reference contains a link to the full official text.

Two other types of references also make up the research conducted for this study. One of them consists of news articles, interviews, public letters, manifestos, and similar documents that supported all the accounts presented on the political crisis faced in Brazil, as well as the economic and managerial challenges that affected Science & Technology in the country. Once again, the choice was always to use primary sources, including recorded public statements from various actors mentioned throughout this work.

Among the cited interviews are recent records to support the mentioned accounts, but historical documents are also included. For example, the text "CAPES 60 years: Six decades of evolution in graduate education" incorporates transcriptions of interviews with former CAPES presidents, directors and coordinators, as well as evaluators, graduate education experts, etc. Interviews, testimonies, and online seminars also served as sources for capturing the opinions of relevant actors, for example, events organised by the Brazilian Society for the Advancement of Science and Brazilian Academy of Sciences (SBPC and ABC, 2020a), and lecture transcriptions such as one from a seminal speech by Schwartzman (1989). Speaking of opinions, some articles in this category were also used, written by CAPES leaders and other important actors in the construction of Brazilian graduate education. Among them, it is worth mentioning works by Barata (2020), Castro and Soares (1983), and Sucupira (1980).

The last type of additional reference consists of the databases used. The number of sources is relatively extensive, but the most important ones are listed below.

i) **Data from Brazilian graduate education –** Most of these data were obtained from CAPES itself. Information has been collected by the agency for several decades, using dedicated platforms whose design and evolution processes are described in detail in Appendix A. The main system used by CAPES is the Sucupira Plataform (CAPES, 2021c), where data on graduate programs, research projects, faculty, student body, intellectual production, and much more are available. Much of the platform's data are open for online consultation, but obtaining large volumes of data is only possible with administrative access. Although I have the necessary access through my affiliation with the agency, whenever possible, preference was given to using data extracted from the CAPES Open Data platform (CAPES, 2021a), where it is possible to download the microdata collected via the Sucupira Plataform after processing, cleaning, and validation by the agency. With this choice, the reproducibility of the studies presented here is facilitated. Finally, data from the GeoCapes platform (CAPES, 2021e), where partially aggregated official data are presented, were also used at times.

ii) **Additional Data from the Brazilian Scenario –** A variety of other national data sources were used to supplement the information obtained from CAPES. Among them, one can mention the educational census conducted by INEP, whose microdata are available on the institution's platform. Geographic data were sourced from Brazilian Institute of Geography and Statistics (IBGE), and economic data were collected from the Atlas of Human Development.

iii) **International Data on Science and Education –** The Brazilian data was supplemented with a range of international data sources, including the Organisation for Economic Co-operation and Development (OECD), the European Comission's Eurostat, National Science Foundation (NSF), ministries of education, science, and technology (The Netherlands, Japan, China), among others. Each of these platforms features a dedicated Application Programming Interface (API) or query systems that allow for the export of data in comma-separated values (CSV) format. This kind of information, for instance, enabled a comparison of publication rates per doctoral graduate in Brazil and various other countries, as presented in Chapter 5.

iv) **International Bibliographic Databases –** The most extensively used bibliographic database in this dissertation was the Web of Science Core Collection, particularly the in-house version available on the CWTS servers (Clarivate Analytics, 2022). Although this database has coverage issues regarding Brazilian production, as discussed in Chapter 8, there are still some advantages to its use, especially due to the quality of the metadata and the refined affiliation mapping developed over the years. To a lesser extent, consultations were made to the Scopus and Dimensions databases, and data from the recent OpenAlex integrate the information portal about Brazilian science — BrScience — presented in Appendix B. All these databases are available in a relational format on the CWTS server.

v) **Regional Bibliographic Databases –** Among the supplementary databases used are Scientific Electronic Library Online (SciELO), Network of Scientific Journals of Latin America and the Caribbean (RedALyC), and Regional Cooperative Online Information System for Scholarly Journals from Latin America, the Caribbean, Spain and Portugal (Latindex). In most cases, information had to be sourced directly from the websites or the Application Programming Interface (API) provided by these bases, offering a broader overview of journals used for publication not only in Brazil but throughout Latin America.

vi) **Open Science Databases –** To enrich many of the studies conducted, databases such as the Directory of Open Access Journals (DOAJ) and Unpaywall were also used, enabling, for instance, the analysis of the percentage of open access publications and associated Article Processing Charges (APCs) presented in Chapter 9. The DOAJ database was sourced directly from the organization's website, while Unpaywall was available on the CWTS server.

Regarding analytical methods, Atlas.ti was the primary tool for coding interviews and document analysis. For example, Chapter 10 includes a mapping of valued technical and technological products, as well as the main criteria and indicators adopted for the evaluation of these products in each of the 49 distinct evaluation areas in Brazil. About 100 documents needed to be coded to survey the reported information, and Atlas.ti was used to execute the strategies for document analysis.

To conduct the scientometric analyses here included, together with other quantitative studies, SQL was the language used for database queries and the calculation of various indicators presented in the different chapters. The integration of data not available from relational databases, and further data processing and analysis, was conducted using R Studio. Several algorithms created for this research are available in the project's GitHub repository. Among them, for example, are algorithms capable of integrating multiple data files obtained from the CAPES Open Data platform. Many other algorithms, developed especially to deal with the different databases of the GeoCapes system, were integrated into an R package freely available on Comprehensive R Archive Network (CRAN).

Finally, the various charts and figures presented throughout the dissertation were partly developed using R and Flourish Studios. However, most of the visualizations were created using Tableau Desktop, allowing for the publication of interactive online dashboards that are referenced throughout this work.

In conclusion, by combining the adoption of a multi- and interdisciplinary approach with mixed research methods, this study seeks to integrate theoretical principles, academic literature, and empirical observation. This combination aims to not only understand the context of the evaluation of research and graduate education in Brazil but also to propose general principles to guide future actions for its advancement.

References

Agostini, R. (2019). MEC cortará verba de universidade por 'balbúrdia' e já enquadra UnB, UFF e UFBA. *Estadão*. Retrieved from https://bit.ly/3pYA0KW

Andrews, L. (2017). How can we demonstrate the public value of evidence-based policy making when government ministers declare that the people 'have had enough of experts'? *Palgrave Communications*, *3*(1), 11. https://doi.org/10.1057/s41599-017-0013-4

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

Bali, A. S., Capano, G., & Ramesh, M. (2018). Anticipating and designing for policy effectiveness. *Policy and Society*, *38*(1), 1–13. https://doi.org/10.1080/14494035.2019.1579502

Barata, R. B. (2020). Programas de pós-graduação profissionais: Por que precisamos deles? *International Journal of Business & Marketing*, *5*(2), 30–34.

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

Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy and Systems*, *15*(1), 35. https://doi.org/10.1186/s12961-017-0192-x

Capano, G., Regini, M., & Turri, M. (2016). *Changing Governance in Universities*. London: Palgrave Macmillan.

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/

Castro, C. d. M., & Soares, G. A. D. (1983). Avaliando as avaliações da Capes. *Revista de Administração de Empresas*, *23*(3), 63–73. https://doi.org/10.1590/s0034-75901983000300007

Clarivate Analytics. (2019). *Research in Brazil: Funding excellence*. Retrieved from https://bit.ly/3gdf5vY

Clarivate Analytics. (2022). Web of Science Core Collection. CWTS in-house version. Retrieved from https://bit.ly/3rHmCJT

Colglazier, B. (2016). Encourage governments to heed scientific advice. *Nature*, *537*(7622), 587–587. https://doi.org/10.1038/537587a

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

CWTS. (2022). Leiden Ranking. Retrieved May 24, 2023, from https://www.leidenranking.com/

David, P. A. (1985). Clio and the Economics of QWERTY. *The American Economic Review*, *74*(2), 332–337.

Davidovich, L., & Ribeiro, R. J. (2021). Manifestação da ABC e SBPC contra suspensão da Avaliação Quadrienal 2021. Retrieved March 10, 2022, from https://bit.ly/3q12v8j

Dias, M. (2019). Bolsonaro diz que manifestantes contra cortes na educação são idiotas úteis e massa de manobra. *Folha de São Paulo*. Retrieved from https://bit.ly/4386EZ0

Donnelly, C. A., Boyd, I., Campbell, P., Craig, C., Vallance, P., Walport, M., . . . Wormald, C. (2018). Four principles to make evidence synthesis more useful for policy. *Nature*, *558*(7710), 361–364. https://doi.org/10.1038/d41586-018-05414-4

Farmer, R. (2010). How to Influence Government Policy with Your Research: Tips from Practicing Political Scientists in Government. *PS: Political Science & Politics*, *43*(4), 717–719. https://doi.org/10.1017/s1049096510001368

Furlaneto, A. (2019). Acadêmicos de universidades em todo o mundo assinam manifesto contra cortes de verbas para Ciências Humanas no Brasil. *O Globo*. Retrieved from http://glo.bo/42Y5QGE

Gadd, E., Holmes, R., & Shearer, J. (2021). Developing a Method for Evaluating Global University Rankings. *Scholarly Assessment Reports*, *3*(1), 1–19. https://doi.org/10.29024/sar.31

Gaieck, W., Lawrence, J., Montchal, M., Pandori, W., & Valdez-Ward, E. (2020). Opinion: Science policy for scientists: A simple task for great effect. *Proceedings of the National Academy of Sciences*, *117*(35), 20977–20981. https://doi.org/10.1073/pnas.2012824117

Giolo, J., Leher, R., & Sguissardi, V. (2020). *Future-se: Ataque à autonomia das instituições federais de educação superior e sua sujeição ao mercado*. São Carlos, SP: Diagrama Editorial.

Gluckman, P. (2014). Policy: The art of science advice to government. *Nature*, *507*(7491), 163–165. https://doi.org/10.1038/507163a

Hathaway, O. A. (2003). Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.384595

Hazelkorn, E., & Gibson, A. (2017). Global science, national research, and the question of university rankings. *Palgrave Communications*, *3*(1), 21. https://doi.org/10.1057/s41599-017-0011-6

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

Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, *29*(4), 507–548. https://doi.org/10.1023/a:1007113830879

Marques, J. (2021). Crise na Capes ameaça qualidade da pesquisa e divisão de verbas para pós-graduação. *Estadão*. Retrieved from https://bit.ly/3o12nb2

Martins, P. (2019). Cortes nas universidade federais são 'colapso anunciado com calma e precisão'. Retrieved May 19, 2023, from https://bit.ly/3MEbeZF

Montana, J., & Wilsdon, J. (2021). Analysts, advocates and applicators: three discourse coalitions of UK evidence and policy. *Evidence & Policy*, *20*(20), 1–14. https://doi.org/10.1332/174426421x16112601473449

Moura, M. (2019). Universidades públicas realizam mais de 95% da ciência no Brasil. Retrieved November 26, 2022, from https://bit.ly/3gEVbxw

Nader, H., & Davidovich, L. (2021). Crise na Capes e a batalha pela Ciência brasileira. *Piauí*. Retrieved May 13, 2023, from https://bit.ly/44ZdMsw

Ochsner, M., Kulczycki, E., Gedutis, A., & Peruginelli, G. (2020). National Research Evaluation Systems. In R. Ball (Ed.), *Handbook bibliometrics* (pp. 99–106). https://doi.org/10.1515/9783110646610-011

Passarelli, H. (2019). 'Ideia de universidade para todos não existe', diz ministro da Educação. *Valor Econômico*. Retrieved from http://glo.bo/3IBD6uY

Peternelli Jr, R. S. (2019). *PEC 206/2019. Dispõe sobre a cobrança de mensalidade pelas universidades públicas*. Câmara dos Deputados. Brasília, DF. Retrieved May 19, 2023, from https://bit.ly/3BEQ6Mr

Rezende, C. (2020). Weintraub: 'Não quero sociólogo, antropólogo e filósofo com meu dinheiro'. Retrieved May 19, 2023, from https://bit.ly/3WhXmHw

Ribeiro, M. (2021). Ministro da Educação, Milton Ribeiro, é o convidado do Sem Censura. Retrieved May 25, 2023, from https://bit.ly/4391lZU

Ribeiro, R. J. (2022b). Depoimento do presidente da SBPC sobre a Avaliação Quadrienal da CAPES. Rio de Janeiro. Retrieved from https://bit.ly/3tp0FyM

Rossini, M. C. (2021). Entenda o corte de R$ 600 milhões no orçamento da ciência brasileira. *Superinteressante*. Retrieved March 12, 2022, from https://bit.ly/3t1lsJH

SBPC. (2021a). Carta aberta da SBPC sobre a CAPES. Retrieved March 10, 2022, from https://bit.ly/3CnkQ4o

SBPC. (2021b). Entidades se manifestam em defesa da Capes. Retrieved May 14, 2022, from https://bit.ly/38xTKN8

SBPC, & ABC. (2020a). A história da CAPES na voz dos seus ex-presidentes [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

Schwartzman, S. (1989). A ciência no período de pós-guerra. presentation. Brasília, DF.

Sucupira, N. (1980). Antecedentes e primórdios da pós-graduação. *Forum Educacional*, *4*(4), 3–18.

Torfing, J. (2009). Rethinking path dependence in public policy research. *Critical Policy Studies*, *3*(1), 70–83. https://doi.org/10.1080/19460170903158149

Verhine, R. E., & Dantas, L. (2009). Reflexões sobre o sistema de avaliação da CAPES a partir do V Plano Nacional de Pós-graduação. *Revista de Educação Pública*, *18*(37), 295–310.

Verhine, R. E., Lima, A. C. S., Carvalho, D. B. B. d., Leite, D., Bertolin, J. C. G., Dantas, L. V., . . . Rosa, D. L. (2019). *Autoavaliação de Programas de Pós-Graduação*. Brasília, DF: CAPES. Retrieved from https://bit.ly/3jTSRS8

Part I

The what, the why and the how

2 Building a national system of research and graduate education

> *While higher education in Brazil is plagued by many known problems, graduate education is a token of national pride recognised as such by the entire Brazilian society.*
>
> — **Balbachevsky & Schwartzman**

The ancient Greek philosopher Heraclitus once said: "there is nothing permanent except change". The idea of transformation has been a recurrent object of interest for thousands of years, as the cyclical process of birth, growth, decline, and decay has not only determined the course of human life, but has also been a driving question for scholars everywhere (London, 1996). As described by de Biasi (2019), change is ubiquitous, undeniable, inevitable, and irresistible.

In the modern world, resisting change has often been a fast path to failure. Numerous examples exist in the business context, but those of Blockbuster and Kodak have become somewhat emblematic. Both companies were leaders in their domains, but failed to adapt to changing times. Blockbuster refused a partnership with Netflix in 2000, as it considered the new company business model to be profitless; Kodak would not risk its lucrative photography film business and, despite having invented the first megapixel-camera back in 1986,

This chapter was published as: 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. DOI: 10.33871/nupem.2020.12.27.222-253.

fell behind in the digital race. Both companies filed for bankruptcy in the last decade mainly because they could not change, as their organisational culture had become too rigid (Gershon, 2013).

The evolutionary nature necessary to remain relevant also applies to scientists, research centres, and whole science systems (Fealing et al., 2011). The COVID-19 crisis has confirmed this need for continuous change. Although the call for a fast response to fight the pandemic has reaffirmed the importance of science & technology worldwide, scientists had to redefine concepts of collaboration, scientific publishing, influence over policy, etc. (Serafim and Dias, 2020). As a consequence, notions of change that are well-established in business are increasingly present in the scientific environment (de Biasi, 2019).

Despite being so evident at this very moment, the understanding of the dynamics of science and innovation activities, as well as the way they change over time, is not something new. For decades, they have been studied by core social science disciplines such as Economics, Philosophy, and Sociology. On top of that, there is an emerging interdisciplinary field that is trying to establish itself as a recognised academic subject: Science of Science Policy (SoSP). This research area already contributes to the frameworks that provide an understanding of institutional structures that promote or impede scientific progress. Among its core concepts is the idea that science is not divorced from politics, as demands from the political system have been, for decades, the main drivers of science investment and advancement (Fealing et al., 2011; Marburger III, 2011).

Serafim and Dias (2020) argue that alliances between the State and scientific community can positively impact on the advancement of science. SoSP builds on that idea, claiming that governments play a crucial role in this progress, despite often lacking the necessary evidence for the proper design and implementation of S&T policies. As a result, policymakers usually cannot predict how best to manage science investments, much less design or reform science systems within their countries (Marburger III, 2011; Sapolsky and Taylor, 2011).

Brazil is such a country. As this chapter will show, the nation has an extensive research and graduate education system whose design has been the result of firm public policy initiatives over several decades. As the core of Brazilian science, this system has become very peculiar, as it integrates research and education in such an indivisible manner that it became determined even by the country's federal constitution (BRASIL, 1988; M. H. d. M. Castro, 2015). The successful result is knowably expressed by Balbachevsky and Schwartzman (2010, p.87), who state that "while higher education in Brazil is plagued by many known problems, graduate education is a token of national pride recognised as such by the entire Brazilian society".

The aim of this chapter is mainly to provide a contextual understanding of the system mentioned above. The task was conducted from the study of primary sources – including original legislation and policy documents – as well as the analysis of archival interviews and literature exploring the system's history. The intended result is to reach a proper knowledge of: i) what is the Brazilian National System of Graduate Education (SNPG) and how representative it is in the face of the science conducted in the country; ii) how did this system evolve as a result of policy decisions over several decades; iii) why were such decisions made, and how they relate to a broader picture of public policy in Brazil.

After tracing back the history and evolution of the SNPG, this chapter will also provide a panorama of what it has become: how large is the system; how is it organised and distributed across the country; what is the demography of the faculty and student body, etc. Hopefully, understanding the SNPG today, as well as the complexities behind its development, may assist policymakers in their continuous efforts to advance the science system in Brazil.

Despite the main goal of bringing scientific evidence to the policy-making environment, this chapter also aspires to provide a comprehensive understanding of the Brazilian science system to the international academic community. Few studies on the SNPG have been published in English, and none of them is as complete as this one proposes to be. As will be clear, the public policy approach to science and high-level education in Brazil can serve as an inspiration to every country, both from its successes and failures.

2.1 The late start of higher education in Brazil

Higher education was a late phenomenon in Brazil, even compared to other Latin American countries. There were many reasons for this, but two of them were of particular importance: first, the Portuguese colonisation policy presented absolute resistance to this kind of initiative; second, the Brazilian elites would rather pursue their education abroad, mainly in European higher education institution (HEI). Such determinants led not only to inaction, but also to the active establishment of obstacles to those interested in bringing higher education to the colony. One of the first known records of this comes from the 16th century, when the Portuguese Crown denied authorisation for the Jesuits to open the first HEI in Brazil (Fávero, 2006; C. B. Martins, 2018; Moacyr, 1937).

Over the next centuries, other attempts were made without success. An example comes from the Minas Conspiracy, a separatist movement at the end of the 1700s that had the first Brazilian university as part of its independence plans. The conspiracy was foiled, and the hope for higher education was put on hold until the Portuguese Court was transferred to Brazil, escaping the French invasion in 1807. Within a year of their arrival, two medical schools were created: one in the state of Bahia and another in Rio de Janeiro, where a polytechnic school soon followed. (Cunha, 2007; Fávero, 2006; Rothen, 2018).

Within a couple of decades, a few more courses and institutions were created, but all of them with a very applied nature. They were mainly designed to qualify professionals to work for the government or to provide technical training to a select group of the ascending elite. After the country's independence, in 1822, the new Brazilian Empire started to make room for a more academic mentality, and by 1828 two legal courses were installed: one at the Saint Francis Convent in the city of São Paulo and the other at the Saint Benedict Monastery in Olinda. Although these courses had a significant influence on the development of political mentality in the country, the promise of higher education in Brazil would not materialise for several decades (Fávero, 2006; Rothen, 2018).

At the end of the Brazilian imperial times, in 1889, the country had just six HEI and all attempts to create universities were unsuccessful. With the advent of the Republic, although the number of higher education institutions increased to 24, these were mostly professional schools, devoid of scientific research. The dream of a university continued to slip away (Fávero, 2006; C. B. Martins, 2018).

2.1.1 The first universities

Legislation from the newly proclaimed Brazilian Republic would start to pave the way for a university system in the country. The first step came from the 1891 Constitution that, despite maintaining higher education as an attribution of the Federal Government, stated it was no longer its sole jurisdiction (BRASIL, 1891). Also, Law 173/1893 and other legal provisions created the possibility of private ownership of higher education and scientific establishments, including those linked to associations (BRASIL, 1893; Rothen, 2019).

Once the Federal Government lost its exclusive power over matters of science and higher education, the first research institutes were founded (C. B. Martins, 2018). At the time, there was a great need to better understand Brazilian biodiversity, tropical diseases, and other topics relevant to the country (Schwartzman, 1989). This led to the creation of institutions such as the Emílio Goeldi Museum (1885), the Agronomic Institute of Campinas (1887), the Butantan Institute (1899), and the Oswaldo Cruz Foundation – Fiocruz (originally founded as the Manguinhos Institute in 1900). These and other research centres were created isolated from educational institutions and would play an essential role in the building of the country's scientific community (C. B. Martins, 2018).

Parallel to the creation of research centres, several higher education institutions were founded, including the first universities in the country: one in Manaus (1909), another in São Paulo (1911) and the third in Paraná (1912) (Cunha, 2007). As a direct reaction to this movement, the secondary and higher education systems were reorganised by Decree 11.530/1915, which determined the Government would move to establish the first federal university in Brazil "when possible" (BRASIL, 1915). This would take place in 1920, as another decree instituted the University of Rio de Janeiro (BRASIL, 1920).

Despite having a federal university as one of the positive effects of Decree 11.530, the educational reform would have a much more substantial impact on the Brazilian science system. The reason was that it established the university as the home of research and education; and these should be integrated (Rothen, 2019). The drawback is that the approach might have been taken in a very normative and absolute way, and the resolution would become an obstacle for the founding of new research institutions outside of universities.

2.1.2 Early concepts for a graduate system

While Brazil mobilised to build its first universities, prominent organisations emerged in the country. Among them was the Brazilian Academy of Sciences (founded originally in 1916 as the Brazilian Society of Science); and the Brazilian Education Association (ABE) in 1924 (Fávero, 2006). These institutions would play a fundamental role in the design of the country's universities and in how they integrate teaching and research activities (C. B. Martins, 2018). For example, in the document "The problem of the Brazilian university", ABE (1929) stated "the university, to be worthy of that name, should become a focus of culture, of disseminating acquired science and creating new science".

The viewpoint of ABE represents the dominant belief in Brazil at that time: the defence of the university's role in the development of scientific research, in addition to the training of professionals (Fávero, 2006). This belief becomes reality with yet another reform of the higher education system in 1931 (Sucupira, 1980). Francisco Campos, the Minister of Health and Education at that time, led the initiative that included the creation of the National Education Council (Decree 19.850, 1931a), the promulgation of the Statute of Brazilian Universities (Decree 19.851, 1931b), and the creation of the University of Rio de Janeiro (Decree 19.852, 1931c).

Although there had been records of institutions that had already provided doctorate degrees through simple thesis defence (Sucupira, 1980), the comprehensive university project that was launched in 1931 marked the formal beginnings of graduate education in the country (Balbachevsky and Schwartzman, 2010). The project not only created the first regular doctoral courses in the areas of law, exact, and natural sciences (BRASIL, 1931c), but defined the idea of a university that "transcends the exclusive purpose of teaching, involving concerns of pure science and disinterested culture" (BRASIL, 1931a). Even though inspiring, this core belief would not be strong enough for science to find its true space in the university at that time (Schwartzman, 1991).

2.2 The first steps of the SNPG

In the years following the university reform of 1931, the advancement of science through graduate education in universities was extremely limited (Cunha, 2007). Brazil's priority at the time was to implement what turned out to be a successful national development policy. The initiative was quite effective from 1930 to 1945 and focused mostly on industrialisation. In the first years, the country was able to put its idle industrial capacity to use, but the expansion of national production pressured imports and led to inflation and unattractive exchange rates. The result was a reduced interest in private and foreign investment in Brazil (Cano, 2015).

What followed contributed to shaping the primary *modus operandi* of the Brazilian Government in terms of state matters. Since it was not possible to rely on the market to act, the State stepped up to continue to develop, which meant founding its own industries, such as the National Steel Company (1941) and the Rio Doce Valley Company (1942, known today as Vale) (Cano, 2015; Gouvêa, 2012).

While all this happened, few steps towards scientific development were taken. Among the relevant was the creation of the University of São Paulo (USP), in 1934. The institution was conceived as an ambitious project to produce a new intellectual elite in Brazil and, for almost two decades, it would graduate most of the scientific workforce in the country. Furthermore, the HEI was the first in Brazil to really strengthen its research capacity through international outreach, as it became the new home for various European scientists who moved to Brazil to escape the threat of war (da Fonseca, 2015).

Still, World War II erupted, and so did a movement to take science out of its academic isolation to serve economic and social development. Such ideas had been spreading in Europe for the best part of a decade, counting with strong advocates such as J. D. Bernal and his book "The social function of science". In the inevitability of war, those ideas had to be put into practise, as scientists played critical roles in cracking the German secret code, improving radar, and many other relevant activities, including the development of the atomic bomb (Schwartzman, 1989).

As World War II ended, Brazil was sparked by the idea of scientists focused on solving concrete problems, and the country sought inspiration in the success stories of the war (Cano, 2015). As an example of that, physicists and mathematicians who arrived from Italy in the 1930s had helped educate a group of competent researchers in high-energy physics at USP. This group was then called to develop the country's nuclear policy, a project that would lead to the creation of the Brazilian National Council for Scientific and Technological Development (CNPq), in 1951 (Schwartzman, 1989).

Despite the eventual failure of the nuclear project, in part because of the United States veto (Schwartzman, 1991), the initiative integrated a substantial evolution in the Brazilian political, economic, and social progress in the early 1950s. In this new phase of the development project, the State continued to push forward, founding new institutions such as the Commission for Industrial Development (CDI), the National Bank for Economic Development (BNDE), and the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) (Gouvêa, 2012). The last one would be of extreme significance for the birth of the Brazilian National System of Graduate Education (SNPG) (Rothen, 2018).

Created by Decree 29.742/1951, CAPES started as a commission in charge of a national campaign for "ensure the existence of specialised personnel in suitable quantity and quality to meet the needs of public and private efforts aimed to develop the country" (BRASIL, 1951). For that, a mixed group of policymakers, academics, and even representatives of the financial sector received unprecedented autonomy to implement pioneering programs for graduate education in Brazil (CAPES, 2010).

2.2.1 CAPES: the first years

The decree that created CAPES (BRASIL, 1951) established a series of objectives for the initial campaign. They can be summarised in three points:

i) Study the country's needs in terms of the highly qualified personnel necessary to work towards social and economic development;

ii) Address such needs by mobilising, in cooperation with public and private institutions, existing resources to provide training opportunities to the most capable individuals, in particular to those without their own financial means;

iii) Promote the expansion of graduate education and research centres in Brazil;

The study by Gouvêa (2012) described how CAPES was organised in order to carry out the proposed objectives. At the core of the campaign, a Technical and Scientific Program (PQTC) was implemented, with the support of a Statistics and Documentation Service (SED). Their job was to map the research infrastructure already installed in the country and to understand its deficiencies with respect to scientific personnel. To do so, a University Program (PgU) was also implemented, counting with the support of a Scholarship Service (SBE).

The strategy adopted by PgU was to search for talented people in academic institutions and provide them with direct support in research infrastructure and staff (Balbachevsky and Schwartzman, 2010). Such support would include hiring foreign visiting professors, assistants, and even technicians to work in Brazilian institutions, as well as granting scholarships for select national researchers to study abroad (Gouvêa, 2012). Although the statistical data for that period seem not to be completely reliable, C. B. Martins (2018) estimates that, from 1953 to 1959, an average of 1.200 students moved abroad every year.

As a result of these policies, a new generation of Brazilian researchers was formed. Many of them graduated abroad, often in the United States, and most were back in Brazil to assume the scientific leadership in universities by the early 1960s. From their international perception of what a research program could be, they would actively participate in the design of the new master and doctoral courses so needed in the country (Balbachevsky and Schwartzman, 2010; Gouvêa, 2012; C. B. Martins, 2018).

2.2.2 The persistent design of Brazilian graduate education

In the first decade of its existence, CAPES was able to both strengthen the few research departments in Brazilian universities and contribute to the development of a critical mass to advance science in the country (Gouvêa, 2012). Now, the international experience of the researchers who graduated abroad would challenge the established design of the national institutions to achieve the once planned result: the university as an educational and research environment (Sucupira, 1980). According to Fávero (2006), this modernisation movement was clearly seen in 1961, with the creation of the University of Brasília (UnB).

Brasília is the current capital of Brazil. Founded in April 1960, the city was built in just 41 months, in the very centre of the country. Designed in the shape of a plane and erected on the margins of an artificial lake with 80 km of shore length, Brasília was a symbol of progress. Its first university should reflect this modernity, and that was made clear in the HEI's original plan, which stated UnB was "projected on the same bases as the teaching and research centres that are revolutionising the modern world" (D. Ribeiro, 1961).

The plans that led to the structure of UnB included innovations that would soon be reflected in broader legislation: Law 4.024 (LDB), which sets the guidelines and bases for national education. The LDB decentralised the educational system; gave freedom for each institution to organise its curriculum; provided academic, administrative, financial, and disciplinary autonomy to universities; and so on. For this study, the most significant contribution was that graduate education became a permanent task of the university. It constituted a system of regular courses to deepen the training received during undergraduate studies and could lead to academic degrees (BRASIL, 1961; D. Ribeiro, 1961; Sucupira, 1980).

Despite the conceptual advances of UnB and the LDB in graduate education, the lack of accumulated tradition in research at Brazilian universities hindered the expected growth of the system (Balbachevsky, 2004). 1965 would bring a decisive milestone for the change, as the Minister of Education and Culture, Suplicy de Lacerda, asked the Federal Education Council (CFE) to further define and regulate graduate courses, given the imprecise understanding that still reigns about their nature and purposes (C. B. Martins, 2018).

The resulting document, Report 977/1965 (known today as the Sucupira Report in honour of its lead author, Newton Sucupira), was approved by CFE to become the most important document in the history of graduate education in Brazil. Among its core ideas and definitions, we can find (CFE, 1965):

i) Although graduate education might refer to any course that follows an undergraduate degree, specialisation and improvement courses focus on professional development and are to be considered as *sensu lato*. They are not included in the scientific policy initiative at hand;

ii) Graduate education *sensu stricto* is intended to train researchers and professors for higher education courses; to stimulate the development of scientific research; and to prepare high-level personnel to meet the needs of national development in all sectors. It consists of two successive cycles, "equivalent to the master and doctor of the American system";

iii) The master's degree is not a necessary prerequisite for enrolment in a doctoral course. Certain fields of knowledge may even offer only doctoral programs, with direct access after an undergraduate course;

iv) The master's course can be more than a mere preliminary stage in the path to the doctorate. It may also be seen as a terminal degree;

v) The master's candidate must produce a dissertation revealing mastery of the chosen theme and the ability to systematise; the Doctoral candidate must defend a thesis that represents research work, making a real contribution to the knowledge of the subject;

vi) In addition to the thesis or dissertation, candidates must follow a certain number of courses, participate in seminars and research works, and submit to a series of exams;

vii) The master's degree would take at least one year to complete, while the doctorate would take a minimum of two years (the "ideal" duration would soon be established as two and four years, respectively).

The proposed model was strongly influenced by the American graduate system at that time. The Sucupira Report (1965) recognised the inspiration, citing the Robbins Report (1963) to support the decision. Such a document presented research on the conditions for the expansion and improvement of British higher education, recommending the adoption of techniques and processes from the North American system in British graduate studies. Furthermore, it is possible to infer that the choice for the American model would also be grounded on its familiarity to the new generation of Brazilian researchers with international experience, as most of them had graduated in the USA.

By the time the Sucupira Report was approved, there were only 38 graduate courses active in Brazil, 27 of them master's and 11 doctorates. In the decade that followed, the growth was enormous, and by 1975 this number jumped to 429 master's and 149 doctorates (Balbachevsky, 2004; C. B. Martins, 2018). The results came from the direct impact of Report 977/1965, as the document provided a precise shape for graduate education; one that higher education institutions could use as a guide to implement real change in the Brazilian science system (Sucupira, 1980).

From the observed growth, it is relevant to note the significant number of master's courses. At the time, they were seen as the most efficient way to quickly address the pressing need to form new scientists in Brazil. Although the Sucupira Report (1965) recounts objections among its authors on the adoption of the master's level in Brazil, the courses were not only embraced but also shaped resembling a short doctorate. As the model endures today, one may argue that the master's degree is still of high relevance to Brazil, both from the formative and scientific perspectives.

2.3 The research system as public policy

The Sucupira Report (1965) was able to advance previous efforts to develop Brazilian graduate education. Further initiatives recognised its value, either by confirming or strengthening its main concepts (Sucupira, 1980). The university reform of 1968–1969, for instance, established that academic degrees would be among the main criteria for admittance and progression in the teaching career. The same reform also stipulated that universities should institute programs to improve teaching staff, creating a demand that would further stimulate the growth of the graduate system (C. B. Martins, 2018).

Another relevant initiative to advance science in Brazil came with the foundation, within the structure of the Ministry of Education and Culture, of the National Council of Graduate Education (CNPG). Created by Decree 73.411 (BRASIL, 1974a), the council was formed by two ministers (education and planning); presidents and directors of institutions such as CAPES, CNPq, BNDE and CFE; and by presidents from two public and one private universities. These distinguished members were in charge of: i) elaborating a National Plan for Graduate Education (PNPG), ii) proposing the necessary measures to execute and regularly update such plan.

2.3.1 The National Plans for Graduate Education

According to the initial diagnostics of the National Council of Graduate Education, the observed expansion of the Brazilian science system had been partially spontaneous, somewhat unbalanced, and pressured for conjunctural reasons. This differed from the intended coordinated form, although CAPES had been making efforts, from its very foundation, to understand and act on the needs of the country. Changing this scenario to one of stable and balanced growth would require graduate education to move up in the national agenda, by becoming the object of state planning. For that, the idea was to issue continuous and subsequent National Plans for Graduate Education (PNPG) that should be recognised as guides for future initiatives and efforts on the issue (CNPG, 1974).

The PNPG 1975–1979 was then published in 1974, containing contextual analyses of the graduate system and a series of goals established according to the priorities at the time. The document included measures to be taken at all institutional levels of coordination, planning, execution, and regulation of graduate activities. Mostly, the first PNPG reinforced main concepts already presented in this study, such as the need for graduate courses to address development demands; the aim to better train university teachers, highly qualified personnel for all sectors, and researchers for scientific work; and the need to keep education and research integrated within all levels. Nevertheless, some major policy decisions from the PNPG 1975–1979 (1974) would continue to shape the Brazilian National System of Graduate Education:

i) Graduate students are trained professionals who could choose the job market instead of continuing their studies, so the path to master's and doctorate degrees must be attractive. Most students should work full-time on a scholarship robust enough to meet their needs. CAPES, CNPq and other agencies should design a harmonious policy to meet these demands and the funding requirements of research programs;

ii) The primary funding for research and graduate education should come from the higher education institutions themselves. This involves costs with infrastructure, personnel (professors' salaries included), expenses with teacher training programs, etc. Resources from government agencies such as CAPES and CNPq should be seen as complementary;

iii) The expansion plan for graduate education should be based on efficiency. This meant the general rule for new courses was to prioritise established universities, rather than isolated research centres, as these presented only a subset of the activities in the educational-scientific work matrix. From the perspective of investment, the criteria were to consider the greatest possible multiplier effect, meaning that smaller budget increases for universities would probably yield larger results. Finally, the PNPG considered the costs of graduate education too high to be maintained with resources from school fees, while maintaining good quality levels. Thus, government support for graduate courses in private institutions would be considered only for particular fields and situations;

iv) The accreditation of graduate courses had been a plan since the Sucupira Report (1965), which believed it was necessary to guarantee the quality level within the system. At the time of the first PNPG, the Federal Education Council was in charge of this process, but it was carried out *a posteriori*, lacking the appropriate mechanisms and procedures to be successful (Balbachevsky, 2004). The plan identified this limitation and called for an analysis of future alternatives, leading to a transfer of responsibility to CAPES.

In the complexity of a PNPG, some goals and consequent actions were significant for the design of a national science policy. For example, the section about the current state of SNPG will show that items (ii) and (iii) have led the system to a scenario where the majority of graduate education is concentrated in a few institutions, mostly public universities. Additionally, as much as item (iv) has led to immediate action, pushing for a new evaluation system within CAPES, item (i) is a goal yet to be achieved, as only about half of the master's and doctoral candidates have got access to a scholarship, even today.

2.3.2 CAPES, evaluation and the future of the PNPG

From 1964 to 1985, Brazil was under the rule of an authoritarian military dictatorship. While similar regimes in Argentina, Chile and Uruguay dismantled public universities (Hostins, 2006), in apparent contradiction, Brazil strengthened science, technology and higher education (Schwartzman, 1991). The first National Plan for Graduate Education came in 1974 as a critical step in this direction, but that was also an important year for CAPES. At that time, the agency was restructured, acquiring a new level of administrative and financial autonomy that would help carry out its new mission of evaluating graduate courses (BRASIL, 1974b).

In an effort to improve the quality assessment conducted by the Federal Education Council, CAPES conducted the first general evaluation of graduate courses in 1976. At the time, the agency had to decide on the allocation of student scholarships, and, instead of focussing on individual level assessments, it evaluated the quality of graduate programs as a whole, providing block grants to research units according to their achievements. By connecting performance with funding from the very beginning, the evaluation system was built as a powerful science policy instrument in Brazil, capable of steering the SNPG in the direction proposed by the national plans. To this day, a good performance in evaluations improves the chances of substantial support in student scholarships, research infrastructure, and funds (Balbachevsky and Schwartzman, 2010).

The evaluation system launched by CAPES at that time became a periodic, widely publicised process that has been improved at every cycle. Although it has acknowledged shortcomings, the system engages the academic community in committees using mixed methods to determine quality levels considering differences between fields of knowledge (Balbachevsky, 2004; Rothen, 2019).

The complexity of such evaluation cannot be briefly captured, and we shall dedicate a whole study to investigate its evolution and current perspectives, in a similar way to what we have done for this chapter. The important message that should remain for now is that CAPES' evaluation was converted into one of the most effective instruments to build the foundations of the Brazilian scientific community and guide its growth in the decades to come (Schwartzman, 1991).

In addition to the recently acquired evaluative role, by 1981 CAPES experienced yet another progression in its strategic importance in science policy (Hostins, 2006). At the time, Decree 86.791 extinguished the National Council of Graduate Education, transferring all of its responsibilities to CAPES, including the formulation of the new National Plan for Graduate Education (BRASIL, 1981).

The II PNPG was then published for the 1982–1985 period. The document expanded and updated the goals presented in the first plan, but now emphasises the need to further assess the quality of graduate education. The evaluation system under development would then become more expressive with increasing participation of the scientific community. Furthermore, the new plan would provide additional support for research infrastructure, more financial stability, and autonomy for graduate courses, and, for the first time, would contemplate strategies to reduce institutional and regional asymmetries (CAPES, 1981).

By the end of the II PNPG coverage, Brazil had once again become a democracy. That was a time of stability for CAPES, without any significant direct impact from the regime transition. As a consequence, the PNPG 1986–1989 was, as its predecessor, built mainly on a base of updated and incremental goals. The III PNPG hardened the criteria for the accreditation of new courses, while proposing inductive action in strategic areas. It recognised shortcomings of the original scholarship policy, both in coverage and in the loss of purchasing power over the years. The plan also enhanced the role of the evaluation system to fight high student dropout rates, long graduation times, low overall performance, etc. On top of that, evaluation should also guide an increase in investments in the best courses, with the aim of improving their productivity (CAPES, 1985).

In addition to such incremental ideas, the plan registered a concern that would become quite influential in shaping the SNPG: there was an undesired dependency of short-term extra-budgetary resources for research within HEI. According to the PNPG, the issue created instability for research groups, often leading to temporary or permanent interruption of their work. In addition to that, institutions and researchers had spent too much time on the continuous development of fund-raising projects. The proposed solution was to highlight specific funds for research and graduate studies in institutional budgets.

The result was a reinforced idea of research centres within HEI, without base funding concerns, and counting with dedicated resources, infrastructure, and personnel. Additional investments, including scholarships, would come from agencies such as CAPES and CNPq, and were usually linked to the centre's performance. These research centres must include a master's or doctorate degree, often both, and are officially known as graduate programs (PPG).

One of the characteristics of such graduate programs is the overall absence of the "researcher" figure. As peculiar as it seems, one of the consequences of the ever-present integration between education and research is that, in PPG, researchers invariably fit into two categories: students or professors. Considering that most of these programs in Brazil are within public universities, the most common path to become a professor is through public tenders.

The idea that every citizen could have the ability to work for the government solely on the basis of "talents and virtues" was first introduced by the Imperial Constitution of 1824. This concept would be refined over time to determine that access to public employment would depend on prior approval in an open selection based on exams and academic degree analysis, whenever relevant. According to the current constitution, from 1988, such tenders are mandatory for most government jobs, except office nominations or select positions of trust. This process aims to eradicate any gender, race, social, or age discrimination in government hiring. In addition, to ensure security in the face of potential political influences, public servants acquire stability after an evaluation period, usually of three years (BRASIL, 1988). Although not perfect, the process allows professors to be hired according to explicit criteria, with the possibility of gaining tenure after a short period of time.

The constitution of 1988 also formalised other relevant concepts. First, article 207 confirmed that i) "universities enjoy didactic-scientific, administrative, financial and patrimonial autonomy"; and ii) "they will obey the principle of inseparability between teaching, research and extension". Then, article 209 established that iii) "the private initiative can offer educational services, but are required to a) comply with the general rules of national education, and b) submit to authorisation procedures and quality assessment by the Government".

2.4 The latest chapters in designing the SNPG

The 1980s were very prolific for the evolution of the science system in Brazil. The military government had a nationalist ideal of turning the country into a world power, which motivated the push for a Science & Technology agenda (Rothen, 2019). This plan advanced in the first years of the democratic regime, but Latin America had been experiencing what became known as "the lost decade", a period of successive economic crises that dominated the region. In Brazil, that was actually one of the reasons for the political impasse that ended the military dictatorship, and the effects of such crises would eventually reach the science system as well (Ferreira and Moreira, 2002).

A significant sign of that came with the 1990 federal government restructuring, in which CAPES was extinguished (BRASIL, 1990b). Due to the negative repercussions within universities and the decisive action of the academic community, the agency was re-established in less than a month (BRASIL, 1990a) and later acquired public foundation status (BRASIL, 1992). Despite this renewal that even improved CAPES' standing in the federal structure, in the early 1990s Brazil was facing hyperinflation, reaching the mark of 80% a month, at one point. That was a time for restraint and, although the main characteristics of the current SNPG had already been shaped, the next chapters of this history would be put on hold for a few years.

The IV PNPG, for example, would only come into development in 1996. The new document would try to resume the evolution of the system, deal with structural factors hampering its performance, focus on the identification of asymmetries that need to be addressed, push for greater integration among S&T organisations. Despite the extensive discussion around a preliminary version, the enduring crisis and the lack of articulation among the funding agencies at the time would lead the final plan to be abandoned (CAPES, 2004).

Although no official PNPG would be published until 2004, CAPES acted on a series of recommendations from the plan's draft (Hostins, 2006). Among them, the proposal to diversify the graduate education model would lead to one last significant addition to the PNPG before the turn of the century: the professional master's courses (CAPES, 2004).

2.4.1 The professional modality of graduate courses

The Sucupira Report (1965) determined master's courses could be seen as a terminal degree, helping to advance the abilities of professionals, or as a preliminary stage in the path to the doctorate. After performing an analysis of how such courses developed over three decades, CAPES was concerned that the dynamics of graduate education had led these courses predominantly to the second category, and the master's degree was not fulfilling their whole intended purpose (CAPES, 1995b).

After three years of discussion within the agency and with the academic community, CAPES issued legislation creating a new type of graduate course: the professional master's. The modality was designed to apply the high-quality level *stricto sensu* approach to the training of professionals interested in problem solving by advancing existing knowledge (CAPES, 1998). To understand the different perspectives of academic and professional modalities, the Stokes (1997) quadrant model of the relations between basic science and technological innovation, seen on Figure 2.1, can be applied.

		Considerations of use	
		No	Yes
Quest for fundamental understanding	Yes	Pure basic research (Bohr)	Use-inspired basic research (Pasteur)
	No		Pure applied research (Edison)

Figure 2.1.: Quadrant model of scientific research, as proposed by Stokes (1997, p.73)

Stokes (1997) categorises research according to the search for fundamental understanding and by considerations of use. From this perspective, academic graduate courses in Brazil can be classified in Bohr's Quadrant, as they strive for the advancement of scientific knowledge without pressing concern with immediate application. On the other hand, the professional master's were designed to actively pursue knowledge toward immediate application, thus existing within Pasteur's Quadrant. The final quadrant, Edison's, might be appropriate to describe the *latu sensu* courses, as they focus on the training of professionals based on already established knowledge.

Over the years, there has been much debate in academia about the validity of the professional modality, with advocates on both sides of the discussion (Ferreira and Moreira, 2002). Despite the opposition, these courses have become quite representative within the SNPG, even in their role of producing more applied research outputs (Brasil, 2018). As a consequence, in 2017, the Ministry of Education authorised the professional modality at the doctoral level (MEC, 2017). By August 2020, CAPES had already accredited 53 of these courses, complementing the 873 professional master's in the system. (CAPES, 2021c).

2.4.2 Centralising access to scientific literature

From what has been presented so far, it becomes evident that the Brazilian SNPG is organised in a top-down structure, where government institutions have the power to enact their ideas by controlling the budget, the evaluation system and the legislative authority to promote changes. A significant example of the potential and dangers of such reality may be found in the Portal of Journals.

The Ministry of Education created, in 1990, a national program to support higher education institution libraries. Five years later, a funding program for journal acquisition was designed as a joint initiative of MEC, CAPES, CNPq and the Funding Authority for Studies and Projects (FINEP). The goal was to extend access to scientific literature for libraries related to graduate education. CAPES would deal with editors and subscribe to journals that would be sent directly to each participating institution (CAPES, 2020c; de Almeida et al., 2010).

Despite delays in journal acquisitions, mostly because HEI requests would often take longer than desired, the model worked well until 1999. At that time, a 53% cut in the program's budget enhanced challenges from the ongoing devaluation of Brazilian currency against the US dollar. There was a need for reform, and the Internet became a possible solution, as most scientific publishers were already digitising their collections (de Almeida et al., 2010).

From the restructuring of the acquisition program, the Brazilian Portal of Journals was launched in November 2000. With it, the 72 institutions offering graduate education at the time gained access to 1.419 digital journals at first. The Portal adopted a centralised acquisition model to optimise processes, improve bargaining power over publishers, and generate economies of scale. In addition to that, it also helped reduce asymmetries, as institutions would have access to the same collections, regardless of their size, budget, or location (CAPES, 2020c).

By the end of 2019, the Portal already provided access to 426 HEI, covering 49.247 journals; 331.565 documents such as books and reports; hundreds of databases of publications, patents, statistics and media; and more (CAPES, 2020c). Data from Geocapes (2021e) show nearly 190 million accesses to the Portal in 2019 alone, including over 59 million full-text downloads; all that with a yearly budget of around 100 million US dollars (CAPES, 2020b). Despite this success, at the time of its launch, there was little support from academia, as only the Brazilian Society for the Advancement of Science favoured the idea (de Almeida et al., 2010). Without the centralised, top-down structure of the Brazilian science system, the Portal would probably never exist.

Although direct action from the government and its top agencies has been able to shape and develop the SNPG, experience shows that such power comes with risks. For example, high managers at CAPES have consistently reported how changes at various levels of government have led to questions about the very existence of the Portal of Journals (SBPC and ABC, 2020a).

An example of such threats came in 2015, with the announcement of budget cuts to the Portal in the following year. At the time, SBPC, ABC and seven other scientific entities sent a letter to the minister of education, stating "it would be inconceivable for the country to be devoid of an instrument so relevant to the regular functioning of our education, science, technology and innovation system" (Nader et al., 2015). Fortunately, such initiatives helped contain threats not only to the Portal of Journals, but also to several core components of the SNPG. Nonetheless, there is always the apprehension that misguided actions could change all that.

2.5 The SNPG today

The historical panorama shows that the evolution of the Brazilian National System of Graduate Education has resulted from decades of coordinated work from CAPES, CNPq, MEC, higher education institutions and other actors. Although they have continued to be active in the past twenty years, the core design of the graduate system was established, and most of the actions not covered in this study had only incremental effects on the evolution of the system.

For instance, in the 21st century, the PNPG have continued to give macro-political direction for research and graduate education, through diagnoses and the establishment of conceptual and practical goals. So far, two additional plans have been published, one for the period 2005–2010 (CAPES, 2004) and the other with a decade-long coverage of 2011–2020 (CAPES, 2010). Both have proposed strategies for the induction of vital research areas, to improve the performance of the system, to expand financial support to programs, to evolve evaluation for better quality assurance, etc.

Although the most recent PNPG did not play the same role in shaping SNPG as its predecessors, they have been essential to guide the impressive growth of the system presented in this section. Here, we examine information obtained from different national databases to see what the system has become. In particular, we will work with CAPES ' integrated system of graduate education data – Sucupira Platform (2021c); and the Georeferenced Information System of the Agency, Geocapes (2021e).

2.5.1 Graduate courses and programs

All basic research conducted in Brazil takes place within the realm of graduate education, and the same is true for around 95% of all S&T research in the country (SBPC and ABC, 2020b). As we have discussed, such a graduate system is organised around graduate programs (PPG): research centres offering either or both a master's and a doctorate, always in the same modality (academic or professional). Such PPG are necessarily part of higher education institutions.

As Figure 2.2 shows, this century has been quite significant for the growth of this system. From 2001 to 2020, the number of graduate courses has gone from

3.292 to 7.146: a 117% increase. At the time the data extraction took place, 4.601 of these courses were master's (3.728 academic and 873 professional) and 2.545 were doctorates, 53 of those in the professional modality.

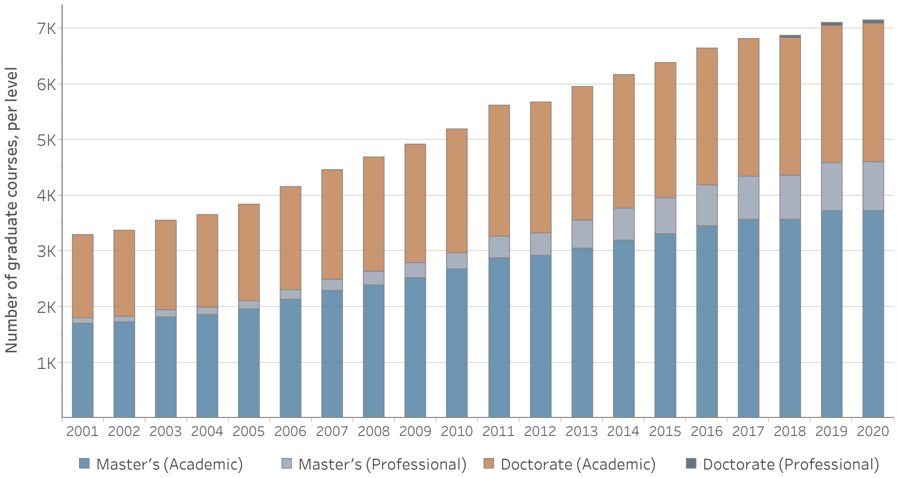


Figure 2.2.: Number of graduate courses active per year (2001 – 2020)

From the PPG perspective, the 7.146 courses active in 2020 are organised into 4.690 graduate programs, according to what is displayed in Table 2.1.

Table 2.1.: Organisation of active courses into graduate programs (2020)

Modality \ Level	**Master's**	**Doctorate**	**Master's & Doctorate**
Academic	1.322	86	2.406
Professional	823	3	50

Although the Sucupira Report (1965) established isolated doctoral courses to be acceptable, even desirable for specific fields, Table 2.1 shows that there are only 89 programs in this situation. On the other hand, the number of PPG with only master's courses is quite significant in the Brazilian National System of Graduate Education, serving as evidence of the importance of such courses for the development of science in Brazil.

Taking into account the perspective of graduate programs, Figure 2.3 shows the geographical growth of the system, contrasting the current distribution of PPG with that of 1975, the first year of coverage of the original PNPG.

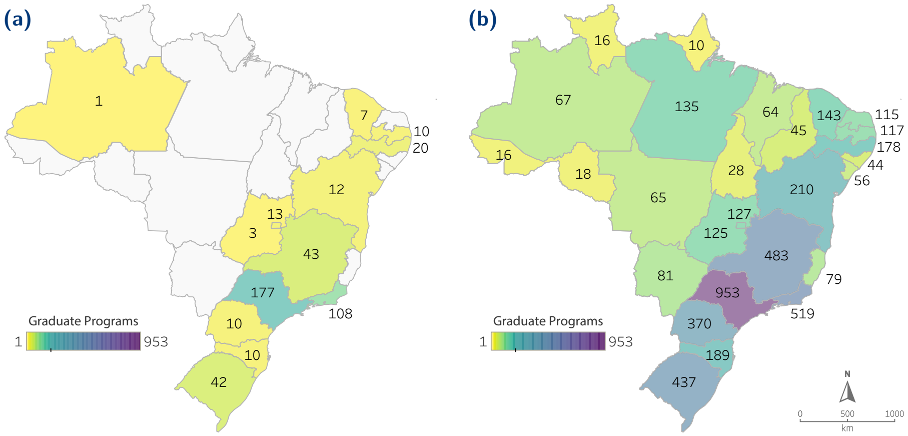


Figure 2.3.: State distribution of graduate programs in Brazil, in: (a) 1975; and (b) 2020. Data source: Sucupira Platform (2021c).

The contrasting maps included in Figure 2.3 show a considerable improvement in the distribution of graduate education. In 1975 (a), we can see that more than half of the Brazilian states did not count with a single program, and by 2020 (b) all of them were contemplated. If the dispersion of the population is considered, the current distribution is even more balanced, and we can use the state with the most PPG as an example: São Paulo hosts 953 programs (20,3% of the total), while containing 21,9% of the country's 210 million inhabitants, according to recent estimates from the Brazilian Institute of Geography and Statistics (2021). On the other extreme, at the north of the second map, we see the state of Amapá, with only 10 programs. Although a 0,2% representation is far from adequate, the discrepancy seems less extreme if we consider the fact that the state accounts for only 0,4% of the country's population.

Despite the scenario presented by a state-based analysis, a city-based distribution of graduate programs makes an unbalanced configuration apparent. In Brazil, out of 5.570 municipalities, only 307 count with graduate education, as can be seen in Figure 2.4. This distribution of courses reflects a concern that had already been highlighted by the PNPG 2011–2020: Research centres are still too concentrated in metropolitan regions, notably in coastal areas (CAPES, 2010). This new map makes the fact more noticeable because it shows how states with large numbers of courses have already achieved some interiorisation, while the others register substantial concentrations mostly at the state capitals. So, it is possible to say that Figure 2.3b reflects how far asymmetry reduction policies have come, but Figure 2.4 shows how far they still need to go.

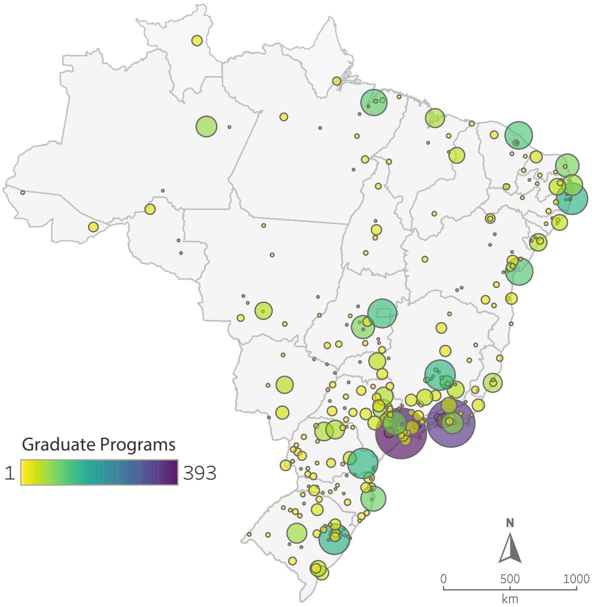


Figure 2.4.: Distribution of graduate programs per city

2.5.2 Higher Education Institutions

Higher education in Brazil had always been quite restrictive and only became a social priority in the 21st century. To increase access to the general population, a series of policies led to a tremendous growth in the number of higher education institutions, mostly through the expansion of private establishments (Rothen, 2019). In 2000, for example, Brazil had 1.180 HEI, but this number more than doubled by 2018, reaching a total of 2.537 institutions (INEP, 2020).

However, access policies that led to institutional expansion focused mainly on the undergraduate training of professionals, while the graduate system continued to be very meritocratic and restricted to a few HEI. As a result, only 432 of the 2.537 institutions currently offer graduate education. More than that, as a direct result of policies that prioritise investments to develop research within universities and large institutions, there is a significant concentration of graduate programs in a small group of these already select HEI.

The Pareto chart in Figure 2.5 confirms such perception, showing the accumulative contribution of every one of the 432 higher education institution to the whole graduate system, in descending order. Each colour represents 20% of the total number of programs.

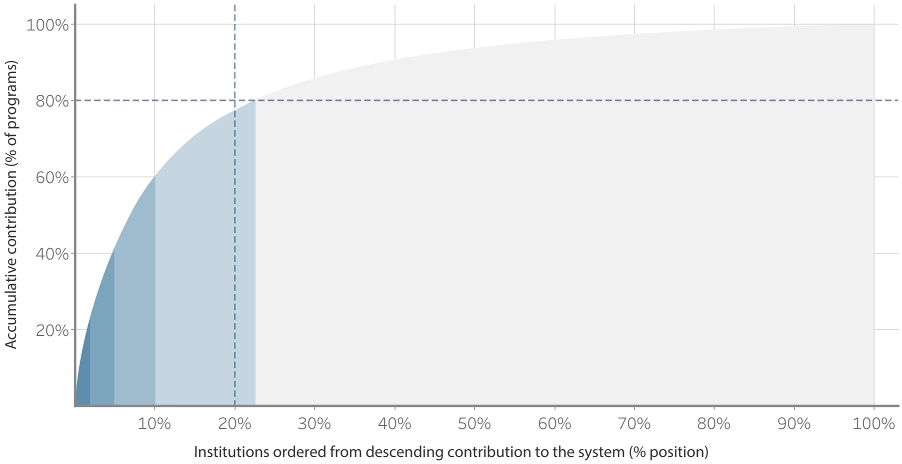


Figure 2.5.: Pareto chart on accumulative number of graduate courses, per HEI

From Figure 2.5, we can conclude that nearly 80% of all graduate programs are offered by 20% of the institutions. Looking at the first colour group, the situation is even more impressive, as 20% of the PPG can be found in just seven universities[1]: USP, UNESP, UFRJ, UnB, UFRGS, UFMG and UFPE.

Taking into account the policies that prioritised the expansion of graduates in public institutions, the top seven HEI were expected to be public. In fact, in the ranking of institutions with the most PPG, there is only one private university within the top 10%, and it comes in the 40th position in the list: The Pontifical Catholic University of Rio de Janeiro (PUC-Rio), which offers 34 distinct programs. For a better understanding of the public versus private dimension of graduate education, Figure 2.6 details the higher education institutions according to type.

[1] Explore the interactive dashboard at https://andrebrasil.net/publication/building-the-snpg/

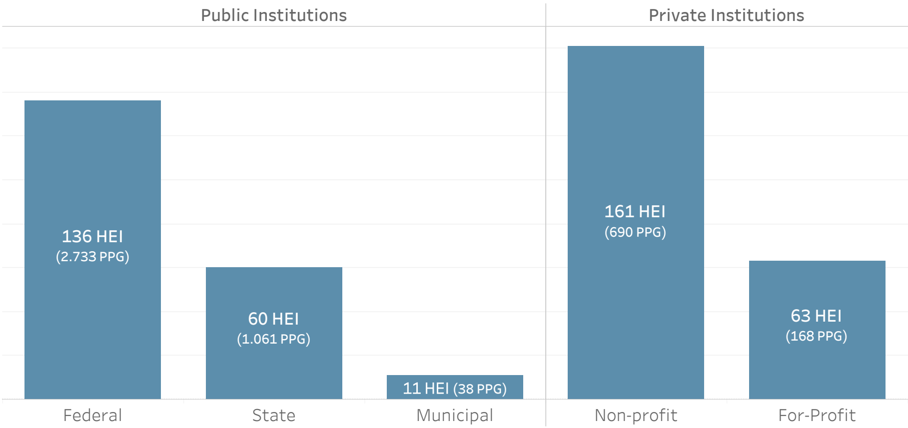


Figure 2.6.: Higher Education Institutions (HEI) in Brazil with active graduate programs in 2020, and the total number of PPG offered.

Of the 2.537 Brazilian HEI active in 2018, 2.238 (88,2%) are private (INEP, 2020). Figure 2.6 shows a different scenario for those involved in research, as this percentage drops to 52%. Although the number may still seem significant when considering the research policies discussed throughout this chapter, it is important to note that only 18,3% of all courses are found in private institutions. Besides that, 69% of all public HEI in the country offer at least a master's or doctoral course, while only 10% of private ones do the same.

The first PNPG stated that the costs to offer graduate courses are too high for institutions that rely on student tuition (CNPG, 1974). In Brazil, these are private organisations subdivided into two categories: non-profit and for-profit. The first group includes philanthropic and community HEI, as well as those with some religious orientation. They hold 80,4% of the programs offered by private institutions: an average of 4,28 per HEI, versus 2,7 in the for-profit group. In addition to the financial aspect, another element that could contribute to the minor involvement of private institutions in research and graduate education is their faculty profile, as discussed in the following.

2.5.3 Faculty

According to the 2018 educational census (INEP, 2020), the Brazilian higher education system counts with 384.474 faculty members, 173.868 working for public institutions and 210.606 for private ones. The workforce in the private sector is relatively smaller, averaging 95 faculty members per institution, while 581 are recorded in the public sector. Another difference is that the public side prefers full-time employment contracts (86,4%), while the private side prefers part-time (42,4%) and hourly contracts (30,1%). Finally, most of the faculty in public higher education institutions hold doctoral degrees (64,3%) while in private ones the master's degree is the predominant higher level (50,1%, against 25,9% with a doctorate). As the workforce available for research activities in private HEI appears to be smaller than in public ones, this can be an obstacle to implementing new graduate programs.

Considering faculty members active in graduate education, data from the Sucupira Platform (2021c) show there were 81.639 professors working in Brazilian PPG during the 2017–2018 biennium. Of these, 99,1% held doctoral degrees, and their distribution by gender and age group can be seen on Figure 2.7.

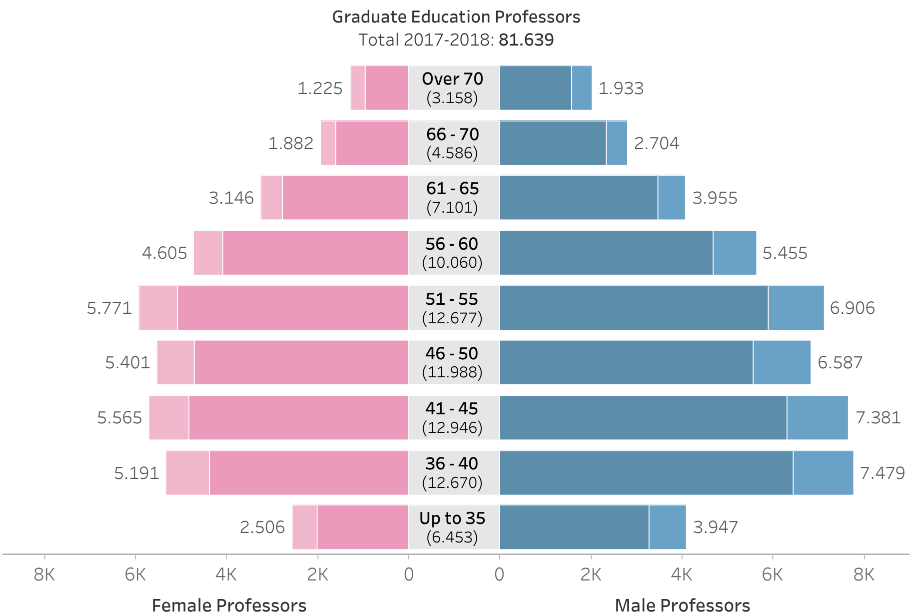


Figure 2.7.: Age and gender distribution of graduate education professors in Brazil (2017 – 2018). Lighter colours highlight those working for private HEI.

One of the first things we notice from the pyramid of faculty members is that the lighter parts of the bars, representing professors working for private institutions, are much smaller than the darker colours. In the 2017–2018 period, 85,7% of all faculty members worked for public institutions (which is not far from their 82% share of courses).

Another evident perception from Figure 2.7 is the skew towards the male side, which represents 56,8% of the total number of professors. The reasons for this are undoubtedly complex and are not the object of the present study. Regardless, we can propose two variables for future investigation:

i) The rules for retirement in Brazil differ by gender, as men generally work longer. They contribute an average of five more years to social security, which can lead to a gender imbalance in academia. This perception might be supported by a study conducted by Waltman et al. (2019), analysing the length of academic life according to gender (1998–2018). Such research shows above average dropout rates for women in Brazil, a fact that might be influenced by age of retirement;

ii) Maternal roles can also be a factor in gender asymmetry, as the relative representation of women increases along the earliest age groups. This hypothesis seems particularly promising when comparing all female professors in higher education with those acting in graduate programs. Considering only professors holding a doctoral degree, women up to 35 years of age represent 49,96% of the entire higher education system, but only to 38,8% of graduate education professors. For the next age group, 36–40, women represent 48% of the higher education system but only reach 41% in graduate education (CAPES, 2021e; INEP, 2020). How much of this difference comes from the choice to keep away from research while still holding a professorship may account for part of the gender imbalance.

2.5.4 Student Body

As the science system in Brazil is connected to the training of highly-qualified personnel, one of the leading indicators to determine its success comes from the number of students enroled or graduating from master's and doctoral courses. Taking into account the explosion of such programs over the past two decades (Figure 2.2), the resulting number of candidates was expected to follow. In this sense, Figure 2.8 shows the last 20 years of official data on enrolment at different levels and modalities of graduate courses (CAPES, 2021e).

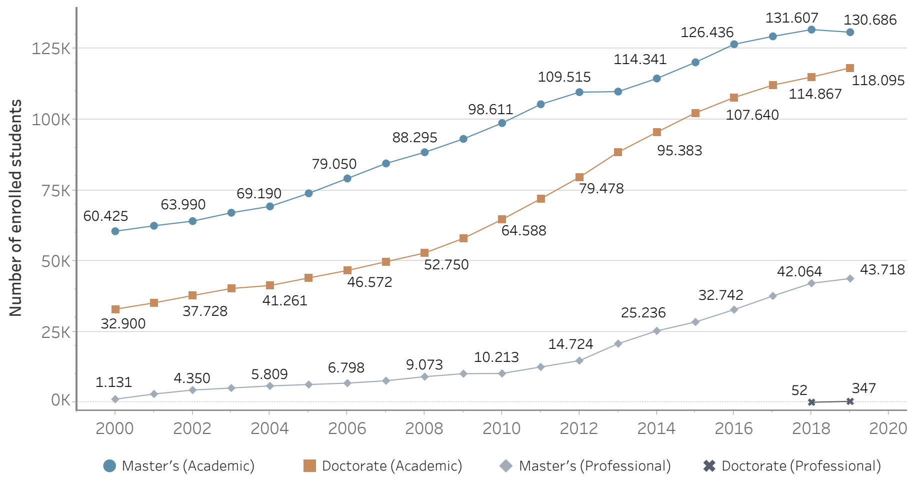


Figure 2.8.: Number of students enrolled in different levels and modalities of graduate education (2000 – 2019)

From 2000 to 2019, the growth in the number of graduate candidates enroled was of 210%. The results are remarkable and the professional courses seem of particular value. The modality was first authorised for the master's level in 1998, and the number of 1131 students in 2000 expanded to nearly 44 thousand in twenty years. One of the reasons this increase is notable is the fact that professional courses, even though tuition-free in public institutions, are not funded by scholarships. It remains to be seen whether the first recorded number of professional doctorate candidates, 52 in 2018 and 347 in 2019, will also experience this growth.

Regarding scholarships, earlier PNPGs proposed to fund all candidates in academic courses, but the plan has never become a reality. In 2019, CAPES granted 44.238 master's and 43.327 doctoral scholarships (excluding international grants), while CNPq provided an additional 13.402 and 11.252 for the corresponding levels. According to data from the National Council of State Funding Agencies (CONFAP), the state system also contributed with 13.536 scholarships (7208 for master's and 6328 for doctoral levels). Together, these agencies covered 51,6% of doctoral and 49,6% of master's students in academic courses, and any additional contribution from other sources, although welcome, would not be representative to impact the overall figure (CAPES, 2021e; CNPq, 2020; Dellagostin, 2020).

An additional aspect to note is that more women benefited from CAPES' scholarships in 2018, since they received around 53% of the master's and doctoral grants provided by the agency (CAPES, 2021e). The impact of this funding is aligned with the results of a bibliometric study considering trends of new researchers entering the science system (Waltman et al., 2019). The investigation indicated that by 2018, approximately 53% of new Brazilian researchers were female, which is quite a significant number considering that the global baseline is around 40%.

Despite the limitations in financial support for all students, the number of graduates each year has also been growing, as Figure 2.9 shows.

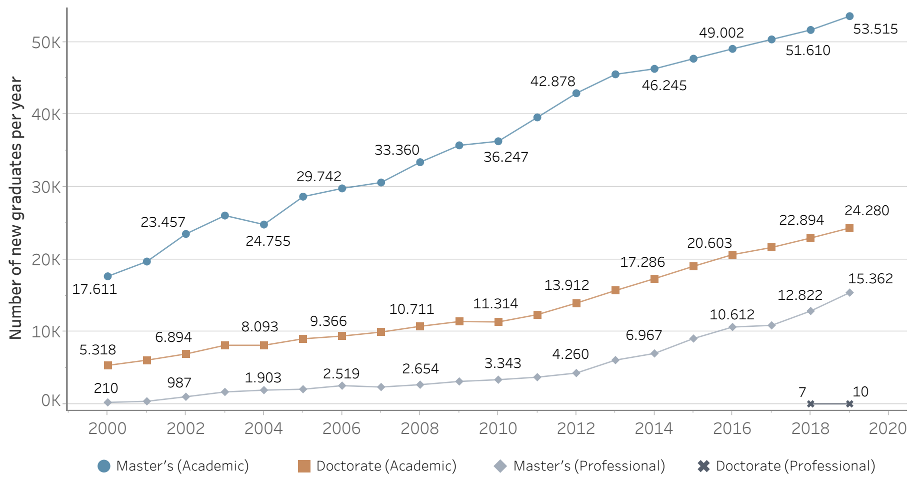


Figure 2.9.: Number of students graduating at different levels and modalities of graduate education (2000 – 2019)

As defined by the National Education Plan (BRASIL, 2014), Brazil established a goal to be reached by 2024: to graduate a minimum of 60.000 masters and 25.000 doctors per year. As Figure 2.9 shows, the desired number for the master's level had already been surpassed by 2019, with 68.877 graduates. For the doctoral level, the number was slightly below the mark, with 24.290 graduates. Although the COVID-19 pandemic might have an impact on the number of degrees awarded in 2020, considering the recorded number of PhD candidates and the four-year duration of the courses, the goal might still be reached before 2022.

2.6 A brief overview of the SNPG's impact

In recent years, several studies have discussed the evolution of one of the most evident results of the Brazilian National System of Graduate Education: its scientific production. The topic has been explored by this century's National Plans for Graduate Education (CAPES, 2004; CAPES, 2010), as well as by authors such as Balbachevsky and Schwartzman (2010) and de Almeida and Guimarães (2013) and others. These studies show a significant growth in Brazilian production, which eventually led the country to be among the top 15 in the world in the number of scientific publications.

A recent report from Clarivate Analytics (2019) analysed Web of Science data (2013–2018), ranking Brazil as 13th in the world in the output of research papers, ahead of countries such as the Netherlands (14th) and Russia (15th). Data from the report corroborate the narrative and findings presented in this research, for example, by showing that around 60% of the country's scientific production comes from only 15 institutions, all of them public universities. The report also shows a growth of 30% in the number of papers published by Brazil in the analysis period (twice the average growth in the world). This is consistent, for instance, with the observed increase in graduate education enrolment and the number of graduates in the same period.

The growth of Brazilian science above the world average is a positive result for the SNPG, but absolute numbers should be interpreted with caution. For example, while Brazil has a slightly larger scientific production than the Netherlands, the country immediately behind it in the Clarivate ranking (2019), Brazilian population surpasses 210 million people, more than 12 times more than the European nation. From this point of view, in the 2013–2018 period, Brazil produced around 130 papers per 100.000 population, while the Netherlands published more than 1.500. This relative perspective shows that Brazilian science is further from the top than we would expect from looking at such a simple measure as the number of papers published.

Another insight from the comparison of the scientific production from Brazil and the Netherlands relates to publication impact. An analysis of WoS data for the two countries shows that they perform quite differently in terms of the PP(top 10%) indicator: a common proxy of research excellence based on the percentage of publications in the world's top ten percent of highly cited papers.

From 2013 to 2018, Brazil performed under the expected value of 10% (6,9%), while the Netherlands exceeded it (15,3%).

The analysis also shows that internationalisation is an important component in increasing the impact of Brazilian publications. Taking into account papers published in collaboration with international partners, Brazilian PP(top 10%) increased to 12,2%. When cooperating with select countries such as the Netherlands, Australia, and Japan, the results were even better, surpassing 20%. That may suggest that finding strategic international partners can increase the impact of Brazilian research, at least from a scientometric point of view.

Nevertheless, this chapter indicates the SNPG may not be a typical representation of a standard science system. Thus, its impact must be measured from a broader perspective, for instance, by considering graduate education degrees granted annually. Research by the Centre for Strategic Studies and Management (CGEE, 2016) compared the number of doctoral graduates in Brazil with 27 OECD countries. In 2013, the year of data collection, Brazil issued 7,6 doctoral degrees per 100,000 people. This performance was ahead of Mexico (4,2) and Chile (3,4) only, and it was quite far from the group's median of 25,4 graduates per year. By 2019, Brazil had graduated 11,6 new PhDs per 100,000 population, much closer to what Japan (12,9), Turkey (11,5), and Hungary (10,8) were doing in 2013. If considered that CAPES' official Portuguese name translates as "Coordination for the Improvement of Higher Education Personnel", the results are not only positive, they are a reflection of the primary goals that contributed to the birth of the SNPG.

Another core concern discussed as a motivation for the design of the SNPG was Brazilian development. By 1995, CAPES analysed the evolution of graduate education and considered that there was a need to improve the participation of scientific research in the country's economic and social development (CAPES, 1995b). The envisioned solution became a professional modality of graduate program (CAPES, 1998). Whether due to the effect of this new modality or the mentality that led to its creation, Clarivate's report (2019) shows how, after decades of slow expansion, industry collaboration in academia started to grow exponentially at the turn of the century. Looking at the ten higher education institutions that collaborate the most with authors from the industrial sector, joint publications went from around 160 per year, by 2000, to nearly 1600 in 2017. The numbers are still modest, but the observed growth offers promise.

From an industry perspective, Petrobras was responsible for 14% of the collaborative output (2015–2017), due to a strong integration program with universities. However, dozens of other national and multinational companies have been collaborating with Brazilian academia in the pharmaceutical, agricultural, and many other sectors (Clarivate Analytics, 2019). More than that, there are many success stories of strong bonds formed between industry and the SNPG, and among them is the strategic partnership between the Aeronautics Institute of Technology (ITA) and the Brazilian Aviation Company (Embraer).

With the commercial success of its first regional jets, Embraer started the 21st century investing in a specialisation program for the development of its aeronautical engineers. Due in part to the elevated costs of the program, a better solution was in order, and it came from a collaboration with ITA, a study centre founded in the 1950s inspired by the model adopted at the Massachusetts Institute of Technology (MIT). ITA became a national benchmark in developing aeronautical technologies, and its joint professional master's degree with Embraer was accredited in 2003. The course has been so successful that Boeing wanted a guarantee of its maintenance in recent negotiations for the purchase of the Brazilian company (Barata, 2020; de Andrade et al., 2005).

Although the growing industry collaboration with academia may be a recent development, its impact is becoming more evident every year. From a multidimensional perspective, its combination with the social, economic, scientific, publishing impacts may add to a better perception of a more precise picture of the contribution of SNPG to Brazil and the world.

2.7 Conclusion and perspectives for the future

The Sucupira Report (1965), while building the core definitions of the future SNPG, stated that Brazil was yet to create a tradition in graduate education. The present study aimed to achieve a better understanding of the long-term historical process crafting such tradition. As comprehensive as it was intended to be, it is far from complete, as the subject could easily justify a lifelong investigation. What was possible within this chapter was to recount critical moments in history that shaped how science is done in Brazil, especially in its relation to graduate education. From this research, we can understand some of the reasons behind the design of the system, as well as its most peculiar aspects. The following paragraphs summarise part of these findings.

First, **the Brazilian science system was not a spontaneous creation**. It was built as a result of mostly consistent public policy, developed over multiple regimes and governments for several decades. The resulting system is a unique product of social, political, and economic conjunctures, and the path behind it can never be ignored moving forward.

Secondly, **the SNPG is the science system in Brazil**. For more than a century, the core idea of having the university as the house of science has guided policy in the country, to the point that the indivisibility of research and education became a constitutional matter. As a direct consequence, the entire science system was structured around graduate education, and the master's and doctoral courses are responsible for the absolute majority of the national S&T research.

In the third place, **the SNPG has, until recently, been too academic**. Although the pursuit of development has motivated much of the significant science policy in Brazil, the idea of the type of research that is applied to solve social and economic problems began to expand only in the past two decades. This is probably the result of funding models, faculty hiring methods, limits of the evaluation model, and pure tradition. A professional modality of the graduate course was implemented in 1998, with the aim of promoting balance. The results have been promising at the master's level. Despite lingering opposition, the first professional doctorates have joined the system in 2017.

Fourth, **the SNPG numbers are significant**. There are 432 higher education institutions offering graduate education in Brazil. They are spread over 307 cities across all Brazilian states and hold 7.146 master's and doctoral courses organised into 4.690 graduate programs; 18,7% of which are in the professional modality. Furthermore, almost 80% of the programs are concentrated in only 20% of the HEI, most of them public. In terms of people involved, over 80 thousand professors are active in graduate education, supervising more than 290 thousand master's and doctoral candidates, which leads to over 93 thousand graduates every year (nearly 25 thousand of those at the PhD level).

In the fifth place, **the master's has always been essential in Brazil**. Despite the influence of the United States graduate system in the design of the SNPG, the Brazilian master's did not follow the American path that allowed such courses to represent professional competence in fields such as Engineering or Business Administration. In Brazil, master's degrees were mostly implemented as minidoctorates, and they have performed a significant role in the development of science in the country. Even though the advent of the professional modality has addressed the need for applied approach in these courses, the geographical asymmetry of doctorates in Brazil makes the academic master's still quite relevant as a scientific degree across the country. While top institutions such as the University of São Paulo can justifiably decrease the role of the master's when used as a stepping stone towards the PhD, such experiences cannot be adopted as a policy for the whole country.

Finally, **the SNPG has both weakness and strength in the concept of change**. Throughout this study, we demonstrated how the Brazilian science system is organised under a top-down structure of influential government agencies promoting change through regulation, evaluation, and funding. Such centralised structures have the potential to promptly produce change as decisions can be easily disseminated, often even imposed, throughout the whole system.

From a weakness perspective, the SNPG faces a lingering concern of disruption from political motivations. For instance, the government elected to run the country from 2019 to 2022 has consistently criticised the value of basic science, in particular the Social Sciences and Humanities. As a result, priority plans for CAPES, MEC, CNPq and others have impacted whole disciplines ranging from Mathematics and Physics to Sociology and Philosophy. Several funding cuts led to the mobilisation of scientific entities who were hoping to have a voice capable of minimising long-term damage to the SNPG (Saldaña, 2020).

Despite the current apprehension around the system's susceptibility to change, the same property has benefited SNPG in the past. According to testimony from four CAPES' ex-presidents, with over 20 years of combined experience leading the agency, positive change comes from respecting critical ingredients (SBPC and ABC, 2020a). The first of them is treating science as a state policy, not as a government agenda. Continuity is key, as a research system cannot really develop when restricted to four-year terms. In addition to that, evolution must be incremental rather than disruptive and decision making should be based on a thorough understanding of the system. Finally, leading the SNPG must be a democratic effort, where policymakers and all levels of academia work together to implement plans for a better future.

Although the SNPG is still far from perfect, looking at the accomplishments and obstacles of its past, we can better shape the improvements for the future. Hopefully, the policymakers in charge of evolving the system will value such lessons, enabling Brazilian science to move forward.

References

ABE. (1929). *O problema universitário brasileiro: Inquérito promovido pela Secção de Ensino Technico e Superior da Associação Brasileira de Educação*. Rio de Janeiro: Encadernadora.

Balbachevsky, E. (2004). Graduate education: emerging challenges to a successful policy. In C. Brock & S. Schwartzman (Eds.), *The challenges of education in Brazil* (pp. 209–228). Oxford, UK: Triangle Journals.

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

Barata, R. B. (2020). Programas de pós-graduação profissionais: Por que precisamos deles? *International Journal of Business & Marketing*, *5*(2), 30–34.

BRASIL. (1891). *Constituição da República Federativa do Brasil*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1893). *Lei nº 173, de 10 de setembro de 1893. Regulamenta a organização de associações que se fundarem para fins religiosos, Moraes, scientíficos, artísticos, políticos ou de simples recreio, nos termos do art. 72 da Constituição*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1915). *Decreto nº 11.530, de 18 de Março de 1915. Reorganiza o ensino secundario e o superior na Republica*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1920). *Decreto nº 14.343, de 07 de setembro de 1920. Institue a Universidade do Rio de Janeiro*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1931a). *Decreto nº 19.850, de 11 de Abril de 1931. Exposição de motivos*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1931b). *Decreto nº 19.851, de 11 de Abril de 1931. Dispõe que o ensino superior no Brasil obedecerá o sistema universitário*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1931c). *Decreto nº 19.852, de 11 de Abril de 1931. Dispõe sobre a organização da Universidade do Rio de Janeiro*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1951). *Decreto nº 29.741, de 11 de julho de 1951. Institui uma comissão para promover a Campanha Nacional de Aperfeiçoamento de Pessoal de Nível Superior*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1961). *Lei nº 4024, de 20 de dezembro de 1961. Fixa as diretrizes e bases da educação nacional*. Diário Oficial da União. Brasília, DF.

BRASIL. (1974a). *Decreto nº 73.411, de 4 de janeiro de 1974. Institui o Conselho Nacional de Pós-Graduação e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1974b). *Decreto nº 74.299, de 18 de julho de 1974. Dispõe sobre a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1981). *Decreto nº 86.791, de 28 de dezembro de 1981. Extingue o Conselho Nacional de Pós-Graduação e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1988). *Constituição da República Federativa do Brasil*. Senado Federal do Brasil. Brasília, DF.

BRASIL. (1990a). *Lei nº 8028, de 12 de abril de 1990. Dispõe sobre a organização da Presidência da República e dos Ministérios, e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1990b). *Medida Provisória nº 150, de 15 de março de 1990. Dispõe sobre a organização da Presidência da República e dos Ministérios e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1992). *Lei nº 8405, de 9 de janeiro de 1992. Autoriza o Poder Executivo a instituir como fundação pública a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (2014). *Lei nº 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação*. Diário Oficial da União. Brasília, DF.

Brasil, A. (2018). Exploring differences in scientific production originated from professional and academic graduate courses in Brazil. In *Proceedings of the 23rd international conference on science and technology indicators* (pp. 27–40). https://doi.org/10.6084/m9.figshare.9443693.v1

BRASIL. Ministério da Educação. (2017). *Portaria nº 389, de 23 de março de 2017. Dispõe sobre o mestrado e doutorado profissional no âmbito da pós-graduação stricto sensu*. Diário Oficial da União. Brasília, DF.

Cano, W. (2015). Crise e industrialização no Brasil entre 1929 e 1954: A reconstrução do Estado Nacional e a Política Nacional de Desenvolvimento. *Brazilian Journal of Political Economy*, *35*(3), 444–460. https://doi.org/10.1590/0101-31572015v35n03a04

CAPES. (1981). *Plano Nacional de Pós-Graduação - PNPG 1982-1985*. Brasília, DF: CAPES.

CAPES. (1985). *Plano Nacional de Pós-Graduação - PNPG 1986-1989*. Brasília, DF: CAPES.

CAPES. (1995b). *Programa de flexibilização do modelo de pós-graduação senso estrito em nível de mestrado*. Brasília, DF. Retrieved from https://bit.ly/3sp5Aic

CAPES. (1998). *Portaria nº 80, de 16 de dezembro de 1998. Dispõe sobre o reconhecimento dos mestrados profissionais e dá outras providências*. Diário Oficial da União. Brasília, DF.

CAPES. (2004). *Plano Nacional de Pós-Graduação - PNPG 2005-2010*. Brasília, DF: CAPES.

CAPES. (2010). *Plano Nacional de Pós-Graduação - PNPG 2011-2020*. Brasília, DF: CAPES.

CAPES. (2020b). Orçamento da CAPES: Evolução em Reais. Brasília, DF: CAPES. Retrieved from https://bit.ly/37NYz0P

CAPES. (2020c). Portal de Periódicos. Retrieved August 17, 2021, from https://periodicos.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/

Castro, M. H. d. M. (2015). Higher education policies in Brazil: A case of failure in market regulation. In S. Schwartzman, R. Pinheiro, & P. Pillay (Eds.), *Higher education in the brics countries* (pp. 271–289). Dordrecht: Springer.

CGEE. (2016). *Mestres e doutores 2015: Estudos da demografia da base técnico-científica brasileira*. Brasília, DF: CGEE.

Clarivate Analytics. (2019). *Research in Brazil: Funding excellence*. Retrieved from https://bit.ly/3gdf5vY

CNPG. (1974). *Plano Nacional de Pós-Graduação - PNPG 1975-1979*. Brasília, DF: MEC.

CNPq. (2020). Portal de Dados Abertos do CNPq. Retrieved from http://dadosabertos.cnpq.br/

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*. Brasília, DF.

Cunha, L. A. (2007). *A universidade temporã: o ensino superior da colônia à era Vargas* (3rd ed.). São Paulo: Editora UNESP.

da Fonseca, M. L. M. (2015). A institucionalização da pesquisa científica brasileira: Os primeiros anos de atuação do Conselho Nacional de Pesquisas (CNPq). *Parcerias Estratégicas*, *18*(36), 253–268.

de Almeida, E. C. E., & Guimarães, J. A. (2013). Brazil's growing production of scientific articles—how are we doing with review articles and other qualitative indicators? *Scientometrics*, *97*(2), 287–315. https://doi.org/10.1007/s11192-013-0967-y de Almeida, E. C. E., Guimarães, J. A., & Alves, I. T. G. (2010). Dez anos do Portal de Periódicos da Capes: Histórico, evolução e utilização. *Revista Brasileira de Pós-Graduação*, *7*(13), 218–246. https://doi.org/10.21713/2358-2332.2010.v7.194

de Andrade, D., Rizzi, P., & de Almeida, S. F. M. (2005). Experiência em parceria estratégica: mestrado profissional em Engenharia Aeronáutica, ITA-Embraer. *Revista Brasileira de Pós-Graduação*, *2*(4), 48–60. https://doi.org/10.21713/2358-2332.2005.v2.77

de Biasi, K. (2019). *Solving the Change Paradox by means of trust*. https://doi.org/10.1007/978-3-658-23912-1

Dellagostin, O. (2020). Semelhanças e diferenças entre os estados no financiamento à pesquisa e na pós-graduação [webinar]. presentation. Retrieved November 13, 2020, from https://youtu.be/JFBKShfJKDM

Fávero, M. d. L. d. A. (2006). A universidade no Brasil: Das origens à reforma universitária de 1968. *Educar em Revista*, *22*(28), 17–36. https://doi.org/10.1590/s0104-40602006000200003

Fealing, K. H., Lane, J., Marburger III, J. H., & Shipp, S. (Eds.). (2011). *The Science of Science Policy: A Handbook*. Stanford: Stanford University Press.

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

Gershon, R. A. (2013). Innovation failure: A case study analysis of Eastman Kodak and Blockbuster Inc. In A. B. Albarran (Ed.), *Media management and economics research in a transmedia environment* (pp. 46–68). New York: Routledge.

Gouvêa, F. C. F. (2012). A institucionalização da pós-graduação no Brasil: O primeiro decênio da CAPES (1951-1961). *Revista Brasileira de Pós-Graduação*, *9*(17), 373–397. https://doi.org/10.21713/2358-2332.2012.v9.312

Hostins, R. C. L. (2006). Os Planos Nacionais de Pós-graduação (PNPG) e suas repercussões na pós-graduação brasileira. *Perspectiva*, *24*(1), 1–28.

IBGE. (2021). Sistema IBGE de Recuperação Automática. SIDRA. Brasília, DF. Retrieved from https://sidra.ibge.gov.br/

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

London, S. (1996). Understanding change: The dynamics of social transformation. Retrieved July 8, 2021, from https://bit.ly/3k0MFq7

Lord Robbins, L. C. (1963). *The Robbins Report*. Retrieved from https://bit.ly/2WrVdeS

Marburger III, J. H. (2011). Why policy implementation needs a Science of Science policy. In K. H. Fealing, J. Lane, J. H. Marburger III, & S. Shipp (Eds.), *The science of science policy: A handbook* (pp. 31–55). Stanford: Stanford University Press.

Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). *Revista Brasileira de Sociologia*, *6*(13), 9–26. https://doi.org/10.20336/rbs.256

Moacyr, P. (1937). *A instrução e o Império: Subsídios para a história da educação no Brasil*. São Paulo: Companhia Editora Nacional.

Nader, H., Palis, J., Sampaio, F., Garcia, F. et al.. (2015). Carta ao ministro da Educação contra os possíveis cortes no Portal de Periódicos da CAPES. letter. Retrieved August 5, 2020, from http://www.sbpcnet.org.br/site/arquivos/arquivo%5C_467.pdf

Ribeiro, D. (1961). Universidade de Brasília. *Revista Brasileira de Estudos Pedagógicos*, *36*(83), 161–230.

Rothen, J. C. (2018). A universidade brasileira na Reforma Francisco Campos de 1931. *Revista Brasileira de História da Educação*, *8*(17), 141–160.

Rothen, J. C. (2019). O ensino superior e a Nova Gestão Pública: Aproximações do caso brasileiro com o francês. *Revista Eletrônica de Educação*, *13*(3), 970–995. https://doi.org/10.14244/198271993549

Saldaña, P. (2020). Em meio a pandemia, governo Bolsonaro investe contra pesquisa em ciências humanas. Retrieved August 10, 2020, from https://bit.ly/2PHEUH8

Sapolsky, H. M., & Taylor, M. Z. (2011). Politics and the Science of Science Policy. In K. H. Fealing, J. Lane, J. H. Marburger III, & S. Shipp (Eds.), *The science of science policy: A handbook* (pp. 9–22). Stanford: Stanford University Press.

SBPC, & ABC. (2020a). A história da CAPES na voz dos seus ex-presidentes [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

SBPC, & ABC. (2020b). Qual o papel da CAPES na construção da ciência e da pesquisa no Brasil? [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

Schwartzman, S. (1989). A ciência no período de pós-guerra. presentation. Brasília, DF.

Schwartzman, S. (1991). *A space for science: The development of the scientific community in Brazil*. University Park, Pennsylvania: Penn State Press.

Serafim, M. P., & Dias, R. d. B. (2020). A importância da ciência e das universidades públicas na resolução de problemas sociais. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, *25*(1), 1–4. https://doi.org/10.1590/s1414-40772020000100001

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Sucupira, N. (1980). Antecedentes e primórdios da pós-graduação. *Forum Educacional*, *4*(4), 3–18.

Waltman, L., Boekhout, H., & van der Weijden, I. (2019). A large scale bibliometric analysis of differences between male and female researchers: Time trends and career paths. In *Proceedings...* Amsterdam, Netherlands.

3 The dynamics of a national evaluation system

> *The remarkable quality of Brazilian graduate education is due, in significant part, to the evaluation system adopted at the national level.*
>
> **— Robert Verhine**

The Brazilian evaluation of research and graduate education is a five-decade effort driven by a complex array of political, social, and economic factors. This system has undergone significant changes since its creation and, to fully comprehend its current format, it is crucial to dive into the historical context in which it was established, including the configuration of the national science system and the evolving aims of evaluation beyond a quality guarantee mechanism (Brasil, 2020; Sguissardi, 2006).

The role of the state in Brazil has been a defining feature of the country's higher education. As explored in detail in a previous study, its influence was overwhelming until the early 20th century, as the state played an important role in preventing the emergence of universities in the country (Brasil, 2020). Authoritarian marks and overregulation persisted even after the founding of the first universities, often at the expense of institutional autonomy. Graduate education is also part of this complex history. Although its origin in Brazil can be traced back to the Statute of Universities of the 1930s, the system did not develop spontaneously, with very few master and doctorate courses established over the next few decades (BRASIL, 1931b; Fávero, 2006; C. B. Martins, 2018).

This chapter has been submitted for review and it is available as a preprint: Brasil, A. (2023) The dynamics of a national evaluation system: The Brazilian model and its role in developing research and graduate education in the country. SocArXiv. https://doi.org/10.31235/osf.io/wf5kn.

It was only during the 1950s and 1960s that graduate education started to be shaped in Brazil, influenced by converging objectives from two government technocracy sectors: economic and educational. The former realised the great shortage of highly qualified personnel to manage large-scale development projects, while the latter aimed to prepare the necessary personnel for multiple fronts and tasks, including the qualification of teaching staff required by the great expansion of higher education (Rothen, 2018; Sguissardi, 2006).

To address the dimension of human resources, the Brazilian government has enacted several initiatives, including the creation of a campaign to acquire the critical mass necessary to expand high-level education, primarily at the graduate levels. The resulting National System of Research and Graduate Education (SNPG) emerged as an object of state planning, formalised in great part after the Ministry of Education empowered the Federal Education Council (CFE) with the task to regulate it. The resulting document, known today as the Sucupira Report in tribute to its main author, Newton Sucupira, was published in 1965 to define the essence of what graduate education would be like in the country (Balbachevsky and Schwartzman, 2010; CNPG, 1974).

As Brazil's experience with graduate education was still in its infancy, CFE drew on international models to inform the development of the national system. The United States model, influenced by Germanic traditions, played a particularly significant role in shaping Brazilian graduate education. From the inspiration, *stricto sensu* graduate courses, encompassing the master's and doctoral levels, were established as the unit responsible for graduate education, which would be facilitated through scientific research, the promotion of high culture, and the training of scholars and university professors (CFE, 1965; Sucupira, 1980).

One key element of the CFE recommendations was the recognition of the need to regulate the implementation of graduate courses. The council believed that the mere existence of accredited undergraduate courses at a higher education institution (HEI) did not guarantee its competence to implement the graduate level. According to CFE (1965), graduate education was essential for the renewal of the Brazilian university and, without adequate guidelines and regulations, its degradation would be at risk. Consequently, the report recommended establishing more than doctrinal principles, but also operational criteria that direct and control the implementation and development of these courses. The country's national evaluation of research and graduate education derived from this, becoming one of the oldest in the world, predating the global trend of evaluations to become a "growth industry" for governments and public sector organisations in the late 20th century (Leeuw and Furubo, 2008).

This research aims to provide a comprehensive understanding of the Brazilian evaluation system departing from its conception, thus contributing to the ongoing academic discourse on the role of evaluation in research systems. To achieve this, primary sources such as original legislation, policy documents, archival interviews, and related literature were examined. The study specifically aims to (i) understand the origins and driving forces behind the implementation of a national evaluation system; (ii) trace its evolution with a focus on pivotal moments and decisions that have driven its development over several decades; (iii) contextualise the current evaluation model, also while examining some outcomes of the most recent evaluation cycle.

3.1 The origins of a National Evaluation System

Leeuw and Furubo (2008) delineate the concept of “evaluation systems”, positing four key criteria to characterise them. The first criterion involves having a distinctive epistemological perspective that fosters consensus among stakeholders on what evaluative activity entails. The second criterion requires that evaluations be performed by evaluators within organisational structures and institutions, as opposed to individual evaluators. The third criterion involves a certain permanence or history of the activities involved, suggesting that the evaluation process is ongoing. The fourth and final criterion requires a focus on the intended use of evaluations, which are planned in advance to be delivered to decision makers during a specific phase of decision making. Evaluation results may serve as the main determinant of the decision, as is the case in accreditation processes.

Taking into account the criteria proposed by Leeuw and Furubo (2008), Brazilian evaluation took approximately a decade after its initial conception to evolve into a system. In 1965, the Sucupira Report established that the accreditation of graduate courses in Brazil would fall under the purview of the Federal Education Council (CFE), and diplomas would possess legal value only if issued by courses accredited with the Ministry of Education (CFE, 1965).

However, the first attempts by the CFE to fulfil this role were hindered by the absence of suitable mechanisms and procedures, and accreditation was only carried out *a posteriori* (Balbachevsky, 2004). To better address the continuous implementation of graduate education, a National Council for Graduate Education (CNPG) was established within the Ministry of Education and Culture structure in 1974 (BRASIL, 1974a). The council had two primary objectives: to formulate a National Plan for Graduate Education (PNPG) and to propose measures to execute and update the resulting plan regularly.

Based on information from CNPG (1974), it is estimated that approximately 7000 students enroled in graduate courses in 1973, increasing the total number to approximately 13.500. Most of the students attended public institutions, with 5000 in federal HEI and 5800 in state and municipal ones. Approximately 7500 professors participated in graduate activities, working across 50 institutions, which included 25 federal, 10 state and municipal, and 15 private HEI. The graduate system had experienced considerable growth, as only around 3500 master's and 500 doctoral degrees had been granted in the country up to that year, with approximately 50% of graduates opting to remain in academia.

Recognising the new dimensions and anticipated growth of the SNPG, it was acknowledged that the country lacked experience in evaluating research and education at the graduate level. Although a National Council for Scientific and Technological Development (CNPq) was established in 1951, the accrued experience focused on assessing individual projects rather than entire research units. The proposed solution involved assigning the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) with the task of organising the evaluation of graduate courses and designing future PNPG (CNPG, 1974).

CAPES was founded in 1951 as a government campaign to create enough critical mass to implement graduate education in Brazil (BRASIL, 1951). The campaign was eventually converted into a coordination and was restructured in 1974 to obtain administrative and financial autonomy to carry out the national evaluation of graduate education (BRASIL, 1974b). The agency then conducted its first evaluation attempts from 1976. These were sporadic and relatively informal, mainly intended for internal use (Verhine and Dantas, 2009).

During the shaping of the first steps of the evaluation, CAPES even considered delegating the process to scientific associations. For instance, Verhine (2008) presents the case of the National Association of Research and Graduate Studies on Education (ANPEd), which was formally established in 1978. At that time, CAPES asked the association to formulate an evaluation model for the area of education, one that other associations could subsequently implement to assess their courses. A proposal was prepared and discussed at an ANPEd meeting in 1979, but both the model and the idea of an association-led evaluation were rejected under the argument that a scientific association should be free to promote academic exchanges and advocate for their member courses.

With that, CAPES embraced the evaluation of graduate education as one of its main roles, established its evaluation directorate, and allocated a portion of its budget to evaluative activities, with the system being implemented in 1980 as an annual exercise (Verhine and Dantas, 2009; Verhine et al., 2021). Such activities would include accreditation and continued evaluation, but this study will focus on the second, as both have rich histories and complex developments. Although accreditation became an essential step for a graduate course to enter the SNPG, continued evaluation determines its permanence through mandatory accreditation renewal and also impacts funding, reputation, and more.

3.2 Shaping evaluation

The evaluation carried out by CAPES was conceived from two main perspectives, one from the National Council for Graduate Education (CNPG) and the other from the agency itself. The CNPG perspective was one of quality, as the council considered it essential to overcome the low overall performance of the graduate courses in the country. To this end, the CNPG proposed comprehensive support measures to improve standards, including financial resources, advisory services, information dissemination, and quality recognition mechanisms (CNPG, 1974).

CAPES faced a more practical challenge: the growth of the SNPG and an increasing number of scholarships granted. In response, the evaluation system was designed with graduate courses as the primary assessment unit. The quality of their work and productivity would be evaluated through a combination of peer review and information collected from institutions. Assessment was based on various criteria, including faculty qualifications, research output, and infrastructure. The outcome was an internal ranking of the courses, with grades ranging from *A* to *E* (with *A* being the highest). The application of the results came through mechanisms linking the aid to a good grade in the evaluation. This translated into increased funding and higher scholarship quotas for courses and additional benefits for HEI with stronger graduate education, as the results of the evaluation soon influenced budgetary discussions at federal universities (Castro and Soares, 1983; Schwartzman, 1982; Verhine and Dantas, 2009).

The evaluation system's approach yielded significant advantages, enabling the efficient management of thousands of scholarships while consuming fewer resources. In the early 1980s, around 8000 scholarships were allocated to existing courses, with only two CAPES employees overseeing the evaluation process and an additional four or five dedicated to constructing an information archive. In contrast, the management of over a thousand international scholarships required the efforts of approximately 20 agency employees. However, beyond practical benefits, many academics regarded the new evaluation system as a vital instrument for quality assurance, and the media played a crucial role in reinforcing this perception (C. d. M. Castro, 2006; Castro and Soares, 1983).

According to Castro and Soares (1983) and Verhine (2008), the results of the first evaluations, conducted annually, were confidential, used only to guide funding and support decisions within CAPES. Gradually, some results were published to recognise the performance of the institutions, allowing the public to access a list of courses evaluated as *A* or *B*. Some HEI began to publicise all of their results, mixing transparency with a marketing strategy, where positive evaluations were seen as a CAPES seal of approval. However, in 1982, the list of the 56 courses graded *E* was leaked to the media, sparking considerable public debate. The directors of many of these courses, particularly from Brazil's largest HEI, the University of São Paulo (USP), attempted to discredit the evaluation. Nevertheless, USP's pro-rector at the time countered that if they received a poor grade, it was because they were underperforming. Castro and Soares (1983) consider this controversy to be the baptism of fire for evaluation.

Another form of criticism emerged at the time, ultimately reinforcing the value of the new evaluation system. A former CAPES director recalled an encounter with a high-ranking bureaucrat from the Ministry of Education who criticised the changes in the agency from the introduction of the evaluation. The bureaucrat recounted that a renowned appeals court judge had requested a scholarship for his son, only to be surprised when it was denied (C. d. M. Castro, 2006). In the absence of an evaluation system, resources were typically allocated based on tradition or political influence. With the evaluation system, CAPES could objectively distribute resources and allocate scholarships to courses with the greatest potential to advance knowledge and innovation (Castro and Soares, 1983; Schwartzman, 1982).

Although objectivity in the evaluation was considered a gain, the downside was that CAPES was never an agency solely dedicated to assessment, but instead became a hybrid institution with potentially contradictory roles to fund and evaluate at the same time (Verhine, 2008).

In the second Brazilian National Plan for Graduate Education (PNPG), this time produced by CAPES, significant emphasis was placed on the need for a critical evaluation of the quality of graduate courses in terms of intellectual production and human resource development. Although acknowledging that evaluation practises required broader institutionalisation within the country's academic community, the plan also highlighted the growing adoption of evaluation by Science & Technology government agencies, also recognising the need for its constant evolution (CAPES, 1981).

3.3 Evolving foundations and the second model

One of the strengths of Brazilian evaluation is that it has never been static. From its conception, it has gone through incremental evolution at every cycle, for instance, with the replacement of the *A* to *E* grading system to a *1* to *5* scale (Barata, 2016), or the review in its periodicity, which changed from annual to biannual in 1984, since one year proved to be too short of an interval for courses to present any radical differences in their performance to justify reevaluation (Castro and Soares, 1983; Viana, 2018).

In the first 20 years of evaluation in Brazil, significant progress was made in developing the system and aligning it with the intended goals. However, the implementation of the evaluation proved to be a significantly difficult task, as recognised by some of the system's proponents (Sucupira, 1980), architects (C. d. M. Castro, 2006; Castro and Soares, 1983), and experts (Balbachevsky, 2005; Schwartzman, 1982; Sguissardi, 2006). Although not every problem could be solved, these experiences helped to form the core principles that would guide the evaluation in Brazil over the next decades.

3.3.1 Core principles of the evaluation

Schwartzman (1982) states that the evaluation implemented in Brazil should not be perceived as a simple rating system carried out by impartial and independent judges, but rather as part of a process through which the academic community gradually explicates its criteria and establishes quality standards. Although a grading system is among the main results of the evaluation, there are no explicit rules connecting specific concepts with particular information about the assessed graduate courses. Evaluators rely on all available information, including data from CAPES data collection, independent sources, or informal exchanges between committee members (Castro and Soares, 1983).

That has not always been the case. The first attempts at evaluating graduate courses sought to develop systems of objective indicators that could be quantified. The advantage of such a system, if functional, would be its freedom from subjectivity of the evaluator. Classic indicators were adopted for this purpose, including scientific publications, approved theses, faculty qualifications, etc. However, it soon became apparent that the outcome of these measurements often contradicted the consensus opinions of experts in various disciplines on the actual quality of courses. Furthermore, subjectivity was only seemingly eliminated since it remained present in the selection of the supposed quality indicators and the weights attributed to them (Schwartzman, 1982).

In that sense, the CAPES evaluation system aims to bring the peer review approach as an additional dimension to the quantitative one. CAPES officials gather data from graduate courses, and experts working in disciplinary committees analyse them. This means that the human element plays a role in contextualising quantitative indicators, while a reliable and well-presented factual foundation will help reduce the so-called "halo effect". This is a phenomenon in which evaluators could fill gaps in data with biased preconceptions about institutional reputation, course longevity, geographic location, and other dimensions that lead to assumptions about what may be good or bad in an assessed unit (Castro and Soares, 1983).

Furthermore, Schwartzman (1982) believes that evaluators can incur two additional types of errors in their judgment during peer review activities. The first involves a discrepancy in the evaluations of various evaluators, meaning that the human dimension of an evaluation process may influence the potential output from the same types of input. The second error would involve being influenced by unclear or debatable criteria. Even if there are no explicit rules linking the evaluation results with the objective characteristics of a course, there may be implicit rules, even unconscious. In the heart of these concerns, it is possible to find a direct relation to the Matthew effect introduced by Merton (1968), who described a phenomenon in which prominent scientists receive disproportionate recognition for their work, while less established scientists are denied such recognition. This effect, which often leads to a "rich gets richer" scenario, is a risk for graduate courses in the Brazilian evaluation system.

Thus, a core principle of Brazilian evaluation is that there should be a dynamic equilibrium between data, peer review, and criteria. While peer evaluation must address the inadequacy of quantification, data collection must leave little room for assumptions, and systematic information should be available to confirm or rebut any initial judgments. As an underlying and structural element of this delicate balance, well-thought criteria should guide data collection and peer review, making the process easier to replicate, being adopted by different evaluators, and being comparable across disciplines.

From a data collection perspective, CAPES employed its first instrument for that purpose in 1977, when all graduate courses were asked to submit an annual data report to the agency. At first, there were significant issues regarding the quality of responses and compliance, but as soon as the connection between evaluation and funding became clear to the entire SNPG, a growing fraction of master's and doctoral courses began to respond (Castro and Soares, 1983). Steps were also taken to improve data quality, including personnel training and refinement of data collection instruments, which evolved from paper-based to computer-generated forms in the early 1980s (CAPES, 1981).

For the peer review part, the CAPES evaluation model has always relied on renowned members of the academic community for expertise and legitimacy (Verhine and Dantas, 2009). Evaluators are invited to integrate committees or what CAPES calls evaluation areas. These areas serve disciplinary purposes, seen in committees such as Sociology, Chemistry, Biotechnology, but also organisational ones, which are apparent in the subdivision of medical graduate courses in Medicine I, II, and III. Each of the areas, of which there are 49 as of 2023, has a coordinator chosen by peers and CAPES for a term of equal duration as an evaluation cycle (Brasil, 2023b).

Regarding the evaluation criteria, it was a challenge to build a comparable process for all areas in the first decades of the system. Areas established their own criteria, dynamically, without institutional instruments of support. Some attempt to connect the work of the various areas was exercised by the Technical Advisory Group (GTC), a small committee of representatives from the country's main government funding agencies. Although the group would play a role in approving the evaluations performed by the area committees, its purpose was not to criticise the evaluation but to familiarise the various agencies with the results derived from it. Consequently, approval was only regimental, something that would change significantly in the 1998 evaluation reform (Verhine, 2008).

3.3.2 The second model

Balbachevsky and Schwartzman (2010) argue that, despite its strengths, the CAPES evaluation model faced challenges that became more evident over time. The small size of the Brazilian scientific community, combined with the visibility of the committees' work, led to unavoidable local pressures, resulting in grade inflation. For instance, around 80% of the courses evaluated in 1996 received the two highest grades, revealing that the original scale and the adopted criteria had lost some of their discriminatory power (Verhine et al., 2021).

In response to the issue, CAPES implemented a new evaluation model in 1998, with improvements inspired by the experience of previous decades (Balbachevsky and Schwartzman, 2010). An important element in the reform addressed the need for adequate criteria in the evaluation, in part aiming at greater standardisation and comparability in the system. The proposed solution was the creation of a single assessment form, made up of predetermined dimensions and subdimensions to guide evaluation in all areas (Verhine, 2008).

Furthermore, the role played by the GTC in the approval of evaluation results now belonged to CAPES' Technical & Scientific Council (CTC). With a composition based on agency and evaluation area representatives, this council went beyond ratification of the results to conduct a thorough review of the evaluations by area committees (Horta and Moraes, 2005). With control over the final grades of the courses and the comparability across different areas, the CTC became the main decision-making body responsible for regulating and coordinating the entire evaluation process (Verhine and Dantas, 2009).

Another relevant change in the evaluation framework established the graduate program (PPG) as the fundamental unit of analysis, rather than individual master's and doctoral courses (Horta and Moraes, 2005). The change addressed a typical situation derived from the growth of the SNPG: institutions that started with master's courses eventually expanded the offer with doctoral courses. A graduate program could include a master's, a doctorate, or both, so the new unit allowed for a more comprehensive evaluation than assessing each course individually (Balbachevsky and Schwartzman, 2010).

Additional changes introduced in the 1998 evaluation reform were the expansion of the evaluation cycle from two to three years (Verhine, 2008), and the expansion of the five-level grading scale (*1 – 5*) to include two additional ones, *6* and *7*. These were reserved for PPG considered to be of excellence by international standards. Furthermore, the grade *3* was established as the lowest acceptable rank in an evaluation, and any performance below this led to the closure of the graduate program (Balbachevsky and Schwartzman, 2010; Verhine et al., 2021). Finally, 1998 also saw the introduction of Qualis, a model for classification of scientific production from PPG by using journals as a proxy (Verhine and Dantas, 2009).

The history and evolution of Qualis have been explored in length by Brasil (2023a). However, it is relevant to highlight that the system arguably led to an increased emphasis by CAPES on schollarly publishing in program evaluations. This shift promoted behaviours considered productivist by faculty and students, despite being an undeniable improvement over the mere quantification adopted at that time (Verhine, 2008). Qualis also marked a shift in evaluative emphasis towards the products of research, particularly qualified bibliographic production. In fact, the message was clear to the academic community, as grades *6* and *7* of the following evaluation would be defined primarily based on a single parameter: international scientific production (Horta and Moraes, 2005).

The evaluation of the first triennium under the new model (1998-2000) was conducted in 2001, and it was organised in four stages: evaluation by the area committees; review and ratification by the CTC; a second round of evaluation, by the area committees, for graduate programs that requested results to be reconsidered; review and ratification of the appealed results by the CTC. The 2001 evaluation assessed 1545 programs, of which less than 10% were classified in the top two grades, and 23% reached grade *5* (Horta and Moraes, 2005).

3.4 The established evaluation system

Despite significant advances in the first two decades, the academic community continued to advocate for further refinement of the Brazilian evaluation (Verhine and Dantas, 2009). External pressure combined with continuous self-diagnosis by CAPES and the evaluation community shaped the dynamics and evolving nature of evaluation, leading to consecutive advances throughout the years. However, the essence of the evaluation still current in Brazil was already established and most of the changes seen in the past decades have been incremental, either improving or broadening the instruments already in place.

The assessment form is one of the instruments that has undergone various changes since its introduction, with dimensions and subdimensions adjusted to steer the evaluation according to relevant priorities. In 2006, for example, the form was revised to reduce its complexity and amplify its emphasis on appraising the value of outputs, rather than concentrating on inputs or processes. The result was the unification of seven dimensions into four: "PPG proposal", "faculty", "student body", and "intellectual production". Shortly thereafter, a fifth dimension named "social inclusion" was added to the form, generating significant controversy at the time. Although some evaluation areas considered the change crucial to induce societal impact, not all areas supported the initiative. As a result, it was decided that the new dimension would be weighted as only 10% of the final grade, limiting its efficacy (Monteiro et al., 2019; Verhine and Dantas, 2009).

After the 2007 evaluation, the form was once again revised. While the five dimensions remained the same, their weight did not. From the 2010 evaluation onwards, "student body" and "intellectual production" would account for 70% of a PPG's final grade, undoubtly steering graduate programs to value more indicators such as the number of PhD graduates, average time of degree, and number of papers published. The weight of "social inclusion" remained unchanged. (Monteiro et al., 2019)

Another relevant change regarding the assessment form was that a separate model was created in 2007 for professional master's courses. The new modality of PPG was implemented in 1998 as an alternative to academic courses, but with the objective of bringing scientific advancement of knowledge into practise, having social impact as a priority (Brasil, 2020; Ferreira and Moreira, 2002).

Although the differences between the forms were minor, the change was a crucial step in diversifying the evaluation to value different models of graduate education (Verhine, 2008). Over time, separate committees for the evaluation of professional programs were established for areas with many PPG in that modality, and the 2017 evaluation would even allow nonacademics to integrate the groups (CAPES, 2017b). Even though there were few of these committee members in recent evaluations, an initiative that opens space for a judgment grounded in professional experience may help overcome an excessive reliance on standardised performance indicators, which do not capture the essence of the applied research conducted in these PPG (Muller, 2018).

The Qualis system also went through changes, including adjustments in classification rules and the scale adopted to rank the journals, with details covered at length in Brasil (2023a). However, it is relevant to mention that a Qualis-Arts, dedicated to assess artistic production, was developed by the Arts evaluation area and implemented with CAPES support in 2005 (Ulhôa, 2017). The agency also encouraged evaluation areas to develop a Qualis-Books, in 2008. The process was designed to allow distinct committees to establish their own criteria to classify this type of research output, with a positive impact in many areas, especially those of the social sciences and humanities (Verhine and Dantas, 2009). Around the same time, a Qualis-Events would also be developed to better value conference proceedings, with the classification being of special value for the area of computer science.

Furthermore, CAPES also made significant investments to improve the annual data collection from PPG, which is not a small challenge considering the size and growth of the Brazilian National System of Graduate Education (SNPG), as reported by (Brasil, 2020). The computer generated form completed by PPG in the early 1980s was supplanted by a specialized data collection system named DataCapes in 1987, followed by different iterations of a new and improved system – Coleta Capes – used from 1996 to 2013 (R. J. Ribeiro, 2008). All these efforts led to the creation of the Sucupira Platform. Launched in 2013, the integrated Current Research Information System (CRIS) not only accepts continuous data submissions, but also allows the general public to access most of the data without restrictions (Siqueira, 2019). Consequently, data collection is kept open and the graduate programs under evaluation can verify their data. This improves transparency in the system according to the relevant principles of the Leiden Manifesto for research metrics (Hicks et al., 2015).

Following a new change in the periodicity of evaluation to a four-year cycle (CAPES, 2014), Morato (2015) reports how the Leiden Manifesto became a significant influence on Brazilian evaluation, sparking extensive debates over the existing model. The process culminated in the establishment of twelve working groups tasked with performing diagnostic analyses and recommending strategic initiatives to help CAPES to effectively carry out its mission within the National System of Graduate Education (SNPG). Each group comprised area coordinators, institutional representatives, and CAPES professionals, in addition to subject matter experts. The working groups covered a wide range of topics, investigating issues such as the CAPES evaluation system, different aspects of the Qualis classification (including journals, books, events, and a new proposal for technical and technological production), professional master's, knowledge area taxonomy, impact evaluation, risk analysis, and information systems.

The debate would inspire a long-term process to restructure evaluation in Brazil, including the appointment of a special committee by the Ministry of Education. The group presented various recommendations to revise the model, touching on several crucial aspects of the evaluation process (Oliva et al., 2017). These proposals can be categorised into four primary areas:

i) Streamlining the process: This involves simplifying aspects of the evaluation, using subcommittees, extending even more the interval between evaluations, simplifying the assessment form and report (data collection), and revising the ranking system (from *1–7* to a simplified version);

ii) Broadening focus on training: Recommendations involve assessing the quality of the training provided by PPG, and incorporating and emphasising self-assessment.

iii) Acknowledging contextual diversity: Proposals suggest separating regulation from evaluation, introducing a range in the scale to value courses consolidated at the regional level, and calculating the relation between expected and observed results.

iv) Valuing impact: Suggestions here encompass integrating criteria *4* and *5* in the assessment form, the design of a system to track alumni, and the connection of the CAPES databases with other national databases.

The proposed changes reflect the desire for a more holistic, context-sensitive, and impact-driven evaluation model. However, these suggestions and the findings from the CAPES working groups emerged around the same time as the first quadrennial evaluation in 2017. Consequently, only incremental adjustments could be implemented then, and transformative changes would need to wait.

The 2017 evaluation marked a step towards the evolution objective. Once it was concluded, various diagnostic reports emerged to provide more elements to continue the search for a better evaluation model. Different groups prepared these reports, acknowledging the progress made so far in the evaluation model and identifying necessary changes. Among them was the CTC-ES,[1] that consolidated the experiences of the council and area coordinators in a dedicated report (Faljoni-Alario et al., 2018). Another report came from PNPG Committee (2018), which produced a crucial document in consultation with more than a dozen entities representing diverse perspectives in the academic and scientific community. These included the Brazilian Academy of Sciences (ABC), National Education Council (CNE), National Council of State Funding Agencies (CONFAP), National Institute of Educational Studies and Research (INEP), National Forum of Pro-Rectors for Research and Graduate Education (FOPROP), and Brazilian Society for the Advancement of Science (SBPC).

The years 2018 and 2019 were very active for the development of Brazilian evaluation, with a distinct focus on substantial reforms and long-term planning. An international seminar series was held to discuss the future of evaluation (CAPES, 2018a), and a new generation of working groups was established, tasked with extending the insights gained from their predecessors. They sought to learn from the evaluation conducted in 2017 and build on the contributions that had emerged from it. However, the discourse broadened to address other themes, with specialised groups created to focus on issues such as internationalisation, the promotion of innovation, and knowledge transfer, as well as self-assessment mechanisms. The intended plan was for the 2021 evaluation, designed to assess the performance of graduate courses from 2017 to 2020, to introduce the first of these changes. This was seen as a stepping stone towards more substantial transformations in the future cycles (PNPG Committee, 2020).

For instance, PNPG Committee (2018), Faljoni-Alario et al. (2018), FOPROP (2018), and Oliva et al. (2017) were some of the many reports that recom-

[1] The Technical & Scientific Council (CTC) was renamed to Technical & Scientific Council for Higher Education (CTC-ES) in 2007, as CAPES aquired a role supporting initial and continued training of basic education (EB) teachers in Brazil. That led to the creation of a second CTC that would be known as CTC-EB (BRASIL, 2007).

mended the adoption of a self-assessment strategy to expand institutional autonomy and foster a more comprehensive and less uniform evaluation. The report of the responsible working group outlined strategies to establish such a system and promote a more multidimensional evaluation process (Verhine et al., 2019). The predicted implementation for the upcoming evaluation of 2021 was modest, reflected in a review of the assessment form that included self-assessment as a subdimension with little consequence on the final grade attributed to the graduate programs (Brasil, 2022). Despite the limited effect at first, the experience of the 2021 evaluation would be valuable to rethink strategies for a greater impact in the future. However, the plans were abruptly disrupted by the advent of a crisis.

3.5 Recent challenges and the future of evaluation

There was much potential for transformative changes in the Brazilian evaluation system that emerged from discussions around the Leiden Manifesto and that was further developed by the collective effort observed before and after the 2017 quadrennial evaluation. However, CAPES has faced an unprecedented crisis, unfolding between 2020 and 2022, which has caused significant turmoil in the Brazilian academic landscape. The influences that underlie this crisis include the global COVID-19 pandemic and substantial turnover in CAPES' higher management – three presidents and multiple directors were appointed and replaced during the Bolsonaro administration. Some of the changes ignited public protests from major scientific entities nationwide (SBPC, 2021a; SBPC, 2021b).

In 2021 a critical juncture was reached with the abrupt dissolution of the Technical & Scientific Council (CTC-ES). This action, interpreted by the academic community as a deliberate attempt to dismantle CAPES, provoked a widespread outcry. However, from a legal perspective, the dissolution was defensible; The council constitution had been technically illegal due to minor alterations made to improve representation in various evaluation areas, yet the supporting regulations had not been updated concurrently (Vasconcellos, 2021). Although it was crucial to rectify the situation to prevent the council's decisions from being invalidated, the handling of this process amplified existing animosities, further fuelling the crisis (Verhine and Souza, 2021).

However, the situation soon became more intricate when, in September 2021, a lawsuit challenged the ongoing reforms and enhancements to the evaluation procedures. The main dispute centred on the idea that the evaluation process compromised the legitimate expectations of the entities being assessed. In essence, the court believed that the evaluation metrics should have been disclosed in advance, with predetermined cutoff values, ensuring that a graduate program (PPG) would receive the expected outcome if it met the standards set by the defined metric. Based on that, the judge in charge issued an injunction, halting CAPES's ongoing evaluation work (Justiça Federal, 2021). In response, CAPES issued a comprehensive legal defence to emphasise its formative and comparative evaluation model. Despite this, the federal judge upheld the initial injunction, pushing the case to a second instance of the Federal Court for further deliberation (Verhine and Souza, 2021).

This legal imbroglio left numerous committee members and coordinators feeling unable to perform their duties, resulting in a wave of resignations often justified by the way CAPES' high management was dealing with the crisis, and also leading to continuous protests from the academic community (Davidovich and Ribeiro, 2021; Monteiro et al., 2021). Eventually, the second instance court authorised the continuity of the evaluation, but required that the results should be kept confidential (Veiga, 2021). The decision created a climate of uncertainty that led to resignations in other areas. In less than ten days, the evaluation director also resigned (Saldaña, 2021b).

Although the judiciary stoppage only lasted ten weeks and some committees were able to resume their work from December 2021, developments such as the wave of resignations delayed the work in other areas, and the evaluation was completed a full year late. Furthermore, the results for the first two phases of the evaluation – with respect to the assessment by the committees and the ratifications by the reinstated CTC-ES – were not made public, but only released to each graduate program so that they could request reconsideration if desired (CAPES, 2022a; Veiga, 2021). Only after the second round with the committees and the CTC-ES were the results made public, in December 2022 (CAPES, 2022b). The Qualis classification, highly influential in PPG grades, was released only in January of the following year, together with a technical document detailing the procedures (CAPES, 2023b; CAPES, 2023c). The results were also met with widespread protests due to a perceived lack of transparency in the evaluation process (Brigatti, 2023; Ferrari, 2023; Yamashita, 2023).

In the midst of this crisis, several entities have publicly acknowledged the role CAPES and its evaluation have always played in the advancement of science, technology and innovation in Brazil (Nader and Davidovich, 2021). But despite the extensive support evaluation received from academia, the historical transparency of the process has been significantly curtailed, due to ongoing judicial constraints mandating the confidentiality of evaluation outcomes. Furthermore, the historical methodology of the evaluation process has been inherently comparative, involving a comprehensive review of the collective performance of all graduate programs within each area to construct benchmarks reflecting the current landscape of scientific advancements. Subsequently, these benchmarks serve as grading parameters, providing a relative measure of performance within the dynamic field of academic achievement. According to current legal directives, the evaluation process is being pushed towards the establishment of *a priori* parameters for its indicators. By providing PPG with specific metrics to pursue, they risk becoming mere targets according to Goodhart's law, thus reducing their value as effective measures (Elton, 2004). If they could achieve the numerical targets associated with a certain grade, they would be assured of receiving it. However, this approach serves to exacerbate the tyranny of metrics, intensifying the undue emphasis on quantifiable outcomes (Muller, 2018).

Furthermore, Veiga (2021) reports how the *a priori* determination of parameters can also represent a risk. For example, predetermined parameters would have been ill equipped to foresee the advent of a pandemic within an evaluation cycle, as much as it could not predict how extensive the funding cuts in Science & Technology would be under Bolsonaro's government, both of which invariably impact the majority of graduate program indicators. Thus, it is crucial to maintain the principle of comparison and parameter definition emanating from the evaluation areas themselves. This bottom-up approach is better suited to reflect the intrinsic dynamism and unpredictability of scientific research.

The turmoil affecting the evaluation is one of many events that have disrupted the SNPG. The budget allocated for scientific research has seen a dramatic reduction of 60% between 2014 and 2022. Scholarships for master's and doctoral programs have not received adjustments for 9 years, resulting in a real-term loss of 66,6% when adjusted for inflation (Dourado, 2023). Therefore, the fact that the evaluation for the 2017-2020 cycle took place at all can be seen as a triumph, given the serious doubts that surrounded its feasibility (Hanzen, 2021). Looking towards 2023 and beyond, the situation appears to be improving, as evidenced by a 44% increase in the budgets for CAPES and CNPq, and a 40% adjustment in master's and doctoral scholarships (CAPES, 2023a). However, if the evaluation process remains inflexible, its directive power will wane, resulting in adverse effects on the scientific system, as highlighted by the president and vice-president of the Brazilian Academy of Sciences (ABC):

> The judiciary should understand the importance of this evaluation and its public disclosure for the development of graduate education and science in the country. CAPES' evaluation, in addition to being a national patrimony, is the great basis for ensuring the sustainability of our excellence in teaching and the training of professionals. It is what points the way and the destination of graduate education, promoting the advancement of science, technology, and innovation in Brazil in all areas of knowledge. A secretive evaluation does not serve the country. And interfering in CAPES is, therefore, interfering in the future of Brazil (Nader and Davidovich, 2021).

3.6 The impact of evaluation advancements

The Brazilian evaluation system has successfully navigated its most significant crisis yet, avoiding the unprecedented risk of not being conducted for the first time since its conception. However, this achievement did not come without repercussions that go way beyond the one-year delay in the publication of results and the limited transparency of the process. Unexpectedly, the judiciary now has a say in a process that used to be exclusively managed by academics and experts who brought almost half a century of institutional experience to the table. This shift could have implications for the postponed reform of the evaluation system.

However, Brazil's evaluation system has come a long way since it was first established. Investigating the impact of these advancements may provide valuable insights for the planning of future reforms, especially now that there is an opportunity to review them before they are put into action. So, this study turns to the results of the 2022 quadrennial evaluation, which relied on 1808 evaluators to assess the performance of 4512 graduate programs. The grade distribution can be seen in Table 3.1 (CAPES, 2022a).

Table 3.1.: Grade distribution of the quadrennial evaluation of 2022

	Final Grade							
	1	2	3	4	5	6	7	Total
Graduate Programs	14	31	980	1786	1030	410	261	**4512**

To examine how the grade distribution observed in Table 3.1 relates to previous evaluations, Figure 3.1 depicts the evolution of the grade in five consecutive cycles that covered the period from 2007 to 2022. The height of each bar represents the number of graduate programs that fall into each grade. This includes PPG accredited between evaluations (identified as "New") and those that have been discontinued due to substandard performance (labelled as "Out").

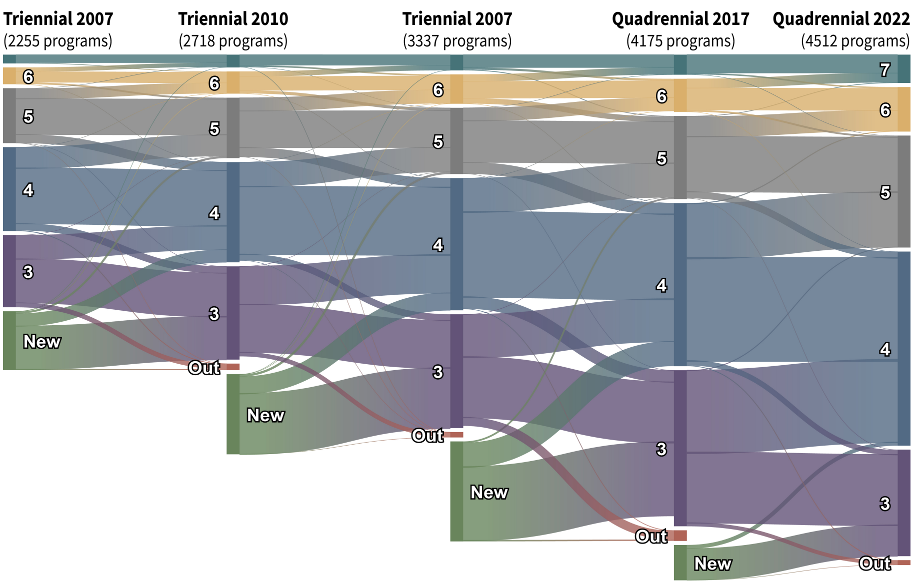


Figure 3.1.: Flow of graduate program grades over the 2007–2022 national evaluations

The visualisation presented in Figure 3.1 provides a dynamic depiction of the evaluation landscape, capturing the flux of grades alongside the PPG accreditation and discontinuation. Observing the thickness of the lines flowing from one evaluation to the next, it is possible to notice that the majority of programs maintain their existing grades, illustrating a degree of stability in the system.

The second most frequent pathway involves a single-level grade increase, indicating the ability of certain programs to improve over time. The third most common occurrence is a single-level grade demotion. Promotions or demotions spanning more than one grade level are comparatively rare.

When comparing the flow between the 2017 and 2022 evaluations with the previous ones, it becomes clear that the most recent cycle has been one of tolerance. The percentage of PPGs discontinued is the smallest in the covered period, with only 45 programs subject to closure for being graded *1* or *2*. In the 2017 quadrennial, 98 out of the 4175 PPG were in the same situation. The same pattern can be seen in all demotions, with most absolute numbers below what has been observed in all previous evaluations.[2] These results are expected, as the 2017–2020 period was affected by the COVID-19 pandemic, funding cuts and the evaluation crisis. However, it is evident that tolerance also affected grade promotions, with a high number of PPG moving to the next level, as in the case of the 751 programs promoted from grade *3* to *4*. Although accreditation is not the focus of the present study, Figure 3.1 also reveals a significant drop in the number of new PPG accredited during the last cycle, possibly due to the same reasons mentioned above.

Having presented the distribution of grades and the general perception of the flow of PPG between evaluations, it is not the intention of this study to explore all possible connections around the quadrennial results, since CAPES has included an extensive dashboard together with the published results (CAPES, 2022a). The objective in this section is to provide an initial exploration of how the diverse dimensions of evaluation influence the assigned grades, and specifically to ascertain whether this influence has evolved. To this end, we refer to an insightful analysis conducted by Schwartzman (1982), who investigated the results of the first evaluations.

According to Schwartzman (1982), in the early 1980s the variable with the highest correlation with the CAPES evaluation was the age of the PPG. The second variable of importance relates to the legal status of the institution where the program is located, more specifically, whether they are public or private. The third variable is the volume of scientific production of the PPG, while the fourth is related to its geographical location. Interestingly, the scholar manifested

[2] All visualisations included in this chapter are available in interactive format at https://andrebrasil.github.io/viz/eval.html. For Figure 3.1, users can interact with all connections to see the exact number of PPG flowing between them.

concern with the results found, as the observed dimensions do not necessarily reflect quality, focussing mostly on the context of the PPG evaluated.

In this study, the variables initially identified by Schwartzman (1982) are mapped in the context of the evaluation results of 2022, with the objective of discerning whether the numerous reforms implemented in the evaluation model have altered the relations observed decades ago. The initial analysis focusses on the age of the PPG at the time of evaluation, and a density graph illustrating the distribution of programs relative to the grade they received is shown in Figure 3.2.

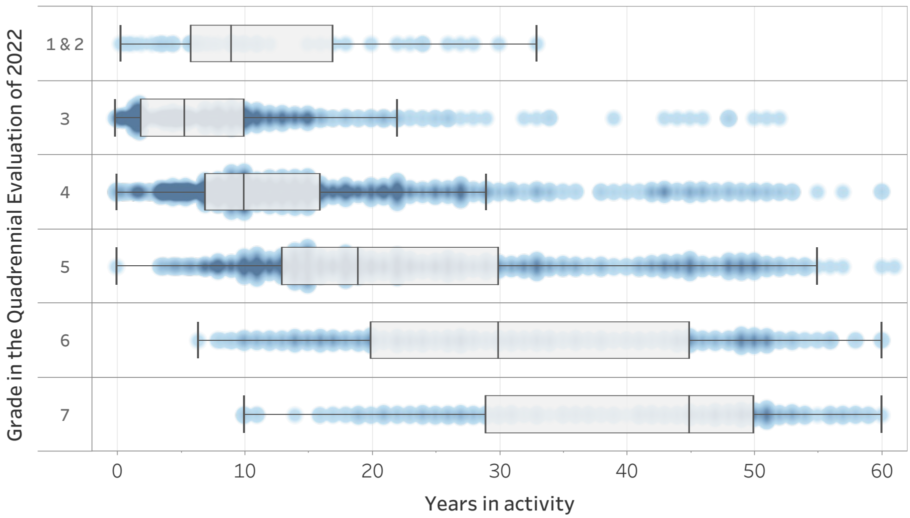


Figure 3.2.: Relation between national evaluation grade (2022) and age of PPG

Figure 3.2 reveals there is a strong link between the age of the PPG and its evaluation results, although not a determinant one. While there appears to be little relationship between PPG age and poor performance that leads to discontinuation (grades *1* & *2*), it seems that maturity is a relevant element to allow programs to obtain better grades. The median age of a grade *6* PPG is close to 30 years, reaching almost 45 for grade *7*. However, that does not mean that a program could not reach the highest grades with less than ten years of activity or that it could be operating for more than 50 years while still being assessed with a grade *3* or *4*.

For the following analysis variable, Figure 3.3 shows the average grade of PPG per HEI in relation to the number of programs in each institution, displayed in blocks based on a logarithmic scale. The public or private legal status of the HEI is highlighted. The scatter plot reveals that most institutions with extensive graduate education are public, with a single private HEI offering more than 30 PPG (Pontifical Catholic University of Rio de Janeiro, with 34 PPG in 2022). Among public HEI, it is also possible to notice that there is an overall trend for higher average grades as the number of PPG per institution grows.

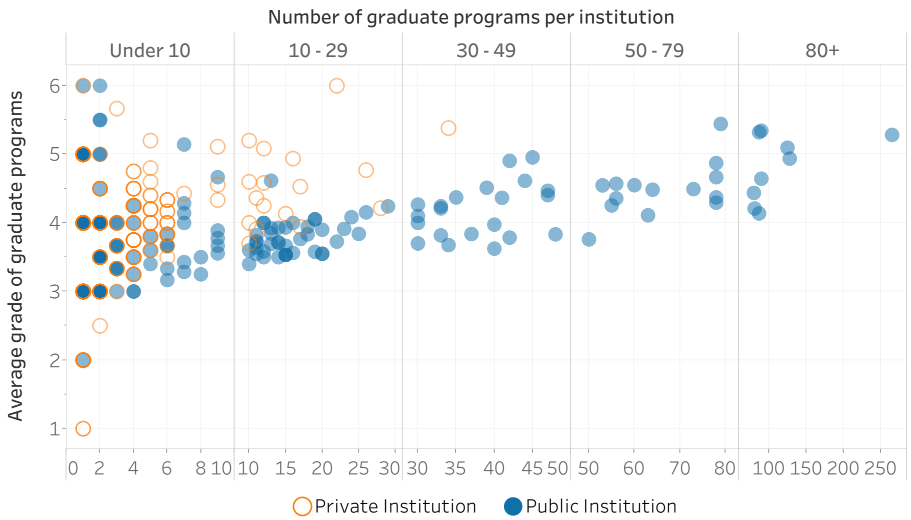


Figure 3.3.: Average grade of PPG vs total number of programs per institution (National Evaluation of 2022)

Two additional conclusions can be drawn from Figure 3.3. The first is that there is a high variation in profiles and performance in institutions with very few PPG, up to four or five. For these, it is possible to see private and public HEI ranging from very low to very high grade averages, without any clear trend. The second is that, for institutions within the 5-29 PPG range, private HEI seem to overperform the public ones. A possible explanation is that the private institutions in this range may have the potential to continue expanding but have chosen to keep their graduate level activities limited to areas of established excellence. The fact evaluation can capture this result may represent a true advancement in methods since the 1980s.

Another improvement captured in the evaluation process can be seen in Figure 3.4, where the relation between grade and number of publications per graduate program is displayed. However, when this relation was found to be significant by Schwartzman (1982), Qualis was not yet introduced as a quality proxy of the publications. So, in the displayed graph, it is possible to analyse all publications per PPG (Total), but also a subdivision according to the Qualis classification used for the 2022 evaluation, grouped in quartiles for easy interpretation.

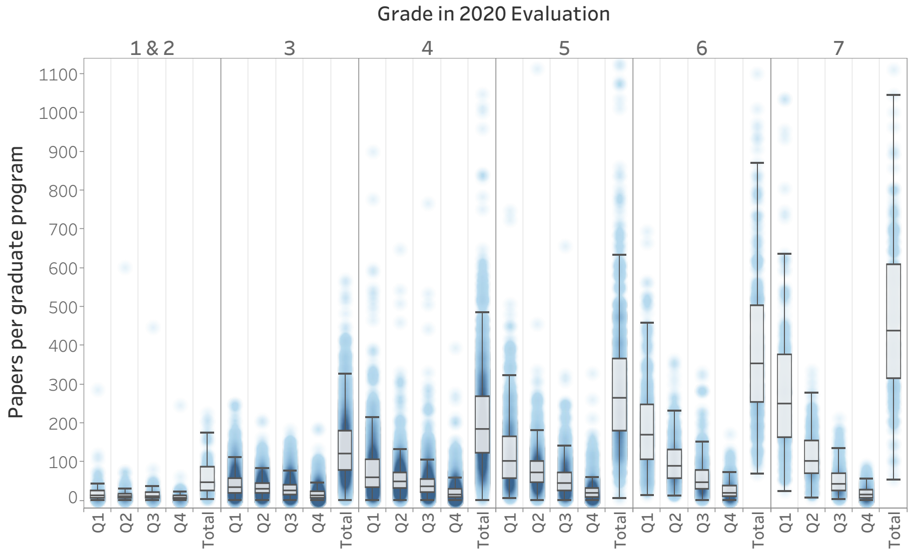


Figure 3.4.: Number of publications per PPG (per Qualis quartile and in total), according to grade obtainned at the 2022 evaluation

Figure 3.4 confirms that the grade can have a direct relation to the number of publications from a graduate program, regardless of any additional quality perspective or even considering a relative approach, such as the ratio of publications per faculty member. However, once again there are outliers in the data, including highly productive PPG with grades *4-5*, and others not as productive but graded *6-7*. The relative perspective may play a role here, but an analysis of the total broken down according to the Qualis quartiles – where Q1 represents publications in journals classified in the top 25% – shows how the classification of scientific publishing plays a role in the grade distribution.

Moving on to the geographical perspective, Figure 3.5 shows the grade evolution per Brazilian state, grouped according to the five regions of the country. The numbers displayed include the average grades in 2007 and 2022, and the overall improvement is clear for the whole country, even if it is more significant in some of the states.

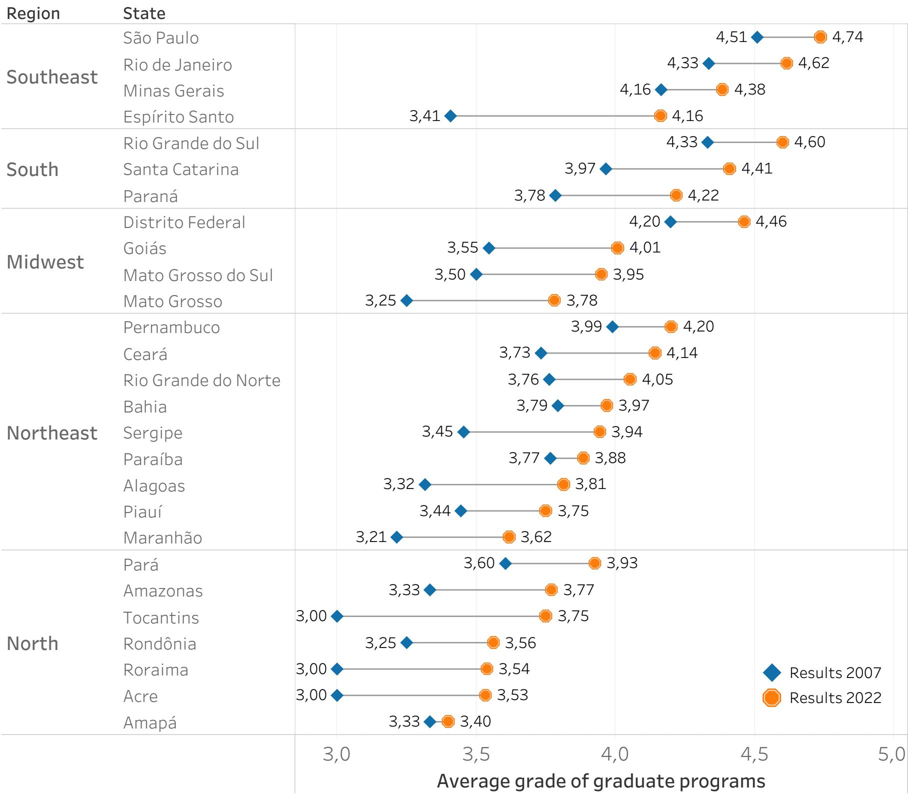


Figure 3.5.: Variation in grade average, per state, from the evaluations of 2007 to 2022

Considering the evolution seen in Figure 3.5 combined with the previous analysis in this study, it is possible to assume that graduate education has been active the longest in the Southeast and South regions and in the country's capital (Distrito Federal) than in the rest of Brazil. These regions concentrate the largest institutions, with the highest number of PPG, and as shown in Figure 3.6, they also hold most of the programs evaluated with top grades.

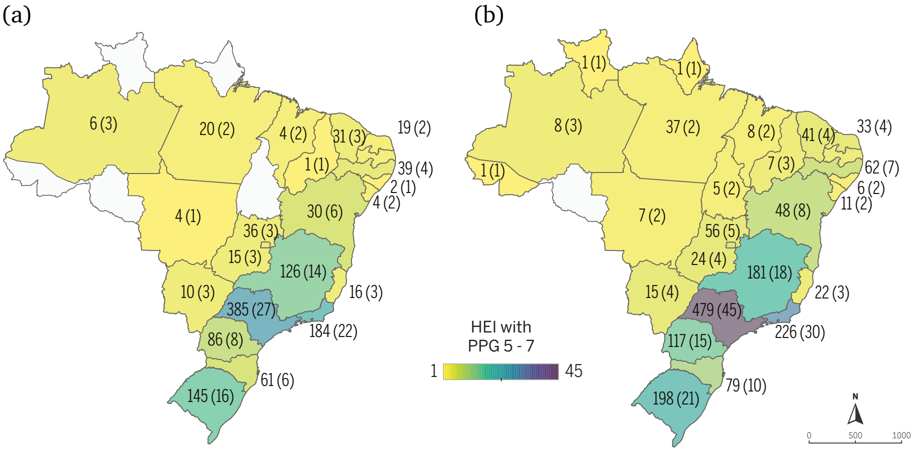


Figure 3.6.: Geographical distribution of graduate programs (grades 5-7) and the number of institutions offering them in (a) 2007 and (b) 2022

As revealed by Figure 3.6, the geographical distribution of the highest-rated PPG in Brazil shows a pronounced imbalance. This pattern is consistent with the general distribution of graduate programs in the country, as reported in an earlier study (Brasil, 2020), and is in part due to the delayed development and interiorization challenges faced by graduate and higher education in Brazil. In that sense, it may be difficult to attribute to a geographical indicator a determinant role in evaluation results.

Taking into account the changes seen from Figure 3.6a to Figure 3.6b, the most concerning part is that absolute asymetries have increased in the last 15 years. While the state of São Paulo went from 385 to 479 PPG grade *5* or above, Pará went from 20 to 37, Amazonas from 6 to 8, Acre, Roraima, and Amapá went from nothing to a single program. Even in the cases where the relative gain is higher, the absolute discrepancy is growing.

The CAPES evaluation model, as noted by Barata (2019) and Verhine and Dantas (2009), potentially facilitates the emergence of this problem, given that it promulgates a standardised evaluation procedure, which often discriminates emerging programs without taking into account regional specificities. As far back as four decades ago, Castro and Soares (1983) stressed that evaluation parameters often fail to recognise the unique history, efforts, or challenges of a PPG. Therefore, an evaluation model with flexible criteria could be needed to address asymetries in less developed regions. Yet, the standing policy of CAPES is to uniformly apply the same evaluative "thermometer". Departing from an objective diagnosis through evaluation, it could be plausible to craft different trajectories of therapeutic intervention grounded in the particularities of individual cases, thus making room for policies aimed at mitigating existing disparities. Nonetheless, the most recent evaluation outcomes suggest that the concerns articulated by Schwartzman (1982) persist.

3.7 The complexity of Brazilian evaluation

Undoubtedly, this study has not captured the entire complexity of CAPES evaluation, and certain instruments in the system were not addressed due to a relatively lower impact or limited application over time. For example, site visits to graduate programs have not been mentioned, despite being one of the strategies that could benefit from a more consistent application in the future.

According to Gatti et al. (2003), CAPES implemented site visits in 1980. The initiative was driven by the perceived challenge of adequately assessing the intricacies of graduate education through written reports alone. However, only 200 visits were conducted in 1981, prioritising newer programs, particularly those outside major academic centres. There were always three main challenges to allow site visits to become a widespread and structural component of the evaluation: the number of evaluators needed for the job, the costs, and the logistical aspects involved with the task.

In the Brazilian system, CAPES organises the logistics of all site visits to PPG, also covering associated costs. This makes the activity depend on the agency's budgetary and organisational capacity every year. Thus, site visits were used mainly to support accreditation of new PPG and only in particular cases, when requested by the evaluation committee. In 2008, an attempt was made to strengthen the role of site visits, so that their outcomes could be used to enhance the educational and formative aspects of the evaluation model. Once again, given the financial and logistical impracticality of visiting every PPG, the activity has been restricted to situations such as by special request from evaluation committees or the CTC-ES, or when recommended by CAPES itself, for example, when PPG seem unable to move up from a grade *3* after three consecutive evaluations. However, the integration of site visits for classification purposes remained ambiguous, as it was not clear how the results of these visits could be integrated when they were limited to only special cases (Verhine, 2008).

The report by Oliva et al. (2017), proposing advances for the evaluation system in Brazil, highlighted that some dimensions of graduate education could not be comprehended through an evaluation exclusively based on information derived from the data provided annually by PPG, as they can be effective in providing product and outcome data, but are not equipped to reveal dynamics and processes. Therefore, the use of well-structured and systematically planned expert-led visits was recommended. Ideally, all PPG should receive a visit close to a quadrennial evaluation, preventing unfairness if incorporating the results as part of the grade assigned to the programs. However, the prospect of visiting nearly 5000 PPG is unrealistic given the financial and logistical burden borne by CAPES. Therefore, Oliva et al. (2017) proposes that each program should receive at least one visit over an eight-year period, representing a more feasible 600 visits per year, approximately.

The main benefit of this model is that it would not serve a classification purpose but a formative one. The approach aligns with (Barata, 2019) recommendation to consider follow-up visits as one of the most effective evaluation strategies. The idea here is to make use of an important outcome of the evaluation process, which is the individual evaluation report that accompanies the grade for a PPG. Although the academic community has focused mainly on the final grades of the evaluation – due to its direct consequences on funding, accreditation, and reputation – these reports are provided since the early 1980s, and they represent the constructive side of the results, being useful to help PPG improve their work in the following years (Sguissardi, 2006). With a site visit guided by those reports, recommendations could be effectively negotiated with the PPG and also with the higher administration of their respective HEI (Barata, 2019).

Together with site visits, the overall evaluation process benefits from a series of additional strategies, such as midterm seminars organised to discuss the panorama of each area with the graduate programs, in preparation for the quadrennial evaluation to come. The whole process is complex and costly, but the results cannot be taken lightly, as they are central for the quality of the Brazilian National System of Graduate Education.

3.8 Conclusion

This study provides a comprehensive exploration of the evolution, impacts, and challenges of Brazilian evaluation, an instrumental mechanism for shaping research and graduate education in the country. The model can be considered a catalyst to improve the quality of graduate education, as evidenced by its integral role in the allocation of funding and the conferral of legitimacy on graduate programs in Brazil (Marques et al., 2020).

The research examined the development of the national evaluation system in Brazil, providing a comprehensive account of the genesis of the system, while also outlining the critical steps taken from its conception to the current model. The analysis begins in 1965 with the publication of the Sucupira Report, which shaped graduate education in the country, while pointing out the pressing need for an evaluation mechanism to manage the birth and expansion of a national system of research and graduate education.

The proposed evaluation faced difficulties to properly develop in its first decade, mainly due to the initial hurdles faced by the lack of appropriate mechanisms to implement the process. As a solution, the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) was tasked with coordinating the evaluation of graduate courses and formulating the future National Plans for Graduate Education (PNPG). This study provides a detailed analysis of how CAPES navigated its responsibilities, from the handling of accreditation and continuing evaluation, to the mitigation of challenges related to course quality, resource allocation, and public perception. Ultimately, the research accentuates the transformative influence of the system in fostering transparency, fairness, and academic quality within Brazil.

However, the model has been subject to criticism for potential drawbacks, such as prioritising publication volume over societal, academic, and scientific impacts, the possible disregard of the training dimension, and increased competition for resources between programs and institutions. Barata (2019) even points out that evaluation has led to unanticipated outcomes, such as the artificial proliferation of programs within HEI, excessive specialisation, homogeneization, and an unbalanced focus on scientific output rather than on the education of new scientists and highly qualified human resources.

A combination of external pressures and internal critique from CAPES and the evaluation community drove systematic improvements. These include periodic adjustments to the assessment form, diversification of methods to recognise different graduate education models, investments to improve data collection, and the promotion of dedicated working groups to foster the development of various facets of the evaluation process. One such example is the reconfiguration of the Qualis system, a key component of Brazilian evaluation designed to consider research quality, not just quantity. While still a work-in-progress, these enhancements underscore the system's dynamic nature and the efforts for a more holistic, context-sensitive, and impact-driven evaluation model.

Despite its merits, the CAPES model has also faced significant challenges in recent years. The Brazilian academic environment has been marked by turbulence, mainly due to funding cuts and insufficient support for the science system, further complicated by the COVID-19 pandemic and a severe crisis that affected the evaluation system. Political and managerial instability raised concerns about an attempt to dismantle CAPES, fears amplified by legal interference in ongoing reforms and evaluation procedures. This turbulent period precipitated resignations and protests, exacerbating an already complex situation. However, the completion of the evaluation for the 2017-2020 cycle, amid this tumult, represents a notable achievement for CAPES and the academic community, and there is hope that the efforts to accelerate the evolution of the system can resume.

While Brazil reclaims the valorisation of science, there is room for CAPES to continue to push away from one-size-fits-all evaluation solutions. The path forward will require a multifaceted approach combining different strategies that could involve adjustments to current evaluation processes, incorporating site visits, the advancement of self-assessment strategies and perhaps even considering inspiration from international models. Future research should continue to critically examine the CAPES model, fostering a culture of continuous self-reflection and progression toward academic excellence. This analytical approach can ensure that the evaluation process will become formative, being used to identify and address issues rather than serving as a punitive mechanism.

However, the viability of such changes heavily relies on the assurance of CAPES's existence as a state initiative, with stable and adequate funding to address the inherent challenges. If these are secured, there is hope for a future where the evaluation of graduate programs in Brazil not only achieves more equitable results, but also fosters an environment of improvement and continuous growth in the Brazilian National System of Graduate Education that is indispensable to address the challenges and asymmetries in the country.

References

Balbachevsky, E. (2004). Graduate education: emerging challenges to a successful policy. In C. Brock & S. Schwartzman (Eds.), *The challenges of education in Brazil* (pp. 209–228). Oxford, UK: Triangle Journals.

Balbachevsky, 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). Rio de Janeiro: Nova Fronteira.

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

Barata, R. B. (2016). Dez coisas que você deveria saber sobre o Qualis. *Revista Brasileira de Pós-Graduação*, *13*(30), 13–40. https://doi.org/10.21713/2358-2332.0.947

Barata, R. B. (2019). Necessary changes in the evaluation of graduate programs in Brazil. *Interface - Comunicação, Saúde, Educação*, *23*. https://doi.org/10.1590/interface.180635

BRASIL. (1931b). *Decreto nº 19.851, de 11 de Abril de 1931. Dispõe que o ensino superior no Brasil obedecerá o sistema universitário*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1951). *Decreto nº 29.741, de 11 de julho de 1951. Institui uma comissão para promover a Campanha Nacional de Aperfeiçoamento de Pessoal de Nível Superior*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1974a). *Decreto nº 73.411, de 4 de janeiro de 1974. Institui o Conselho Nacional de Pós-Graduação e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1974b). *Decreto nº 74.299, de 18 de julho de 1974. Dispõe sobre a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (2007). *Decreto nº 6316, de 20 de dezembro de 2007. Aprova o Estatuto da CAPES* (Decree No. 6316/2007). Diário Oficial da União. Brasília, DF. Retrieved May 13, 2023, from https://bit.ly/42oR4Z3

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2022). Multidimensionality through self-evaluation: from theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation*, (53). https://doi.org/10.22163/fteval.2022.546

Brasil, A. (2023a). Between bibliometrics and peer review: The evolution and challenges of Brazil's Qualis system. preprint. Retrieved from https://bit.ly/412rlnA

Brasil, A. (2023b). Rethinking a national classification of research and graduate education. *Profesional de la Información*, *32*(2), e320224. https://doi.org/10.3145/epi.2023.mar.24

Brigatti, F. (2023). Economistas se rebelam contra mudança em ranking de revistas científicas. *Folha de São Paulo*. Retrieved from https://bit.ly/3AesptT

CAPES. (1981). *Plano Nacional de Pós-Graduação - PNPG 1982-1985*. Brasília, DF: CAPES.

CAPES. (2014). *Resolução nº 5, de 11 de Dezembro de 2014. Altera a periodicidade da avaliação do SNPG* (Resolution No. 5/2014). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3BK5Ywu

CAPES. (2017b). *Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal*. Diário Oficial da União. Brasília, DF.

CAPES. (2018a). Avaliação comparada da pós-graduação. Seminários: Repensando a Avaliação. Retrieved February 2022, from https://bit.ly/3HyExbc

CAPES. (2022a). CAPES divulgará os resultados parciais. Retrieved May 13, 2023, from https://bit.ly/3W0LObG

CAPES. (2022b). Resultado da Avaliação Quadrienal 2017-2020. Retrieved January 5, 2023, from https://bit.ly/3Gj9BfE

CAPES. (2023a). Aumento no valor das bolsas marca ações da CAPES. Retrieved May 14, 2023, from https://bit.ly/3nWhkep

CAPES. (2023b). *Documento técnico do Qualis Periódicos*. CAPES. Brasília, DF. Retrieved April 12, 2023, from https://bit.ly/3KTmUGY

CAPES. (2023c). Metodologia do Qualis Referência: Quadriênio 2017-2020. Retrieved April 12, 2023, from https://bit.ly/3my156n

Castro, C. d. M. (2006). A CAPES na visão de um ex-diretor. *Análise*, *17*(2), 360–376. Retrieved from https://bit.ly/2VNnRGY

Castro, C. d. M., & Soares, G. A. D. (1983). Avaliando as avaliações da Capes. *Revista de Administração de Empresas*, *23*(3), 63–73. https://doi.org/10.1590/s0034-75901983000300007

CNPG. (1974). *Plano Nacional de Pós-Graduação - PNPG 1975-1979*. Brasília, DF: MEC.

Comissão Especial de Acompanhamento do PNPG 2011-2020. (2018). *Proposta de aprimoramento do modelo de avaliação da PG*. CAPES. Brasília, DF. Retrieved from 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*. CAPES. Brasília, DF. Retrieved from 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*. Brasília, DF.

Davidovich, L., & Ribeiro, R. J. (2021). Manifestação da ABC e SBPC contra suspensão da Avaliação Quadrienal 2021. Retrieved March 10, 2022, from https://bit.ly/3q12v8j

Dourado, C. (2023). Fuga de cérebros, a diáspora de cientistas brasileiros. Retrieved May 14, 2023, from https://bit.ly/3MabZs6

Elton, L. (2004). Goodhart's Law and Performance Indicators in Higher Education. *Evaluation & Research in Education*, *18*(1-2), 120–128. https://doi.org/10.1080/09500790408668312

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.

Fávero, M. d. L. d. A. (2006). A universidade no Brasil: Das origens à reforma universitária de 1968. *Educar em Revista*, *22*(28), 17–36. https://doi.org/10.1590/s0104-40602006000200003

Ferrari, L. (2023). Pesquisadores criticam novo 'ranking' de revistas científicas da Capes, o Qualis. *Estadão*. Retrieved from https://bit.ly/41EGuMY

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

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

Gatti, B., André, M., Fávero, O., & Candau, V. M. F. (2003). O modelo de avaliação da CAPES. *Revista Brasileira de Educação*, (22), 137–144. https://doi.org/10.1590/s1413-24782003000100012

Hanzen, E. (2021). Maior crise da história da Capes é agravada com renúncia coletiva de coordenadores de área e consultores. *Jornal da Universidade*. Retrieved from https://bit.ly/3nYkApy

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

Horta, J. S. B., & Moraes, M. C. M. d. (2005). O sistema CAPES de avaliação da pós-graduação: Da área de educação à grande área de ciências humanas. *Revista Brasileira de Educação*, *10*(30), 95–116. https://doi.org/10.1590/s1413-24782005000300008

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

Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation Systems: What are they and why study them? *Evaluation*, *14*(2), 157–169. https://doi.org/10.1177/1356389007087537

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. https://doi.org/10.22347/2175-2753v12i37.2825

Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). *Revista Brasileira de Sociologia*, *6*(13), 9–26. https://doi.org/10.20336/rbs.256

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

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*. Brasília, DF: CAPES. Retrieved from https://bit.ly/2PmAf0q

Monteiro, A. L., Meneghetti, S. M. P., Suarez, P. A. Z., Rubira, A. F., Zarbin, l. J. G., Rezende, C. M. d., ... Coltro, W. K. T. (2021). Carta aberta dos Coordenadores e Consultores da Área de Química na CAPES. Retrieved May 14, 2022, from https://bit.ly/3sBhVkO

Morato, N. (2015). Estratégias e perspectivas para a pós-graduação são debatidas em encontro. Retrieved May 13, 2023, from https://bit.ly/44TT2ST

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton & Oxford: Princeton University Press.

Nader, H., & Davidovich, L. (2021). Crise na Capes e a batalha pela Ciência brasileira. *Piauí*. Retrieved May 13, 2023, from https://bit.ly/44ZdMsw

Oliva, G., Tourinho, E. Z., Neto, E. d. M., Audy, J. L. N., Bevilacqua, L., Laplane, M. F., ... Báo, S. N. (2017). *Comissão Especial para Análise do Sistema e Processo de Avaliação da Qualidade da Pós-graduação Brasileira*. Brasília.

Ribeiro, R. J. (2008). Quality in Research: CAPES' point of view. In J. L. N. Audy & M. C. Morosini (Eds.), *Innovation and quality in the university* (pp. 127–169). Porto Alegre: EDIPUCRS.

Rothen, J. C. (2018). A universidade brasileira na Reforma Francisco Campos de 1931. *Revista Brasileira de História da Educação*, *8*(17), 141–160.

Saldaña, P. (2021b). Diretor de Avaliação da Capes pede demissão após renúncia de pesquisadores. *Folha de São Paulo*. Retrieved from https://bit.ly/42SWT0M

SBPC. (2021a). Carta aberta da SBPC sobre a CAPES. Retrieved March 10, 2022, from https://bit.ly/3CnkQ4o

SBPC. (2021b). Entidades se manifestam em defesa da Capes. Retrieved May 14, 2022, from https://bit.ly/38xTKN8

Schwartzman, S. (1982). Avaliando a pós-graduação: a prática da teoria. *Instituto Universitário de Pesquisas do Rio de Janeiro*, (10).

Sguissardi, V. (2006). A avaliação defensiva no 'modelo CAPES de avaliação': É possível conciliar avaliação educativa com processos de regulação e controle do Estado? *Perspectiva*, *24*(1), 49–88. Retrieved from https://bit.ly/2VKO88E

Siqueira, M. B. (2019). Sucupira - A Platform for the Evaluation of Graduate Education in Brazil. *Procedia Computer Science*, *146*(2019), 247–255. https://doi.org/10.1016/j.procs.2019.01.081

Sucupira, N. (1980). Antecedentes e primórdios da pós-graduação. *Forum Educacional*, *4*(4), 3–18.

Ulhôa, M. T. d. (2017). As consequências do QUALIS artístico. *ARJ – Art Research Journal / Revista de Pesquisa em Artes*, *3*(3), XLII–LI. https://doi.org/10.36025/arj.v3i3.12079

Vasconcellos, J. (2021). Presidente da Capes nega a dissolução de conselho. *Correio Braziliense*. Retrieved from https://bit.ly/3KGMk7F

Veiga, J. M. (2021). Decisão judicial autoriza Capes a retomar Avaliação Quadrienal; retrieved May 13, 2023, from https://bit.ly/3BoBaSE

Verhine, R. E. (2008). Avaliação da CAPES: Subsídios para a reformulação do modelo. In D. Mancebo, J. d. R. Silva Jr, & J. F. de Oliveira (Eds.), *Reformas e políticas: Educação superior e pós-graduação no brasil* (pp. 165–188). Campinas: Alínea.

Verhine, R. E., & Dantas, L. (2009). Reflexões sobre o sistema de avaliação da CAPES a partir do V Plano Nacional de Pós-graduação. *Revista de Educação Pública*, *18*(37), 295–310.

Verhine, R. E., & Souza, Â. R. d. (2021). Compreendendo a Crise Recente na Pós-Graduação Brasileira. Retrieved May 13, 2023, from https://bit.ly/42F0dNq

Verhine, R. E., Lima, A. C. S., Carvalho, D. B. B. d., Leite, D., Bertolin, J. C. G., Dantas, L. V., . . . Rosa, D. L. (2019). *Autoavaliação de Programas de Pós-Graduação*. Brasília, DF: CAPES. Retrieved from https://bit.ly/3jTSRS8

Verhine, R. E., Faljoni-Alario, A., Federizzi, L. C., Ferreira, F., & Oliveira, T. M. d. (2021). *Equivalência entre conceitos e notas*. CAPES. Brasília, DF. Retrieved June 21, 2023, from https://bit.ly/3NF4tXP

Viana, W. F. (2018). *Sistema CAPES de avaliação da pós-graduação stricto sensu: Um estudo de caso da área de Administração Pública no Brasil* (Doctoral dissertation, Universidade de Brasília).

Yamashita, M. (2023). Lista Qualis e a duvidosa qualidade dos critérios de qualidade. *Questão de Ciência*. Retrieved from https://bit.ly/3UShMq1

Part II

The Brazilian evaluation system in perspective

4

Research evaluation in Brazil and the Netherlands

> *A performance-based funding system, because it encourages competition, may also encourage a shift towards the 'homogenization' of research.*
>
> **— Aldo Geuna & Ben Martin**

Data from the Web of Science reveal that Brazil and the Netherlands are very close in terms of their indexed scientific output (Clarivate Analytics, 2022). Between 2017 and 2021, Brazil ranked 13th among the top-producing countries in the database, publishing a total of 289.562 papers and reviews. With 241.863 publications, the Netherlands followed in 14th place. Despite the proximity in absolute numbers, there are significant differences between the countries when the results are observed from a relative perspective. For instance, while the Netherlands has a population of around 17.6 million (CBS, 2021), Brazil has already exceeded 213 million people (IBGE, 2021). That means the Latin American country produced 136 publications per 100.000 people, ten times less than the European counterpart, at 1374 publications.

Another relevant distinction between the scientific production of the two countries is evident from the analysis of impact indicators. For example, considering the percentage of publications of each country in the upper 10% percentile of the citation distribution in the same fields (PP top 10%), Brazil performs below the average of the database, at 7,7% of the expected 10% value. The observed

This chapter has been submitted for review and it is available as a preprint: Brasil, A. & Trevisol, J.V. (2023) Research evaluation in Brazil and the Netherlands: a comparative study. SocArXiv. https://doi.org/10.31235/osf.io/er9fa.

impact is significantly higher for the Netherlands, since 17,3% of the country's publications are in the top 10% (see Bornmann, 2014).

Differences in impact and relative productivity may suggest that there is a higher level of efficiency in the Dutch science system, which has attracted the attention of Brazilian researchers (e.g., Marcovitch et al., 2018; Verhine and de Freitas, 2012), policymakers (e.g., Barbosa, 2020), and major funding organisations (e.g., CAPES, 2018a; CAPES, 2018b). Based on the views on research governance and the role of evaluation presented by Molas-Gallart (2012), Brazil seeks inspiration from the stable and long-standing Dutch evaluation system, which has been recognised as a critical factor in the quality assurance of the country (van Drooge et al., 2013, p. 17).

Although countries may benefit from international experiences to improve their practises, it is also necessary to carefully reflect on potential learnings. According to Faljoni-Alario et al. (2018, p. 5), Brazil already has one of the most sophisticated performance-based evaluation systems in the world, an argument explored in depth by Brasil (2023c). Potential changes should, therefore, consider what has already been achieved. Furthermore, science systems can be as distinct as the social-economic circumstances, established governance, and cultural realities of each country. Potential disparities should be considered before replicating any strategy that has been successful elsewhere. For instance, the design of the Brazilian science system must account for geographical realities in very particular ways, since the South American country, at 8.5 million km^2, faces challenges that would be more comparable to the whole European Union – which is about half the size of Brazil, with 4.2 million km^2 – than to the Netherlands, 205 times smaller at 41.543 km^2 (CBS, 2021; IBGE, 2021).

This work investigates the nuances and mechanisms of research evaluation in Brazil and the Netherlands, guided by two primary research questions: i) How have the historical trajectories and policy initiatives of each country shaped their current research evaluation systems? ii) How do the consequences of each evaluation system influence the behaviours and objectives of researchers and institutions within their respective contexts? By critically analysing policy documents, legislation, and existing literature, we aim to compare the architectural frameworks of each country's science and evaluation systems, exploring vital components like the interplay between evaluation and funding, and the repercussions of evaluation outcomes on researcher behaviour. Finally, we conclude by highlighting the inspiring methods and approaches of each evaluation system, so that those insights could foster a positive evolution in research evaluation practises not only for these two nations, but also for countries sharing analogous experiences.

4.1 Methodology

According to Galleron et al. (2017), different typologies of research evaluation systems have been proposed over time, but none of the existing frameworks was applicable to compare multiple national systems. Existing research evaluation studies exhibit limitations in the typologies used, as they tend to focus on a narrow selection of countries, disregard discipline-specific evaluation method adaptations, and emphasise primarily financial impacts or performance-based funding systems (Ochsner et al., 2018). Taking into account a national research evaluation system as “the particular combination and organisation of evaluation practises in place that affect the researchers in doing research in their country and sets their context of accountability/evaluation” (Ochsner et al., 2019, p. 3), a working group within the European Network for Research Evaluation in the Social Sciences and Humanities (ENRESSH) has designed a new framework alternative to support analyses of systems in Europe and beyond (Ochsner et al., 2018). This article is derived from that larger research project.

The design of the ENRESSH framework was grounded on existing typologies such as those proposed by Coryn et al. (2007), Geuna and Martin (2003), and Hicks (2012). After extensive review by the network's steering committee, a synthesis of typologies was the object of two rounds of surveys with ENRESSH members from 33 different countries, and also of the application of an intermediate questionnaire incorporating suggested dimensions (Ochsner et al., 2018). The resulting framework included three dimensions: institutional evaluation, national career promotion, and grant evaluation, from which only the first will be adopted in this study. Regarding the various analytical categories in the framework, 17 are part of the institutional evaluation: name, level, responsible entity, legal framework, unit of assessment, time framework, method, SSH specificity, bibliometric data, scientometric data, language preference, gender issues, funding link, changes over time, transparency, controversies, and rapporteur's perception of influence on researchers' way of working.

An analysis of the Brazilian institutional evaluation according to those categories was already the object of an ENRESSH report produced by Brasil and Trevisol (2023). This study takes that effort a bit further for Brazil and also replicates it with an extensive analysis of the practise and regulatory framework of the Dutch evaluation system. From the investigation of the Strategy Evaluation Protocol (SEP) that is used in the country (VSNU et al., 2020), as well as its previous iterations and related legislation, the study continues to apply the 17 aforementioned ENRESSH analytical categories to compare the national evaluation models in the two countries. However, it seemed necessary to enrich the adopted framework with two additional categories to account for the particulars of the selected countries: site visit and accreditation effects. Also, only for organisational purposes, in this chapter the categories are grouped into four themes that will be detailed in the following section: organisational framework; methods and data; evaluation stakes; and transparency and controversies.

For each category in the expanded framework, Brazilian and Dutch experiences are contrasted, identifying distinctions and similarities recorded in policy documents and connected legislation (e.g., OCW, 1992; VSNU et al., 2003; VSNU et al., 2009; VSNU et al., 2016), or emerging from an extensive review of the literature conducted on national evaluations and evaluation impact, including works by Capano (2010), Hammarfelt and de Rijcke (2014), Leeuw and Furubo (2008), Molas-Gallart (2012; 2014), Ochsner et al. (2020), Verhine and de Freitas (2012), and others. The intended result is to find the inspiration sought by Brazil, while also identifying suitable lessons for the Netherlands.

4.2 Comparing Brazil and the Netherlands

The Dutch and Brazilian evaluation systems developed from distinct conceptions of assessment, university autonomy, and governance of higher education, science, and technology. Regulatory frameworks and improvements have institutionalised sui generis and unique systems over time, shaping distinct research cultures that may contribute to the previously mentioned impact differences.

In Brazil, the university system is relatively young. Higher education became a reality in the country only in the 19th century, with courses offered by a few small institutes that focus mainly on professional training (Cunha, 2007).

In these institutions, research was rare, incipient and limited to a few areas (Fávero, 2006; C. B. Martins, 2018). The first universities were founded only in the early 20th century, creating a platform for institutionalisation of research and training of researchers in the country. However, a robust national science system only started to be shaped through state policy during the 1950s and 1960s. At that time, Brazil invested in building a graduate education system that has been the house of science in the country (Balbachevsky and Schwartzman, 2010; Brasil, 2020; Cano, 2015; Rothen, 2018). According to C. B. Martins (2018), this has been an important instrument for the modernisation of higher education, the installation of academic competencies in the country, and the institutionalisation of research in universities. Masters and doctorate courses have changed the shape and dynamics of Brazilian higher education.

Data for 2021 show that there were 2574 active higher education institutions (HEI) in Brazil. Almost 88% of those were private – mainly focused on offering professional training – representing nearly 77% of the nearly nine million students enroled at the undergraduate level. The situation is different in graduate education, where around 430 of the country's HEI offered master or doctoral courses. While 52% of these were private, the public sector held 84% of the more than 325.000 students enroled, comprising 192.000 at the master's level and 133.000 at the doctoral level (Brasil, 2020; CAPES, 2009; INEP, 2022).

The Dutch higher education system, with roots dating back to the establishment of Leiden University in 1575, boasts a rich academic tradition that predates even the formation of the Dutch State (Cohen and Steege, 1982). Over the centuries, the system has evolved while maintaining a public good perspective grounded in strong institutional autonomy and has become a diverse and internationally orientated landscape with 14 research universities (*wetenschappelijk onderwijs – WO*) and 43 polytechnic or applied sciences universities (*hoger beroepsonderwijs - HBO*), ensuring a comprehensive approach to education and workforce qualification (Rijksoverheid, 2023; UNL, 2022). Institutions are maintained through public funding and resources from teaching, research, and service provision (Goedegebuure and Westerheijden, 1991).

At the end of 2022, the Universities of the Netherlands (UNL)[1] reported that approximately 350 thousand students were enroled in the country's research

[1] Former *Vereniging van Samenwerkende Nederlandse Universiteiten* (VSNU), which was renamed in 2021 to more accurately reflect the unified position of Dutch universities (Cuppen, 2021).

universities, of which 225 thousand were at the bachelor's level and the remaining at the master's level (UNL, 2022). For HBO institutions, according to Vereniging Hogescholen (2023), the number of bachelor students was around 441 thousand and that of the masters was close to 15 thousand.[2] The number of PhD candidates reported for 2022 in the country was approximately 37 thousand, all in research universities, since HBO institutions started a project for professional doctorates only in 2023 (UNL, 2022; Vereniging Hogescholen, 2023). Although the absolute figures for Brazil are considerably higher, the Dutch numbers are quite impressive in a relative perspective, not only considering geographical and populational differences, but especially regarding the relatively small number of higher education institutions in the European country.

However, a significant distinction that impacts the object of the present study – the comparison of research evaluation systems – relates to master's students. In Brazil, they are an integral part of the science system, with master's programs operating almost as short-doctorates all over the country. Their value is significant, especially because the distribution of doctoral programs is still quite asymmetric, as these are primarily located in metropolitan areas, often in the southeast coastal region. Therefore, when evaluating research, master's are included in the analysis (Brasil, 2020).

The same is not true for The Netherlands, as the Dutch higher education system underwent significant restructuring following the Bologna Process (European Ministers of Education, 1999). The initiative aimed to create a European Higher Education Area, promoting international collaboration, and ensuring degree comparability and quality assurance. The Netherlands swiftly adopted the European Credit Transfer and Accumulation System (ECTS) and the two-cycle degree structure, with bachelors of three years (180 ECTS) followed by a master's degree (60 to 180 ECTS). Before the Bologna Process, the Dutch system was composed of an undergraduate level equivalent – comparable to a bachelor – followed by a doctorate. There was no master's level at all. Thus, Dutch master's programs are considered part of an educational trajectory, not being the object of research evaluation protocols in the country (NUFFIC, 2023; VSNU et al., 2020; Westerheijden et al., 2008).

[2] It is relevant to notice that the proportion of masters students in HBO should not be compared to the ones in WO. As discussed, the missions of the two types of institutions are quite distinct.

In light of the internationalisation perspective, one of the core ideas of the Bologna Process was to facilitate student mobility across Europe. For most signatory countries, this entailed offering programs in English to accommodate international students. The Netherlands emerged as the nation with the highest availability of English-taught higher education degrees in continental Europe, one of the reasons why almost 25% of the students enroled in Dutch research universities are international, with 72% coming from the European Economic Area (CBS, 2022; Elven, 2023). In stark contrast, the Brazilian education census of 2021 reveals that only 17.947 international students were attending undergraduate degrees in the country, representing only 0,2% of the total number of enrolments (INEP, 2022). The disparate levels of internationalisation in the higher education systems of the Netherlands and Brazil undoubtedly influence the production and evaluation of scientific research in both countries.

Taking into account the complexity and significant differences between the Dutch and Brazilian research and higher education systems, the following sections will then present the analytical categories proposed by ENRESSH, thematically grouped and with mentioned additions to account for the particulars of the selected countries.

4.2.1 Organisational framework

Table 4.1.: Organisational comparison of the Brazilian and Dutch evaluation systems

Categories	Brazil	The Netherlands
Name	Quadrennial Evaluation	Strategy Evaluation Protocol
Level	National	Institutional
Responsible entity	National agency	Institutions
Legal framework	Legal ordinances and field-specific documents (both periodically issued)	Higher Education Act (WHW) enforced through the Strategy Evaluation Protocol (SEP)
Unit of assessment	Graduate program	Research unit
Time framework	4 years (syncronous)	6 years (assyncronous)

Source: Brasil and Trevisol (2023), Scholten et al. (2018), and VSNU et al. (2020).

In Brazil, research is assessed through graduate education, as these dimensions became interdependent and complementary from the approval of the Sucupira Report (CFE, 1965), a document considered the foundation of the country's Brazilian National System of Graduate Education (SNPG) (Cury, 2005). The report created legal and structural conditions for institutionalising research in conjunction with master's and doctoral courses – which integrate graduate programs (PPG) – while also laying the foundation for a national evaluation. As a consequence of interdependence, there is no specific regulatory framework that is specific to evaluate research quality at the national level (Brasil, 2020).

In the Brazilian system, the Federal Government plays the role of primary funder and evaluator (Cury, 2005; C. B. Martins, 2018). The country's evaluation system has been in continuous operation since the 1970s, led by the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES), a government organisation linked to the country's Ministry of Education (Sguissardi, 2006). Through CAPES, the Brazilian state designs, implements and evaluates graduate education policies and performance at the national level, guaranteeing the operation, stability and quality assurance of the system from a hierarchical approach (Barroso, 2005; Cury, 2005; Verhine, 2008).

Until the late 1970s, CAPES distributed funding on an individual level, for instance, by assessing proposals from individuals seeking scholarships to attend master or doctoral courses. With the growth of the system, the task was partially transferred to higher education institutions (HEI). CAPES would assess graduate programs (PPG), allocate a quota of scholarships, and the PPG would be responsible for the internal distribution (Ferreira and Moreira, 2002). The first PPG evaluation took place in 1976 and was held annually until 1984, when it became biannual. The periodicity changed twice more, becoming triannual in 1998, and quadrennial in 2014 (Viana, 2018). These changes and other incremental adjustments are introduced primarily as legal ordinances that shape the evolving regulatory framework. Over the past decades, this regulation has led to a model of top-down, centralised, and centripetal organisation (Saviani, 2020). As such, the system is known today as the Quadrennial Evaluation, which is conducted at the national level as an *ex post* exercise, simultaneously for all PPG in the country (CAPES, 2014).

According to Capano (2010), evaluation and accountability have been popular catchwords employed by higher education reformers over the past 25 years.

From an institutional accountability perspective, HEIs are asked to report their own performance to external stakeholders, especially the government and the public, as a way to justify investments and continued support. From the Brazilian perspective, it is evident that the government steers the science system at a distance using evaluation as the main audit tool. However, the following paragraphs show that the main stakeholders of the Dutch evaluation are the institutions and research units themselves.

The Netherlands was one of the first European countries to establish a quality assessment system for teaching and research (van Drooge et al., 2013; Weert and Boezerooy, 2007). According to Goedegebuure and Westerheijden (1991), the Conditional Funding can be considered the first regulatory framework to establish a formal quality assessment system in the Netherlands, establishing general guidelines for its organisation. The system was introduced in 1983 with the objectives of increasing accountability, promoting quality, and improving university research policy. One of its main decisions was to implement an *ex post* evaluation to reallocate the budgets among the universities.

The legal framework in the Netherlands developed over the following years, but the Higher Education and Research Act (OCW, 1992) was of particular relevance, since the Ministry of Education, Culture and Science stopped issuing guidelines and rules referring to the evaluation process. From then on, the regulatory competencies began to be exercised by associations representing the institutions that carry out the evaluated activities (van Drooge et al., 2013; Weert and Boezerooy, 2007).

As part of this new reality, in the 1990s, the structure of the VSNU (currently UNL) featured "chambers", each representing a specific discipline in the Dutch university system. Specialists from various academic ranks filled these chambers, primarily focused on representing Dutch universities and shaping research assessments in their domains. The assessment results originating from these chambers were appraised and ratified by the deans of the corresponding faculties, such as the proposals related to chemistry being reviewed by the deans of the faculties of natural sciences. This approach ensured that research evaluations were handled primarily at the level at which research was actually conducted, with limited participation of university boards.

A pivotal shift in research assessment practises was introduced in 2003, when a new evaluation protocol transitioned decision-making from the college of deans to the local university boards (VSNU et al., 2003). Since then, the Rector's Conference – operating under the former VSNU flag – has been chiefly responsible for shaping academic evaluation protocols, with the collaboration and endorsement of the Netherlands Organisation for Scientific Research (NWO), and the Royal Netherlands Academy of Arts and Sciences (KNAW). Protocols[3] are updated every six years (van Drooge, 2021b).

Over the last four decades, eight important regulatory frameworks have been published in the Netherlands: Conditional Funding; Higher Education and Research Act; VSNU Quality Assessment of Research - Protocol 1993; VSNU Quality Assessment of Research – Protocol 1994; VSNU Assessment of Research Quality - Protocol 1998; Standard Evaluation Protocol (SEP) 2003–2009; Standard Evaluation Protocol (SEP) 2009–2015; Standard Evaluation Protocol (SEP) 2015–2021 and Strategy Evaluation Protocol (SEP) 2021–2027. The first three SEP were defined as Standard Evaluation Protocols, and the most recent became a Strategy Evaluation Protocol, highlighting that the goal is not to evaluate the research itself, but the unit's research strategy in light of its own objectives (van Drooge, 2021b).

The VSNU protocols considered disciplines as units of assessment. From the SEP 2009–2015, the units became research institutes, research centres, research groups, multi- and interdisciplinary research centres, etc. These are, in general, the units of assessment, but institutions have the autonomy to define their organisation and scope[4], considering the following conditions to be evaluated: (i) be recognised, internally and externally, as a research entity; (ii) have clearly defined and shared objectives, goals and strategies; (iii) have, at least, the equivalent of ten full-time researchers in its permanent academic staff, not counting doctoral and post-doctoral students; (iv) have at least three years of operation (VSNU et al., 2020, p. 12).

While in Brazil research integrates graduate programs, in the Netherlands the situation is reversed. Dutch protocols do not establish a separation between research and graduate education, and there are no specific regulatory frameworks for the evaluation of the second. Research units usually include PhD

[3] This study focusses on academic evaluation protocols, but specific protocols for applied sciences universities and non-academic research institutes also exist.

[4] Evaluation mostly focusses on a single unit, but it is possible to assess a discipline at a national or institutional level, or research developed by a faculty, research centre, or group of research units that work in a joint and integrated way, as "umbrella" projects (VSNU et al., 2020, p. 15).

programs, and these are evaluated within the scope of the units. Furthermore, as discussed, master courses are not considered in this evaluation process.

Unlike Brazil, the state does not act as an evaluator in the Netherlands. The coordination and execution of the evaluation process is the responsibility of the institutions and their research units, in a process linked to the principle of autonomy and institutional planning. In this way, the Dutch evaluation system is bottom-up and decentralised (van Drooge et al., 2013; VSNU et al., 2020).

Regarding time frameworks, the evaluation cycle in Brazil consists of four years, in syncronous evaluation exercises. In the country, the whole system is assessed at the same time, referring to the performance of all units during a fixed four-year period. In The Netherlands, evaluation cycles consist of six years, but the evaluation is held asyncronously. This means that a share of the disciplines is evaluated during each year of the ongoing evaluation window. This approach puts less pressure on the system and the institutions, also making the evaluation a continuous effort (CAPES, 2021d; VSNU et al., 2020).

4.2.2 Methods and data

Table 4.2.: Methodological comparison of the Brazilian and Dutch evaluation systems

Categories	Brazil	The Netherlands
Method	Informed peer review	Informed peer review
Language preference	National language	Push to English
Bibliometric data	Web of Science/Scopus/Google Scholar/CRIS	Each research unit selects data sources (and indicators) that better support their self-assessment
Scientometric data	Comprehensive data collection is conducted yearly from every graduate program (microdata level)	A custom-made selection of sources is made by each research unit, according to their self-assessment approach
Benchmarking	Evaluation is comparative at a national scale, but within 49 evaluation areas.	Research units usually decide whether they want to include a benchmark and which other units they would include. In some fields, national benchmarking is also possible (e.g., psychology)

Continue...

Table 4.2 Continued

Categories	Brazil	The Netherlands
SSH specificity	Evaluated field-specifically	Department-based evaluation, with a separate addendum for the Humanities
Site visit	It may be recommended by the assessment committees in special circumstances.	Integral part of the process, mandatory in the current evaluation protocol.

Source: Brasil and Trevisol (2023), Scholten et al. (2018), and VSNU et al. (2020).

Both the Brazilian and the Dutch evaluation systems can be characterised as informed peer review. However, there are significant differences to consider. For instance, in Brazil, the organisation and modus operandi of the system were established based on the principle that evaluation should be external, independent and carried out by peers, based on official evidence collected and audited by a centralising agency. CAPES is responsible for the metaevaluation and macro-efficiency of the SNPG. Quality evaluation, on the other hand, is carried out entirely by experts in peer review committees organised in 49 evaluation (or knowledge) areas (CAPES, 2017b; CAPES, 2021d; Viana, 2018).

Regulatory frameworks and evaluation procedures are defined through a complex system of cooperation and dialogue between CAPES, as the agency that enforces evaluation, and the scientific community (Viana, 2018). Advanced information systems are used to collect data from all graduate programs in the country, meaning CAPES does not depend on limited data sets for the evaluation process. For instance, instead of relying on Web of Science or Scopus data – which have significant coverage variations across disciplines and limitations regarding non-English output (Brasil, 2021b) – committees can rely on a nearly complete list of scientific output from PPG. Additionally, the collection is not restricted to products such as conference and journal papers. It includes patents, reports, translations, teaching materials, and many other types of output. After collection, the data are processed and enriched with regional and international bibliographic data sets, and they are provided to peer review committees as research intelligence necessary for further qualitative analysis (Siqueira, 2019).

After the committee evaluation, the results are discussed within a council composed of evaluation area coordinators and representatives from CAPES and other science & technology organisations in the country, and the final decisions are expressed through grades (1 to 7), with severe funding and accreditation consequences. These will be discussed in the next section, but it is relevant to say that the potential impact of poor results in an evaluation motivates graduate programs and HEI to provide the highest quality data possible during the collection process (CAPES, 2021d; Monteiro et al., 2019).

However, while institutions and graduate programs provide the data for analysis, they do not have a role in the evaluation itself. Self-evaluation has been proposed as a new practise, as described in Brasil (2022), but the results so far have little weight in evaluation. That has limited the motivation of some PPG from institutionalising permanent self-evaluation policies and practises, especially considering that proper guidelines from evaluators are also lacking. This absence has been reported to be one of the main gaps in the SNPG (FOPROP, 2018; D. Leite et al., 2020). Still, even a slight push for self-evaluation practises marks progress from the prior context where no encouragement existed.

Furthermore, the evaluation is comparative at the national level within each evaluation area. Although that can have negative repercussions, as a diversity of units with distinct missions and realities are measured with the same rulers, the committee perspective allows for proper consideration of disciplinary realities, such as those necessary to value research in the social sciences and humanities (SSH). These are reflected in area-level criteria and indicators, which are made publicly available before every evaluation (CAPES, 2020d).

In the Dutch case, evaluation is an internal and participatory practise carried out based on the identity, objectives, and strategies of each evaluated unit. The main purpose of evaluation is to improve quality and societal relevance (Brennan and Shah, 2000; VSNU et al., 2020; Weert and Boezerooy, 2007).

The Dutch system combines internal and external evaluation in complementary steps. The first is carried out through a self-evaluation process, which is the backbone of the system, contributing to a predominantly internal, formative, contextual, and qualitative assessment. Therefore, units are evaluated based on their mission, goals, and strategies, in a practise of self-knowledge that is not designed for external control, state regulation, or accountability. Institutions and units are responsible for the organisation and management of all stages of the process, in a self-management practise (van Drooge et al., 2013; VSNU et al., 2020).

Protocols define general evaluation objectives at the national level, but criteria are not rigid, standardised, or mandatory for all institutions in the country, as they have the autonomy to reorganise them according to the characteristics of each evaluated unit. From the bibliometric and scientometric perspectives, the self-assessment approach allows each unit to present a custom-made selection of relevant qualitative and quantitative indicators that serve as evidence to support the self-assessment narrative. As units define their own strategies and propose indicators to support the narratives presented, the specificity of SSH is always easy to address, and protocols even have a separate addendum for the humanities (van Drooge, 2021b; VSNU et al., 2020).

From the above, it becomes obvious that the purpose of the Dutch evaluation is not to establish rankings or a systemic benchmark. In general, the evaluation results do not allow comparisons between units, as the criteria are not homogeneous and do not generate standardised results for the entire system. However, some disciplines can be considered partial exceptions. In the case of Psychology, for instance, many of the research units throughout The Netherlands have opted for a joint evaluation exercise, according to the decision of their university boards. For those that joined, there is a relevant level of comparability, but not all universities in the country have done so.

Despite the flexibility in evidence and comparisons, there is a push toward English in the Dutch evaluation process and the considered outputs. This can be partially explained by the already discussed high levels of internationalisation of research in the country and is reflected in the language used in the SEP: it is only produced in English, with no Dutch version available (VSNU et al., 2020).

Returning to the flow of the Dutch evaluation, the second step consists of the assessment by an international, impartial, and independent assessment committee. The committee analyses the self-assessment report and meets with the unit representatives during a site visit. The self-evaluation report and the site visit are the basis upon which the committee formulates its assessment. Each unit has the opportunity to propose members of the assessment committee to the university board that formally appoints the committee (VSNU et al., 2020).

The site visit is an important and mandatory stage in the evaluation process, as the external evaluation committee has the opportunity to (i) evaluate the research infrastructure; (ii) meet the researchers and members of the unit; (iii) hold meetings and interviews with management, researchers, staff, PhD candidates, and stakeholders, and (iv) request additional information and documents (VSNU et al., 2020, pp. 13, 21 and 42). At the end of the visit, the external committee presents the main conclusions, insights, and recommendations, offering elements for the continuous improvement of the unit. After this phase, the self-assessment report is made public, together with the committee report and a position document, in which the unit has the possibility to reflect, agree, and even disagree with the committee's views.

In the Brazilian case, site visits were implemented in the early 1980s, but could not be established as mandatory components of the evaluation. While in the Dutch system, visits are organised and funded by the evaluated unit, in Brazil, all costs and logistics are under CAPES purview. Currently, in the scope of periodic evaluations, site visits are mostly carried out by recommendation of committees, often for PPG strugling with performance. During the grade discussions within CAPES' council, it is also possible to request visits, especially when more information is required to reach an agreement of the evaluation results. Visits are also organised in the case of particular CAPES policies, for instance, when a PPG performance does not improve after being considered "regular" in three subsequent evaluation cycles (Viana, 2018).

4.2.3 Evaluation stakes

Table 4.3.: Stakes of the Brazilian and Dutch evaluation systems

Categories	Brazil	The Netherlands
Evaluation type	Performance-based	Formative
Accreditation effects	Results determine if accreditation of graduate program is renewed	Results do not impact accreditation
Funding link	Strong	Weak
Rapporteur's perception of influence on researchers' way of working	Strong	Weak

Source: Brasil and Trevisol (2023), Scholten et al. (2018), and VSNU et al. (2020).

Some of the main objectives of the Brazilian Quadrennial Evaluation, according to its current regulation, are: (i) understanding the current panorama of Brazilian graduate education in a given evaluation cycle; (ii) assessing the performance of graduate programs; (iii) quality assurance; (iv) evaluating the training of masters and doctors; (v) analysing the intellectual production of PPG and its social, economic, and cultural impact; (vi) contributing to the evolution and improvement of Brazilian graduate education (CAPES, 2021d).

Although not listed, a well-known additional objective is central to the design of the system, which is the establishment of comparative rankings between PPGs, which impact status, reputation, funding, and continuity (Sobrinho, 2003; Verhine, 2008; Verhine and de Freitas, 2012). Part of that goal is achieved by attributing grades to graduate programs, using a seven-level scale. Grades 1 and 2 are considered insufficient, leading to the closure of the program (after a grace period to allow enroled students the chance to graduate). Grades above that threshold – from 3 (regular) to 7 (excellence) – guarantee the renewal of the PPG accreditation for the subsequent four-year cycle. Besides the accreditation value, grades allow comparisons between PPG and institutions, being also used to calculate scholarship quota and funding allocation, and to restrict access to select funding streams (Brasil, 2020; CAPES, 2021d).

Another relevant issue is that CAPES not only runs the evaluation system, but also plays a significant role in regulating the SNPG while being the leading funding agency in the country. The combination of tasks and the strong links between the evaluation results, funding, and the very continued existence of a PPG lead to a high-stakes evaluation model, predominantly normative, standardised, and performance-based. Geuna and Martin (2003, p. 296) state that "a performance-based funding system, because it encourages competition, may also encourage a shift towards the 'homogenisation' of research, discouraging experiments with new approaches, and rewarding 'safe' research, irrespective of its benefits to society". That is the reality in Brazil, where the strong consequences of the evaluation lead to a reliance on quantitative methods, despite the ever present participation of peer review committees in the process.

As many evaluation metrics consider the number of faculty members as a denominator, individual researchers are also expected to contribute positively to the numerators. This includes the number of articles published in qualified journals, supervised students, taught classes, and more. As a consequence, research unit metrics become part of the faculty hiring and firing process and the very concept of a researcher's worth, undoubtly influencing their way of working.

The Netherlands experiences a different reality. According to the Strategy Evaluation Protocol (VSNU et al., 2020, p. 6), "the main goal of the SEP is to maintain and improve the quality and societal relevance of research, as well as to facilitate a continuous dialogue on research quality, societal relevance, and viability in the context of research quality assurance". From this, the evaluation is contextual and essentially formative, being carried out by the institutions themselves with the purpose of analysing the results, identifying strengths and weaknesses, and defining the changes and improvements for the future (van Drooge, 2021a; van Drooge, 2021b; VSNU et al., 2020).

In keeping with the principles above, the current evaluation model does not assign scores to the assessed units. However, this is a recent development. The evaluation protocols active between 1994 and 2015 adopted a five-tier scale, denoted 5–1. The nomenclature ascribed to each score varied per evaluation cycle, with the highest number signifying excellence and the lowest reflecting poor or unsatisfactory performance (van Drooge et al., 2013). In the subsequent protocol (2015-2021), the scale was revised and reversed, leading to four categories: world leading/excellent (1); very good (2); good (3); unsatisfactory (4) (VSNU et al., 2016). The most intriguing aspect compared to Brazil is that the Dutch model attributed scores independently for each criterion. In the protocols up to 2015, they were quality, productivity, relevance, and viability. In the 2015-2021 SEP, they became research quality, societal relevance, and viability. Although the average score was also a result of the evaluation, the individual score for each dimension was an integral part of the observable results.

Furthermore, the Dutch evaluation process aims to preserve and strengthen the values and purposes of academic activity, especially autonomy, academic freedom, scientific quality, societal relevance, transparency, and participation of the academic community (van Drooge, 2021a; van Drooge et al., 2013; VSNU et al., 2020). Although universities and the research system are mainly financed with public resources, the country has not implemented a performance-based university research funding system as described by Hicks (2012) and Ochsner et al. (2018), so there is a clear separation between evaluation and funding (van Drooge et al., 2013; VSNU et al., 2020). In this sense, the results of a unit in the evaluation are not linked to its accreditation.

In the Netherlands, evaluation is not seen as an instrument of external control, state regulation, or accountability. Funding institutions may help guide evaluation, such as with NWO's involvement in the SEP design, but HEI are empowered to conduct their evaluations and determine what to do with the results. This means that the protocols do not define rules and guidelines on the consequences of evaluation and no national institution is responsible for defining or applying sanctions, rewards, and incentives. That is a prerogative of each HEI, following their internal policies. Therefore, the assessment is not regulatory or aims to establish classifications and create indicators of comparability among units. Evaluation respects differences and specificities, and instead of standardising and homogenising the system, it preserves diversity and promotes differentiation (van Drooge et al., 2013; VSNU et al., 2020).

As for the influence of the national evaluation system on researchers' behaviour, our analysis is that very little direct influence can be seen in the Netherlands. However, the academic system in the country is strongly influenced by the constant search for funding streams, permanent or tenured positions, and external project resources. In this sense, the pressure imposed by the national evaluation of Brazil on researchers is seen for other reasons, which are mostly absent in the Latin American country, as previously explored by Brasil (2020).

4.2.4 Transparency and controversies

Table 4.4.: Stability of the Brazilian and Dutch evaluation systems

Categories	Brazil	The Netherlands
Changes over time	Fluctuates	Few
Transparency	Strong	Weak
Controversies	Fluctuates	Few

Source: Brasil and Trevisol (2023), Scholten et al. (2018), and VSNU et al. (2020).

Since the 1970s, evaluation has played a strategic role within the Brazilian System of Research and Graduate Education (SNPG). After nearly five decades of continuous operation, the evaluation system has earned the recognition of the academic community for its importance, reliability, and strategic role

(Balbachevsky, 2005; C. B. Martins, 2018; Saviani, 2020; Verhine and Dantas, 2009). An open letter from the Brazilian Society for the Advancement of Science (SBPC) and the Brazilian Academy of Sciences (ABC) supports the system in place, stating that CAPES evaluation was the main protagonist in strengthening the Brazilian graduate system, decisively contributing to the growth and quality improvement of the country's science in all areas of knowledge. Without the evaluation system, Brazil does not have evidence to guide future decisions for the further development of the SNPG (Davidovich and Ribeiro, 2021).

As evaluation is periodic in Brazil, every new cycle traditionally includes advances in processes, methods, and data. However, change leads to new problems or to the recognition of persisting flaws in need of evolution. Advancement plans are usually created after every evaluation, being the result of collaboration between CAPES, evaluation area committees, and the scientific community. These plans often consist of incremental evolution measures, but some major changes may occur each few cycles. For instance, in 1998 CAPES implemented the Qualis journal classification system, which significantly shaped the evaluation over the following decades (Brasil, 2023a). However, the need for improvement does not affect the legitimacy of the evaluation process as a whole or its recognition by both the evaluators and those evaluated regarding the modus operandi and the consequences of the process.

The situation in the Netherlands is only slightly different, as evaluation protocols are also reviewed at every cycle, which is six years instead of four years in Brazil. Reviews aim to improve regulatory frameworks and establish new guidelines for the next cycle. Although dimensions and evaluation criteria have been adapted over the decades, the system is considerably stable, and changes introduced over time were incremental. In this sense, there are clear lines of continuity between regulatory frameworks.

As an example of these changes, the SEP 2009-2015 revised the concept of "relevance", which became "societal relevance". In addition to scientific impact, the evaluation should consider contributions from research to the development of other areas such as economy, innovation, culture, public management, etc. (van Drooge et al., 2013; VSNU et al., 2009). Similarly, the SEP 2015-2021 reduced the number of analysis dimensions from four to three, excluding "productivity" from the list. Since the change, the evaluation has begun to take into account three dimensions: quality, societal relevance, and viability (VSNU et al., 2016). In addition to this, the SEP 2021-2027 has abolished the use of impact factors and considerably limited the use of the h-index as possible measures for evaluation (VSNU et al., 2020). All these changes have derived from the influence of significant international movements towards responsible evaluation, in part represented by DORA, the San Francisco Declaration on Research Assessment (ASCB, 2012), and the Leiden Manifesto for research metrics (Hicks et al., 2015).

Regarding transparency, in an open letter to the CAPES leadership, the president of SBPC, Renato Janine Ribeiro (2022a), states that "CAPES Quadrennial Evaluation seeks the greatest possible transparency in the definition and exposure of the criteria used, as well as in the disclosure of the data on which the criteria are applied. Mechanisms such as area documents, Qualis, the Sucupira Platform, and the Lattes Platform were major steps in the search for this transparency". The letter praised transparency in the face of controversy, which would linger over evaluation and science policy during the 2019-2022 period.

Brazil emerged from decades of mostly uncontroversial evaluation to face a difficult period after the 2017 Quadrennial Evaluation. Until then, most controversies were localised, focused on methods, and raised by PPG unhappy with the results, especially those that were closed as a consequence of low grades. After 2018, following a significant effort by CAPES and the scientific community to improve the evaluation for the 2017-2020 period, a series of developments kept the system in check, leading even to the legal suspension of the quadrennial evaluation, originally planned for 2021.

The suspension came from the initiative of the Federal Justice system (2021), which questioned the evaluation process, adding to many controversial episodes never before seen in the Brazilian science system, including successive cuts in the number of grants and scholarships, particularly in the social sciences, humanities, and basic sciences (Saldaña, 2020); budget cuts that led CAPES to delay payments to thousands of scholarship holders (Saldaña, 2021a); and definitive suspension of funding for the 101 National Institutes of Science, Technology and Innovation (INCT) in the country (Holanda, 2021).

Furthermore, dozens of scientific societies throughout the country protested the appointment of new leadership to CAPES (SBPC, 2021b), and more than 100 members of peer review committees, including the coordination of several evaluation areas, resigned between 2021 and 2022 (Machado, 2022). In their resignation letter, the 29 members of the chemistry committee state that the success of the country's graduate system comes from its approach as a state policy rather than as a government policy. However, they justified their departure, in part, by the fact that this claim was no longer true (Monteiro et al., 2021).

The Quadrennial Evaluation of 2017 was marked by transparency. A dedicated website was created to cover the evaluation while it was being conducted and even included the data sets that the evaluation committees would use to guide their work (CAPES, 2017a). The underlying data in question had been collected by advanced current research information systems (CRIS), with transparent methods and auditable processes to support the evaluation. The results, as well as the analytical processes behind them, were available beforehand, allowing those evaluated to verify not only their own data, but the entire system, going even beyond what is proposed by the related principles of the Leiden Manifesto (Hicks et al., 2015). The Qualis classification of journals, which is significant in the evaluation results, was also published before the evaluation started.

The Quadrennial Evaluation of 2021, held after a year of legal delays, lost some of that transparency. The dedicated website was shut down and the evaluation section on the CAPES website did not include the data sets for the evaluation to come (CAPES, 2021b), although some of the information could be found in the public interface of the Sucupira Platform (CAPES, 2021c). The 2017-2020 Qualis classification with its adopted methodology were not released until after the evaluation results were published (CAPES, 2023c).

Although controversies dominated the Brazilian evaluation in the past few years, The Netherlands had few problems in that area. While low-stakes evaluation may be a mandatory and formative part of the national science system, it has limited impact on research units. The lack of direct consequences for funding, accreditation, and other relevant aspects places much less pressure on evaluation. Furthermore, the modifications in the most recent evaluation protocol align with national movements on reward and recognition in science, which have been expressed in initiatives such as the one described in the position paper "Room for everyone's talent" (VSNU et al., 2019). In that sense, leaving the use of indicators such as the impact factor and the h-index behind may raise opposing voices, especially among those who had to build their careers under the publish or perish threat. However, universities continue to move toward a more multidimensional approach to institutional and individual evaluations.

Finally, returning to the transparency front, that is an issue in which The Netherlands has not yet achieved a representative level in its processes. As the evaluation is decentralised, there is no central information source to support the research units in their evaluation, and only the self-assessment report (with select appendices) should be made public on the higher education institutions' websites. Finding and collecting them is a difficult, if not impossible, task that makes it unlikely that an overview of the quality of research in the country can be drawn from the evaluation. Furthermore, it is not possible to know which institutions have implemented measures to remedy the deficiencies identified by the external committee (van Drooge et al., 2013).

4.3 Conclusions

In this investigation, we have closely examined and compared research evaluation systems in Brazil and the Netherlands, focussing on their main objectives, methodologies, and the consequences of the evaluation results. This study was partially motivated by Brazil's search for inspiration in international models to help improve its own. Among these models is the Dutch system, known for its long-standing stability and the significant role it has played in the development of the country's science system.

The Brazilian evaluation, led by CAPES, is primarily performance-based, which influences the researchers' way of working and limits the evaluation design that can be implemented (Hicks, 2012). In Brazil, the high-stakes model affects more than funding; it also refers to the accreditation of assessed units, which may be revoked in cases of sub-par performance. Trust remains a relevant issue, and evaluation is perceived as an audit procedure, leading to a focus on quantitative methods and performance indicators (Ràfols et al., 2016). This may encourage research homogenisation and discourage experimentation with innovative approaches (Geuna and Martin, 2003).

In contrast, the Dutch evaluation system, guided by the Strategy Evaluation Protocol (SEP), is formative and low stakes, aiming to maintain and improve research quality and societal relevance while empowering higher education institutions (HEIs) to carry out their evaluations and determine how to use the results. This means that the Dutch formative evaluation, based on self-assessment practises that allow for a more contextualised and multidimensional evaluation that promotes diversity and differentiation, would face significant challenges in Brazil without a reorientation of the assessment objectives.

The main takeaway from examining these distinct evaluation systems is that their primary characteristics are firmly rooted in the core decisions underlying each of them, which may have been shaped by historical trajectories, geographical challenges, and policy initiatives that have influenced the evolution of science, research, and graduate education in each country. The result is that no analytical category presented in this comparative study can be interpreted in isolation, as each of them has dependencies and consequences, creating an interconnected mesh that cannot simply be unmade. Therefore, while some inspiration could be drawn from the positive experiences of each country, no part of these evaluation systems can be seamlessly transposed into the other, even with adaptations in mind.

However, that does not mean that lessons can not be learnt from each other. Despite recent controversies and legal issues that have somewhat affected the transparency of the Brazilian evaluation, the country's high-stakes system has led to several positive developments regarding advanced current research information systems, open science, valorisation of nonbibliographic research output, and more. Some of these experiences may be helpful in inspiring the Dutch in the development of tools and strategies to further incorporate multidimensional assessment strategies into its evaluation system. Similarly, Brazil can look for inspiration in the Dutch self-assessment protocols, which matured over the past decades and that can help the country improve its approach to institutional governance and autonomy.

In conclusion, this comparative analysis of the Brazilian and Dutch research evaluation systems reveals the intricacies and challenges of tailoring evaluation mechanisms to suit the unique contexts of each country. Although the study emphasises that it is not feasible to simply copy and paste solutions from one system into another, identifying strengths and potential areas for improvement in each system can help inform and inspire the ongoing evolution of research evaluation strategies. Such advances in research evaluation will ultimately contribute toward developing more robust, flexible, and context-sensitive evaluation systems that support scientific excellence, innovation, and 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/

Balbachevsky, 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). Rio de Janeiro: Nova Fronteira.

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

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

Barroso, J. (2005). O Estado, a educação e a regulação das políticas públicas. *Educação & Sociedade*, *26*(92), 725–751. https://doi.org/10.1590/s0101-73302005000300002

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation*, *23*(2), 166–173. https://doi.org/10.1093/reseval/rvu002

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2021b). Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases. In *Proceedings of the 18th international conference on scientometrics & informetrics* (pp. 193–204). Leuven, Belgium: KU Leuven. Retrieved from https://bit.ly/3jOHQBR

Brasil, A. (2022). Multidimensionality through self-evaluation: from theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation*, (53). https://doi.org/10.22163/fteval.2022.546

Brasil, A. (2023a). Between bibliometrics and peer review: The evolution and challenges of Brazil's Qualis system. preprint. Retrieved from https://bit.ly/412rlnA

Brasil, A. (2023c). The dynamics of a national evaluation system: The Brazilian model and its role in developing research and graduate education in the country. *SocArXiv*. https://doi.org/10.10100/xxx.202xxxx

Brasil, A., & Trevisol, J. V. (2023). *ENRESSH country report: Brazil*. OSF. https://doi.org/10.17605/OSF.IO/ZH465

Brennan, J., & Shah, T. (2000). Quality assessment and institutional change: Experiences from 14 countries. *Higher Education*, *40*(3), 331–349. https://doi.org/10.1023/a:1004159425182

Cano, W. (2015). Crise e industrialização no Brasil entre 1929 e 1954: A reconstrução do Estado Nacional e a Política Nacional de Desenvolvimento. *Brazilian Journal of Political Economy*, *35*(3), 444–460. https://doi.org/10.1590/0101-31572015v35n03a04

Capano, G. (2010). A Sisyphean task: Evaluation and institutional accountability in Italian higher education. *Higher Education Policy*, *23*(1), 39–62. https://doi.org/10.1057/hep.2009.19

CAPES. (2009). Capes lança aplicativo gráfico para consulta de dados. Retrieved December 7, 2021, from https://bit.ly/3IsaOBw

CAPES. (2014). *Resolução nº 5, de 11 de Dezembro de 2014. Altera a periodicidade da avaliação do SNPG* (Resolution No. 5/2014). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3BK5Ywu

CAPES. (2017a). Avaliação Quadrienal 2017: Hotsite. Retrieved July 4, 2020, from http://avaliacaoquadrienal.capes.gov.br/

CAPES. (2017b). *Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal*. Diário Oficial da União. Brasília, DF.

CAPES. (2018a). Avaliação comparada da pós-graduação. Seminários: Repensando a Avaliação. Retrieved February 2022, from https://bit.ly/3HyExbc

CAPES. (2018b). Avaliação da produção intelectual de programas de pós-graduação. Seminários: Repensando a Avaliação. Retrieved February 2022, from https://bit.ly/3AYGFqb

CAPES. (2020d). Sobre as áreas de avaliação. Retrieved April 9, 2021, from https://bit.ly/3s0XB9f

CAPES. (2021b). Avaliação Quadrienal. Retrieved June 18, 2021, from https://bit.ly/35wv16S

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

CAPES. (2021d). *Portaria nº 122, de 5 de Agosto de 2021. Consolida os parâmetros e os procedimentos gerais da Avaliação Quadrienal de Permanência da pós-graduação stricto sensu no Brasil*. Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/384xDgE

CAPES. (2023c). Metodologia do Qualis Referência: Quadriênio 2017-2020. Retrieved April 12, 2023, from https://bit.ly/3my156n

Centraal Bureau voor de Statistiek. (2021). *Nederland in cijfers: Aan de hand van 36 vragen verbeeld*. Den Haag: CBS. Retrieved February 2022, from https://bit.ly/3gv46h6

Centraal Bureau voor de Statistiek. (2022). 40 percent international first-year students at Dutch universities. CBS. Retrieved May 2, 2023, from https://bit.ly/40Y9YEG

Clarivate Analytics. (2022). Web of Science Core Collection. CWTS in-house version. Retrieved from https://bit.ly/3rHmCJT

Cohen, A., & Steege, M. v. d. (1982). An Historical Overview of the State and Higher Education in the Netherlands. *European Journal of Education*, *17*(3), 271. https://doi.org/10.2307/1502985

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*. Brasília, DF.

Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and Mechanisms for Evaluating Government-Funded Research. *American Journal of Evaluation*, *28*(4), 437–457. https://doi.org/10.1177/1098214007308290

Cunha, L. A. (2007). *A universidade temporã: o ensino superior da colônia à era Vargas* (3rd ed.). São Paulo: Editora UNESP.

Cuppen, H. (2021). Opeens nieuwe naam voor vereniging universiteiten. Retrieved August 8, 2023, from https://bit.ly/468FByR

Cury, C. R. J. (2005). Quadragésimo ano do parecer CFE nº 977/65. *Revista Brasileira de Educação*, (30), 07–20. https://doi.org/10.1590/s1413-24782005000300002

Davidovich, L., & Ribeiro, R. J. (2021). Manifestação da ABC e SBPC contra suspensão da Avaliação Quadrienal 2021. Retrieved March 10, 2022, from https://bit.ly/3q12v8j

Elven, M. v. (2023). Hoe het Nederlandse hoger onderwijs populair werd bij internationals. Retrieved May 3, 2023, from https://bit.ly/44lwv10

European Ministers of Education. (1999). *Bologna Declaration*. Bologna. Retrieved from https://bit.ly/3OxS8lR

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.

Fávero, M. d. L. d. A. (2006). A universidade no Brasil: Das origens à reforma universitária de 1968. *Educar em Revista*, *22*(28), 17–36. https://doi.org/10.1590/s0104-40602006000200003

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

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

Galleron, I., Ochsner, M., Spaapen, J., & Williams, G. (2017). Valorizing SSH research: Towards a new approach to evaluate SSH research' value for society. *fteval Journal for Research and Technology Policy Evaluation*, *44*, 35–41. https://doi.org/10.22163/fteval.2017.274

Geuna, A., & Martin, B. R. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva*, *41*(4), 277–304. https://doi.org/10.1023/b:mine.0000005155.70870.bd

Goedegebuure, L., & Westerheijden, D. F. (1991). Changing balances in Dutch higher education. *Higher Education*, *21*(4), 495–520. https://doi.org/10.1007/bf00134986

Hammarfelt, B., & de Rijcke, S. (2014). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, *24*(1), 63–77. https://doi.org/10.1093/reseval/rvu029

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*(2), 251–261. https://doi.org/10.1016/j.respol.2011.09.007

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

Holanda, L. (2021). Capes encerra repasses de verbas a institutos e surpreende cientistas. *Agência Brasil*. Retrieved from https://bit.ly/3szBkTb

IBGE. (2021). Sistema IBGE de Recuperação Automática. SIDRA. Brasília, DF. Retrieved from https://sidra.ibge.gov.br/

INEP. (2022). *Censo da Educação Superior 2021: Notas estatísticas*. INEP/MEC. Brasília, DF. Retrieved May 2, 2023, from https://bit.ly/42h1Yje

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

Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation Systems: What are they and why study them? *Evaluation*, *14*(2), 157–169. https://doi.org/10.1177/1356389007087537

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. https://doi.org/10.1590/s1414-4077/s1414-40772020000200006

Machado, R. (2022). Mais 3 pesquisadores ligados à Capes renunciam e aumentam debandada. *Folha de São Paulo*. Retrieved from https://bit.ly/3MMCEdW

Marcovitch, J., Goldemberg, J., Axel-Berg, J., Righetti, S., dos Santos, S. M., de Oliveira, L. N., … Guimarães, J. A. C. (2018). *Repensar a universidade: desempenho acadêmico e comparações internacionais* (J. Marcovitch, Ed.). https://doi.org/10.11606/9788571661868

Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). *Revista Brasileira de Sociologia*, *6*(13), 9–26. https://doi.org/10.20336/rbs.256

Ministerie van Onderwijs, Cultuur en Wetenschap. (1992). *Wet op het hoger onderwijs en wetenschappelijk onderzoek*. OCW. Den Haag. Retrieved from https://bit.ly/33YmT1l

Molas-Gallart, J. (2012). Research Governance and the Role of Evaluation: a comparative study. *American Journal of Evaluation*, *33*(4), 583–598. https://doi.org/10.1177/1098214012450938

Molas-Gallart, J. (2014). Research evaluation and the assessment of public value. *Arts and Humanities in Higher Education*, *14*(1), 111–126. https://doi.org/10.1177/1474022214534381

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*. Brasília, DF: CAPES. Retrieved from https://bit.ly/2PmAf0q

Monteiro, A. L., Meneghetti, S. M. P., Suarez, P. A. Z., Rubira, A. F., Zarbin, l. J. G., Rezende, C. M. d., . . . Coltro, W. K. T. (2021). Carta aberta dos Coordenadores e Consultores da Área de Química na CAPES. Retrieved May 14, 2022, from https://bit.ly/3sBhVkO

NUFFIC. (2023). Higher education in The Netherlands. Retrieved May 2, 2023, from https://bit.ly/3oPGfAf

Ochsner, M., Kulczycki, E., & Gedutis, A. (2018). The Diversity of European Research Evaluation Systems. In *Proceedings...* (pp. 1235–1241). Leiden, NL.

Ochsner, M., Kulczycki, E., Gedutis, A., & Girkontaitė, A. (2019). The Diversity of Research Evaluation Systems. Vilnius, LT. Retrieved April 18, 2023, from https://bit.ly/3KL3HpQ

Ochsner, M., Kulczycki, E., Gedutis, A., & Peruginelli, G. (2020). National Research Evaluation Systems. In R. Ball (Ed.), *Handbook bibliometrics* (pp. 99–106). https://doi.org/10.1515/9783110646610-011

Ràfols, I., Molas-Gallart, J., Chavarro, D., & Robinson-Garcia, N. (2016). On the dominance of quantitative evaluation in peripherall countries: Auditing research with technologies of distance. *SSRN Electronic Journal*, 1–22. https://doi.org/10.2139/ssrn.2818335

Ribeiro, R. J. (2022a). Carta da SBPC à CAPES sobre audiência pública para debater a Avaliação Quadrienal. Retrieved from https://bit.ly/3Ki9QYx

Rijksoverheid. (2023). Tertiary (higher) education. Retrieved May 2, 2023, from https://bit.ly/3nmJChP

Rothen, J. C. (2018). A universidade brasileira na Reforma Francisco Campos de 1931. *Revista Brasileira de História da Educação*, *8*(17), 141–160.

Saldaña, P. (2020). Em meio a pandemia, governo Bolsonaro investe contra pesquisa em ciências humanas. Retrieved August 10, 2020, from https://bit.ly/2PHEUH8

Saldaña, P. (2021a). Capes tem déficit de R$ 124 mi para pagar bolsas de formação docente até fim do ano. *Folha de São Paulo*. Retrieved from https://bit.ly/3a0WGm1

Saviani, D. (2020). Meio século de pós-graduação no Brasil: Do período heróico ao produtivismo pela mediação de um modelo superior às suas matrizes. *Movimento - Revista de Educação*, *7*(14). https://doi.org/10.22409/mov.v7i14.46475

SBPC. (2021b). Entidades se manifestam em defesa da Capes. Retrieved May 14, 2022, from https://bit.ly/38xTKN8

Scholten, W., van Drooge, L., & Diederen, P. (2018). *Excellence is extra-ordinary: Thirty years of focus on excellence in Dutch science policy*. The Hague.

Sguissardi, V. (2006). A avaliação defensiva no 'modelo CAPES de avaliação': É possível conciliar avaliação educativa com processos de regulação e controle do Estado? *Perspectiva*, *24*(1), 49–88. Retrieved from https://bit.ly/2VKO88E

Siqueira, M. B. (2019). Sucupira - A Platform for the Evaluation of Graduate Education in Brazil. *Procedia Computer Science*, *146*(2019), 247–255. https://doi.org/10.1016/j.procs.2019.01.081

Sobrinho, J. D. (2003). Avaliação da educação superior: Regulação e emancipação. *Avaliação: Revista da Avaliação da Educacão Superior*, *8*(2), 31–47. Retrieved from https://bit.ly/3I5MR2m

UNL. (2022). Universiteiten van Nederland. Retrieved January 21, 2022, from https://www.universiteitenvannederland.nl/

van Drooge, L. (2021a). Research evaluation in context 1: Introducing research evaluation in the Netherlands. Retrieved January 2022, from https://bit.ly/3IsQZJE

van Drooge, L. (2021b). Research evaluation in context 2: One joint protocol, three criteria and four aspects. Retrieved January 2022, from https://bit.ly/3Atb67G

van Drooge, L., de Jong, S., Faber, M., & Westerheijden, D. (2013). *Twenty years of research evaluation* (tech. rep. No. 8). Rathenau Instituut. The Hague. Retrieved from http://bit.ly/3cXp4mK

Vereniging Hogescholen. (2023). Instroom, inschrijvingen en diploma's. Retrieved May 2, 2023, from https://bit.ly/3Hwirbd

Verhine, R. E. (2008). Avaliação da CAPES: Subsídios para a reformulação do modelo. In D. Mancebo, J. d. R. Silva Jr, & J. F. de Oliveira (Eds.), *Reformas e políticas: Educação superior e pós-graduação no brasil* (pp. 165–188). Campinas: Alínea.

Verhine, R. E., & Dantas, L. (2009). Reflexões sobre o sistema de avaliação da CAPES a partir do V Plano Nacional de Pós-graduação. *Revista de Educação Pública*, *18*(37), 295–310.

Verhine, R. E., & de Freitas, A. A. d. S. M. (2012). A avaliação da educação superior: modalidades e tendências no cenário internacional. *Revista Ensino Superior Unicamp*, (7), 16–39. Retrieved from https://bit.ly/2C9herj

Viana, W. F. (2018). *Sistema CAPES de avaliação da pós-graduação stricto sensu: Um estudo de caso da área de Administração Pública no Brasil* (Doctoral dissertation, Universidade de Brasília).

VSNU, KNAW, & NWO. (2003). *Standard Evaluation Protocol (2003-2009)*. The Hague. Retrieved from https://bit.ly/3grtaGP

VSNU, KNAW, & NWO. (2009). *Standard Evaluation Protocol (2009-2015)*. The Hague. Retrieved from https://bit.ly/3wujG32

VSNU, KNAW, & NWO. (2016). *Standard Evaluation Protocol (2015-2021)*. The Hague. Retrieved from https://bit.ly/3iJqtC3

VSNU, NFU, KNAW, NWO, & ZonMw. (2019). *Room for everyone's talent*. The Hague. Retrieved January 30, 2023, from https://bit.ly/3Jt9ilw

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

Weert, E. d., & Boezerooy, P. (2007). *Higher education in the Netherlands*. Center for Higher Education Policy Studies (CHEPS). Enschede, NL.

Westerheijden, D. F., Cremonini, L., Kolster, R., Kottmann, A., Redder, L., Soo, M., ... Weert, E. d. (2008). *New Degrees in the Netherlands: Evaluation of the Bachelor-Master Structure and Accreditation in Dutch Higher Education*. Ministerie van Onderwijs, Cultuur en Wetenschappen. Enschede, NL. Retrieved from https://bit.ly/3tNdAM4

5

The impact of graduate education on academic publishing

> *Our graduate education system is a source of national pride, a model that has been recognised throughout the world.*
>
> **— Helena Nader**

During his first year in office, Brazilian president Jair Bolsonaro (2019–2022) wrongly stated that few of the country's universities conducted research and that most of those engaged in science were private institutions. The nation's academic community responded immediately. For instance, the Brazilian Academy of Sciences (ABC) stated that it was of utmost importance to provide the president with correct information about the country's science system, highlighting findings from Clarivate Analytics (2019) that showed that more than 95% of Brazilian scientific production comes from public universities (Moura, 2019).

In the debate that followed, various actors in the Brazilian science system contributed with relevant information to better inform politicians and avoid misguided policy design, reporting that: (i) the country had more than 400 higher education institutions (HEI) engaged in research; (ii) the 30 universities that published the most in Brazil were public; (iii) predominantly public graduate education played a significant role in the country's research, with 93.167 new

This chapter has been published as: Brasil, A. (2023). The impact of graduate education on academic publishing: A comparative analysis from a Brazilian perspective. In *Proceedings of the 19th International Conference on Scientometrics & Informetrics*. Bloomington, Indiana: University of Indiana.

researchers graduating in that year: 24.290 PhD and 77.795 masters (Brasil, 2020; CAPES, 2021e; CGEE, 2019; Moura, 2019; SBPC and ABC, 2020b).

Much of the information publicised at that time was grounded in evidence, and the numbers helped support the correct narrative on the reality of the science of the country. However, recurring statements on the substantial contribution of graduate education to scientific production were largely accepted, but rarely supported. For example, in a seminar organised by ABC and the Brazilian Society for the Advancement of Science (SBPC), it was reported that approximately 95% of Science & Technology research in the country is carried out within its graduate programs (PPG), which are composed of master's and doctoral courses (SBPC and ABC, 2020b).

Extensive research on the literature and official data provided by the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) revealed that there are no concrete data to support these claims. Therefore, departing from the hypothesis that such empirically accepted claims properly reflect the reality of the Brazilian system, this study was carried out to verify, through a bibliometric analysis, the extent of the role of graduate education in the Brazilian science system. A comparative exploration of international practises was also carried out to verify how different the Brazilian graduate experience is from that of other countries.

5.1 Additional Background

As reported by Herculano-Houzel (2013), the "scientist" profession does not exist legally in Brazil. Legislation to regulate the profession was proposed in 2015 in the national parliament, arguing that the country's science would not reach its potential impact since it relied on professors who teach as a main activity and research in spare time, or by masters, PhD and postdoctoral students who are not paid or who are awarded a scholarship with a value that is incompatible with their level of education or the expected work regime (Tenório, 2015).

Despite the parliament initiative, science was not regulated as a profession in Brazil. Not only did the proposed law not receive enough public support to advance in the legislative process, but it also faced what appeared to be unexpected opposition. In a public hearing at the parliament, representatives from HEI and scientific societies, such as SBPC and ABC, expressed concerns that regulating the profession could lead to serious negative consequences. The main fear is that strict labour regulations could lead to bureaucratization of the activity, affecting the training of highly qualified human resources. On the occasion, the SBPC's president also highlighted the graduate model as a source of national pride, being one of the most prominent determinants of growing scientific production in the country. Once again, the role of graduate education was central to the discussion (Monteiro, 2015; Tenório, 2015).

The mid-20th century design of the Brazilian National System of Research and Graduate Education (SNPG) has been comprehensively described by Brasil (2020), who reported how the country's science system was the object of public policy linking research and education in nearly inseparable ways. One of the direct consequences of this fact has been reflected in the work of Balbachevsky and Schwartzman (2010). Inspired by Burton R. Clark (1993) "The research foundations of graduate education", the authors discussed the Brazilian case with an alternative title: "The graduate foundations of research in Brazil". The name change was intentional, as it reveals how the Brazilian account differs from most, if not all, other experiences worldwide.

However, is the historical design of the SNPG and its empirical perception by academics reflected in the scholarly publication practises in the country? This research in progress aims to find bibliometric answers to that question.

5.2 Methods and Data

The first approach adopted to verify the contribution that graduate education in Brazil makes to the overall scientific output of the country was to match publications from the Web of Science with the reported output from Brazilian graduate programs. For this, the CWTS internal version of the WoS Core Collection (Clarivate Analytics, 2022) was used, as the Centre for Science and Technology Studies makes significant investments to enrich the data set with improved affiliation data, which makes a more robust country identification of authorships possible.

For the Brazilian data set, the country conducts a high-stakes national evaluation through a government agency, CAPES. The agency implemented a data collection system to support its evaluation process, and the resulting data set is robust and reliable due to the impact evaluation has on funding and the renewal of accreditation of PPG; it is in their interest to provide accurate and comprehensive information (Brasil, 2020). At the time of this research, the most recent microdata available from CAPES included publications up to 2018, so that year was selected for the matching process.

With the WoS and CAPES datasets available, all publications from 2018 with at least one author of Brazilian affiliation were extracted from the first and matched to the second. The process was based on an algorithm described in detail in a previous study, considering DOI, author names, titles, ISSN of the journal, volumes, issues, and page numbers (Brasil, 2021b). The matched publications were then analysed according to a broad area classification adopted by CAPES and detailed by Brasil (2022).

Although the first approach identified the percentage of WoS-indexed papers reported as graduate program output in Brazil, a second approach investigated the rate publications from different countries in relation to their PhD graduates per year. The CWTS in-house version of WoS was used once again, and data on PhD graduates per country were collected from Eurostat (2022) and OECD (2021), complemented with information from CAPES (2021e), the Ministry of Education: People's Republic of China (2022), and the US Census Bureau (2020). Due to the range of data available on PhD graduates per country, this analysis was conducted for the period 2013-2020. A limitation of this approach is that master's courses contribute significantly to the development of the Brazilian science system (Brasil, 2020), but the number of graduates at this level was not considered in the analysis, as there is a lot of heterogeneity in the modality between countries, especially since the changes in Europe sparked by the Bologna Declaration (European Ministers of Education, 1999).

5.3 Results and Discussion

The Web of Science indexes 53.604 articles and reviews published in 2018 by authors with Brazilian affiliation. When these individual publications are matched to the country's database of graduate programs' scholarly output, 40.634 records coincide. That means that more than 75% of the Brazilian publications indexed by WoS in that year are reported by PPG as results of their research. That is a significant share of the indexed output, which is potentially even higher, as CAPES records sometimes lack detailed information on publications (e.g., DOI, ISSN, journal issue).

Considering that Brasil (2021b) showed that coverage of Brazilian academic publishing in the Web of Science is unbalanced between different research areas, an additional analysis of the publications was performed comparing WoS papers according to the major classification of the Organisation for Economic Co-operation and Development (OECD) with the broad area classification system adopted by CAPES for the country's graduate programs.

Figure 5.1 displays the results of this analysis, using fractional counting on both sides of the visualisation. For the OECD classification, this means that the weight recorded in WoS was considered in the distribution seen on the left of the figure. For the right side, authorships from different PPG were also considered (e.g., a publication coauthored by an author from a multidisciplinary program with another from an engineering one has been counted as 0,5 for each broad group).

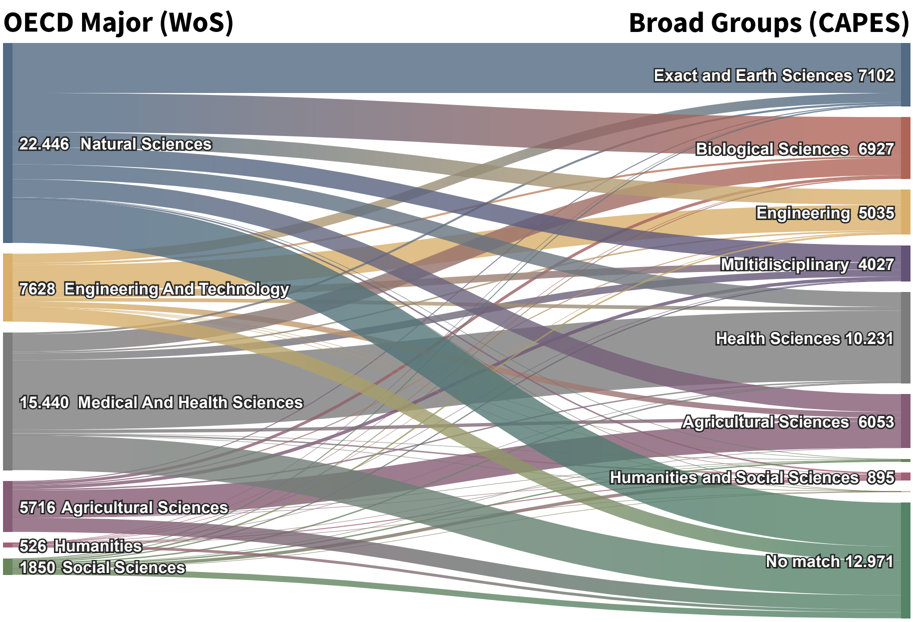


Figure 5.1.: Articles of Brazilian authors in WoS matched to PPG in the country, organized by CAPES broad group classification (2018).

The Figure 5.1 distribution of publications according to OECD's major classification aligns with the findings of Brasil (2021b), showing that the coverage of Brazilian output in WoS is high in the natural and life sciences, but below average in the social sciences and humanities (SSH). The results also reinforce the findings in Brasil (2023b), which shows that the classification system adopted in the country needs to be reviewed, as many graduate programs appear to conduct research in areas not aligned with the local classification.

Taking into account the 12.971 publications without a match in Figure 5.1, a disproportionate distribution according to the OECD classification could indicate that scientific publishing occurs outside of academia. For instance, if numerous papers come from engineering and technology, that could include publications from industry authors. However, the distribution seems proportional in all areas, so two further steps are planned for future analysis: an improvement of the matching algorithms, and a sample analysis of the unmatched ones to identify additional authorship sources (e.g., from within universities but originating from outside graduate programs).

Complementing the perspective of the matching analysis, Figure 5.2 shows the average publication rate per country in relation to the number of PhDs they graduate each year, from 2013 to 2020. The interactive visualisation available at https://tabsoft.co/3i6XJES can be filtered to display the results per year, so any fluctuations can be observed.

Figure 5.2 reveals that most countries have a rate of 0,1 to 0,2 graduating PhDs per publication. For instance, the USA had a per-year average of around 430k papers indexed by WoS, and of 55k graduated PhDs. The respective numbers for China are 360.000 and 58.000; for the UK they are 130.000 and 27.000; and for the Netherlands they are 43.000 and 4600. These and most countries included publish four to ten articles per PhD graduating each year.

However, the situation in Brazil is substantially different. As can be seen, the rate for the country is above 0,4, with an average number of 49.000 indexed papers and around 20.000 new PhDs graduating every year. Figure 5.2 also shows that Mexico and Russia have an even higher rate of graduates per publication, reaching nearly 0,8 in the Russian case (38.000 publications for 28.000 graduates). The graduate systems of both countries are already being studied to understand their dynamics, and the results will be reported in the final version of this study.

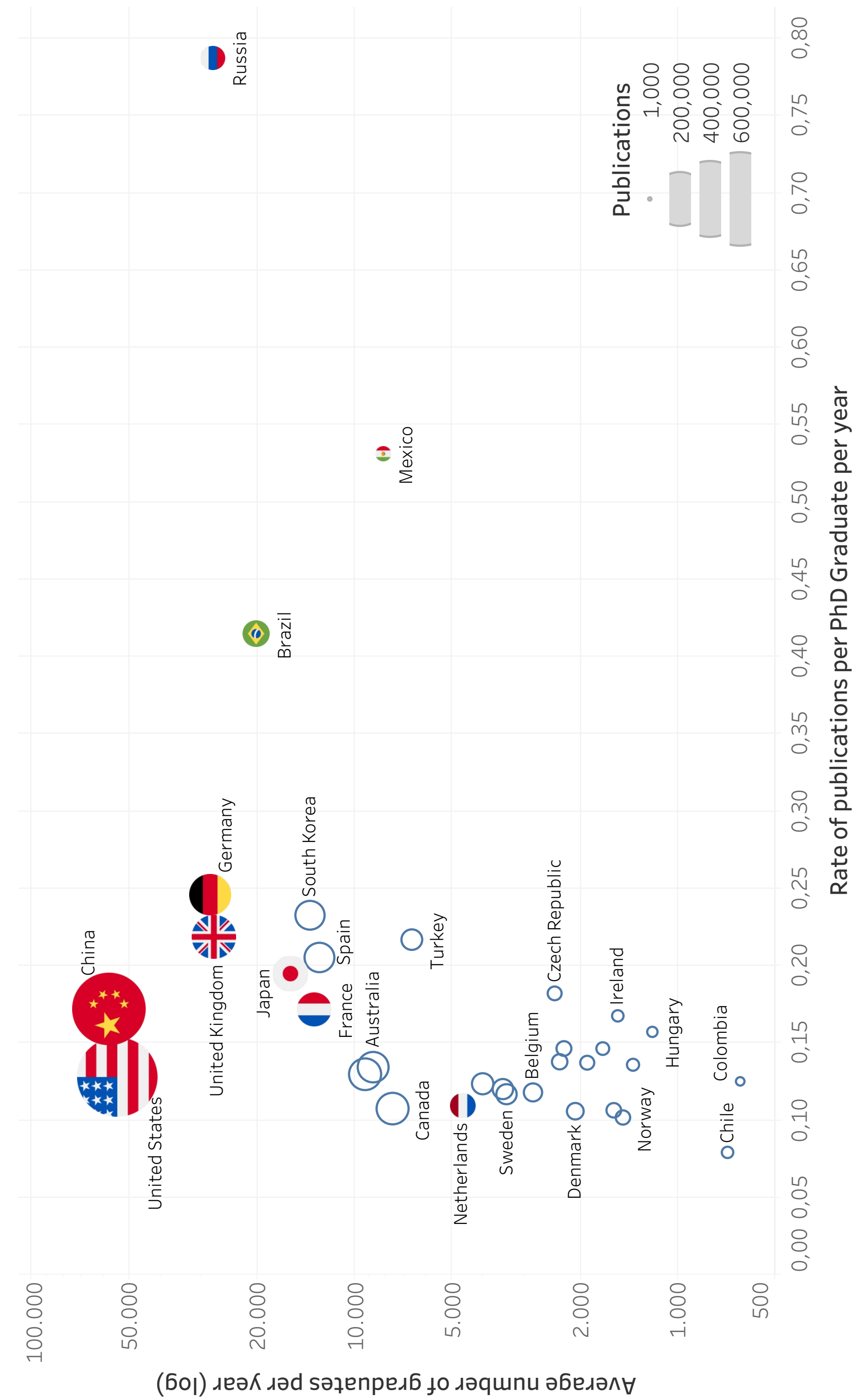


Figure 5.2.: Average number of PhD graduates per year in relation to the yearly rate of publications per PhD graduate (2013-2020).

In the case of Brazil, the analysis may suggest that graduate education has indeed a larger role in the national science system than what can be seen in most countries, supporting the empirical perception of Brazilian academics. However, the indication is still far from being considered a solid conclusion. Low levels of English literacy may make publishing in WoS harder for Brazilians, explaining partially the results. However, that same effect would probably be visible for some of the other Global South countries listed, which does not seem to be the case. The same reasoning could be used in regard to other challenges to publish in the WoS core collection, or regarding practices to publish in journals indexed by local databases such as SciELO.

5.4 Conclusion and next steps

This research in progress aims to produce evidence around a recurring argument of Brazilian academics who state that the country's graduate education can be considered the main or even the single pillar of the country's science system. A first analysis of legislation, policy documents, and related literature shows that the Brazilian research system was designed around a broader graduate education. That perception seems to be contrary to how graduate systems have been implemented around the world, where they are often one dimension of a broader research system. As an ongoing part of this study, a comprehensive literature analysis is being performed for a comparative perspective, including additional records from the Brazilian case, but also covering other countries such as China, Germany, Japan, Mexico, The Netherlands, South Africa, UK, USA (Bawa, 2008; Becher, 1993; Ben-David and Zloczower, 2009; Deem and Dowle, 2020; Gaughan and Robin, 2004; Gellert, 1993; Kobzar and Roshchin, 2020; Kyvik and Tvede, 2010; Lim et al., 2020; Nerad, 2008; Schwartzman et al., 2015; Westerheijden et al., 2008).

Regarding the bibliometric analyses, a first exploration has revealed that at least 75% of all WoS-indexed output from authors of Brazilian affiliation are reported as PPG research output. It was noticed that there appears to be no sign of research areas that register numerous contributions from authors outside of academia. The initial analysis was performed with 2018 data, but as algorithms are refined and more recent data become available, it will be possible to expand the full study data set over a four-year period (2017-2020).

Using data from the Web of Science and multiple international sources on PhD graduates per country, it was also possible to identify that Brazil has a high rate of graduates per publication, which may reinforce the idea that graduate education plays a central role, above world average, in Brazilian science. Preliminary results may indicate there is truth behind de claims, but further investigation is necessary, including the analysis of additional databases with broader coverage of Brazilian output (e.g., Scopus or Dimensions), addition of more Global South countries that may face similar challenges to publish in international journals, and also including master's students in the comparisons, as they are also a driving force in Brazilian graduate education.

References

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

Bawa, A. (2008). South Africa. In M. Nerad & M. Heggelund (Eds.), *Toward a global phd? forces and forms in doctoral education worldwide* (pp. 117–130). Seattle, WA: University of Washington Press.

Becher, T. (1993). Graduate Education in Britain: The View from the Ground. In B. R. Clark (Ed.), *The research foundations of graduate education: Germany, britain, france, united states, japan* (pp. 115–153). Berkeley, CA: University of California Press.

Ben-David, J., & Zloczower, A. (2009). Universities and academic systems in modern societies. *European Journal of Sociology*, *3*(1), 45–84. https://doi.org/10.1017/s0003975600000527

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2021b). Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases. In *Proceedings of the 18th international conference on scientometrics & informetrics* (pp. 193–204). Leuven, Belgium: KU Leuven. Retrieved from https://bit.ly/3jOHQBR

Brasil, A. (2022). Multidimensionality through self-evaluation: from theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation*, (53). https://doi.org/10.22163/fteval.2022.546

Brasil, A. (2023b). Rethinking a national classification of research and graduate education. *Profesional de la Información*, *32*(2), e320224. https://doi.org/10.3145/epi.2023.mar.24

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

CGEE. (2019). *Brasil: Mestres e Doutores 2019*. CGEE. Brasília. Retrieved November 26, 2022, from https://mestresdoutores2019.cgee.org.br

Clarivate Analytics. (2019). *Research in Brazil: Funding excellence*. Retrieved from https://bit.ly/3gdf5vY

Clarivate Analytics. (2022). Web of Science Core Collection. CWTS in-house version. Retrieved from https://bit.ly/3rHmCJT

Clark, B. R. (Ed.). (1993). *The research foundations of graduate education: Germany, Britain, France, United States, Japan*. Berkeley, CA: University of California Press.

Deem, R., & Dowle, S. (2020). The UK Doctorate: History, Features and Challenges. In M. Yudkevich, P. G. Altbach, & H. d. Wit (Eds.), *Trends and issues in doctoral education: A global perspective* (1st ed., pp. 152–178). https://doi.org/10.4135/9789353885991.n7

European Ministers of Education. (1999). *Bologna Declaration*. Bologna. Retrieved from https://bit.ly/3OxS8lR

Eurostat. (2022). Graduates by education level, programme orientation, completion, sex and age. Retrieved from https://bit.ly/3yg5kXf

Gaughan, M., & Robin, S. (2004). National science training policy and early scientific careers in France and the United States. *Research Policy*, *33*(4), 569–581. https://doi.org/10.1016/j.respol.2004.01.005

Gellert, C. (1993). The German Model of Research and Advanced Education. In B. R. Clark (Ed.), *The research foundations of graduate education: Germany, britain, france, united states, japan* (pp. 5–44). Berkeley, CA: University of California Press.

Herculano-Houzel, S. (2013). A profissão cientista não existe no Brasil. Retrieved November 26, 2022, from https://youtu.be/CF8UeugZW7c

Kobzar, E., & Roshchin, S. (2020). Russian Doctoral Education: Between Teaching and Research. In M. Yudkevich, P. G. Altbach, & H. d. Wit (Eds.), *Trends and issues in doctoral education: A global perspective* (1st ed., pp. 127–151). https://doi.org/10.4135/9789353885991.n6

Kyvik, S., & Tvede, O. (2010). The Doctorate in the Nordic Countries. *Comparative Education*, *34*(1), 9–25. https://doi.org/10.1080/03050069828315

Lim, H., Kim, S. J., & Shin, J. C. (2020). Rapid Development and Current Rethinking in Doctoral Education in South Korea. In M. Yudkevich, P. G. Altbach, & H. d. Wit (Eds.), *Trends and issues in doctoral education: A global perspective* (1st ed., pp. 363–386). https://doi.org/10.4135/9789353885991.n15

Ministry of Education: People's Republic of China. (2022). Educational Statistics. Retrieved from https://bit.ly/3yiASfp

Monteiro, V. (2015). Regulamentação da profissão de cientista em debate. *Jornal da Ciência*. Retrieved from https://bit.ly/3GNrm8s

Moura, M. (2019). Universidades públicas realizam mais de 95% da ciência no Brasil. Retrieved November 26, 2022, from https://bit.ly/3gEVbxw

Nerad, M. (2008). United States of America. In M. Nerad & M. Heggelund (Eds.), *Toward a global phd? forces and forms in doctoral education worldwide* (pp. 278–297). Seattle, WA: University of Washington Press.

OECD. (2021). Education at a Glance 2021. *Education at a Glance*. The data set is available at https://bit.ly/3nePKoB. https://doi.org/10.1787/b35a14e5-en

SBPC, & ABC. (2020b). Qual o papel da CAPES na construção da ciência e da pesquisa no Brasil? [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

Schwartzman, S., Pinheiro, R., & Pillay, P. (2015). *Higher education in the BRICS countries* (S. Schwartzman, R. Pinheiro, & P. Pillay, Eds.). Dordrecht: Springer.

Tenório, E. (2015). Regulamentação da profissão "cientista"no Brasil. Retrieved November 26, 2022, from https://bit.ly/3AOij3k

US Census Bureau. (2020). Educational attainment in the United States: 2019. Retrieved from https://bit.ly/3dUYh9a

Westerheijden, D. F., Cremonini, L., Kolster, R., Kottmann, A., Redder, L., Soo, M., … Weert, E. d. (2008). *New Degrees in the Netherlands: Evaluation of the Bachelor-Master Structure and Accreditation in Dutch Higher Education*. Ministerie van Onderwijs, Cultuur en Wetenschappen. Enschede, NL. Retrieved from https://bit.ly/3tNdAM4

Part III

Strenghts and weaknesses in Brazilian evaluation

6

Rethinking a national classification of research and graduate education

> *After many centuries of constructive but yet inconclusive search for a perfect classification scheme, the only sensible approach to the question appears to be the pragmatic one: what is the optimal scheme for a given practical purpose?*
>
> **— Wolfgang Glänzel & András Schubert**

The Brazilian science system exists primarily within graduate programs (PPG), composed of master's and doctoral levels. This design was not an accident, but a consequence of a science system that did not develop spontaneously; it was the object of public policy that prioritised the link between research and education. Most of that effort took place from the 1950s, initially by shaping the system and then towards its expansion (Balbachevsky and Schwartzman, 2010; Brasil, 2020). One of the key strategies adopted was to implement a scholarship system to allow Brazilians to pursue degrees abroad, with the aim of building critical mass to materialise the country's graduate education(CNPG, 1974; Gouvêa, 2012).

In the early 1970s, the number of scholarship holders from the leading agency in charge of funding the research and graduate system in the country, CAPES,

This chapter has been published in English, Spanish and Portuguese, and it is available at: Brasil, A. (2022). Rethinking a national classification of research and graduate education. *Profesional de la información*, 32(2), e320224. https://doi.org/10.3145/epi.2023.mar.24

was significantly small. The agency's grant report from 1971 revealed that only 1831 scholarships were awarded for graduate students in the country, and an additional 134 were granted to study abroad. Because of the manageable numbers, scholarships were handled mainly by a deliberative council that would analyse candidates in the face of the available funding. According to Darcy Closs, CAPES' executive director from 1974 to 1979, this process was particularly challenging, as many national figures would pressure the council to award grants to proteges (Castro and Soares, 1983; Córdova, 2001; Ferreira and Moreira, 2002).

To avoid lobbying, CAPES sought inspiration from the peer review experience of accreditation agencies in the United States. The first effort in Brazil, still in 1974, consisted of the installation of a single peer review committee with a small group of experts from broad areas such as engineering and social sciences. Academic merit would guide decisions on scholarship distribution, and the list of awardees would be submitted to the minister of education for endorsement. The task at hand was beyond the certification of the results, as the real challenge was neutralising the complaints of influential people who had their requests denied. Reports on the initiative of the advisory committee acknowledge that positive results were only possible due to the performance of the minister in protecting the newly established merit system (Ferreira and Moreira, 2002).

However, the single committee would not be able to keep up with the number of scholarships granted every year, which grew more than 400% in less than a decade (Castro and Soares, 1983). Therefore, two significant changes were implemented: (i) The evaluation evolved to an institutional model, where CAPES would assess graduate programs instead of individual candidates, granting a quota of scholarships to the programs based on performance. Then, the PPG would distribute the scholarships based on internal criteria; (ii) The original advisory committee was transformed into a series of disciplinary committees, which multiplied according to the growth of graduate programs and consequent increase in demand (Córdova, 2001; Ferreira and Moreira, 2002).

According to the most recent official reports, in early 2021, there were 4691 graduate programs active in Brazil, and CAPES granted around 95.000 scholarships for master's and doctoral courses in the country and 4500 for study and research abroad. The distribution of these scholarships is still heavily based on the evaluation performance of graduate programs, with a system organised around 49 evaluation areas, developed from the original disciplinary peer review committees (CAPES, 2020e; CAPES, 2021e).

This study looks at the current evaluation areas and analyses whether a reorganisation may be necessary. For that, we consider three different perspectives: (i) the dynamics of the expansion of evaluation areas and the observed inconsistencies in how they are organised; (ii) an international comparison of classification systems of research and education; (iii) a recommendation from a special committee in charge of monitoring the Brazilian National Plan for Research and Graduate Education (PNPG)[1]. Finally, after identifying weaknesses in the structure of the Brazilian evaluation areas, the study advances to propose a scientometric approach to rethink such areas and the distribution of graduate programs within them.

6.1 Evaluation areas and their roles

Evaluation areas are a core component of the established evaluation system in Brazil. Each area counts with its peer review committee, coordinated by representatives appointed by graduate programs in each discipline and nominated by CAPES for a four-year term. The coordinator's work is supported by two deputies, one for academic and the other for professional programs. Although broader regulations guide national evaluation, each area has some freedom to determine specific criteria and indicators in its analyses (CAPES, 2016a). For instance, as described in a previous study, areas can choose which types of technical and technological products should be recognised as appropriate research outputs valued by the committees in the evaluation process (Brasil, 2021a).

The configuration in 49 areas also plays a pivotal role in the organisation of the science system. Accreditation of new graduate programs is mandatory in the country. Once a proposal is approved, the new PPG becomes part of the corresponding area, subject to their specific evaluation criteria. Additionally, every four years, accreditation must be renewed in a national evaluation that is comparative within each area. PPGs are graded on a scale of 1 through 7 based on how well they perform compared to the overall performance of the other programs in the same areas (Brasil et al., 2022; CAPES, 2021d).

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

The evident relevance of the evaluation areas is established even in related legislation, where they are given the responsibility to guide CAPES' programs and courses of action (CAPES, 2016a).

Table 6.1 shows CAPES evaluation areas, with unique identifiers in parentheses, aggregated into the nine broad areas and three upper groups adopted by the agency (CAPES, 2020e).

Table 6.1.: CAPES evaluation areas with respective broad areas and upper groups

Upper Group	Broad area	Evaluation area
Exact Sciences	Engineering	Engineering I (10), Engineering II (12), Engineering III (13), Engineering IV (14)
	Exact and Earth Sciences	Astronomy and Physics (03), Chemistry (04), Computer Science (02), Earth Sciences (05), Mathematics and Statistics (01)
	Multidisciplinary	Biotechnology (48), Environmental sciences (49), Interdisciplinary (45), Materials Science (47), Teaching and Learning (46)
Humanities	Applied Social Sciences	Architecture, Interior and Industrial Design (29), Business and Administration, Accounting and Tourism (27), Economics (28), Journalism and Information (31), Law (26), Social Work (32), Town Planning and Demography (30)
	Humanities	Anthropology and Archaeology (35), Education (38), Geography (36), History (40), Philosophy and Ethics (33), Political Science and International Relations (39), Psychology (37), Religion and Theology (44), Sociology (34)
	Linguistics, Literature and Arts	Arts (11), Literature and Linguistics (41)
Life Sciences	Agricultural Sciences	Agricultural Sciences (42), Food Science and Technology (25), Veterinary Medicine (24), Zootechnics and Fisheries (23)
	Biological Sciences	Biodiversity (07), Biological Sciences I (06), Biological Sciences II (08), Biological Sciences III (09)

Continue...

Table 6.1 Continued

Upper Group	Broad area	Evaluation area
	Health Sciences	Dental studies (18), Medicine I (15), Medicine II (16), Medicine III (17), Nursing (20), Nutritional Science (50), Pharmacy (19), Physical Education, Therapy and Rehabilitation (21), Public Health (22)

Source: CAPES (2020e)

Although the names of some evaluation areas shown in Table 6.1 are very descriptive, such as “Environmental Sciences” or “Computer Science”, others are more difficult to understand unless subareas or specialities are considered. For example, CAPES (2020e) shows that electrical and biomedical engineering are subareas included in “Engineering IV”, and that “Medicine I” aggregates specialities such as oncology and cardiology.

In addition to cryptical names, some areas combine broader sets of disciplines with different levels of affinity for their objects, cognitive methods, and instrumental resources. A significant example is in “Anthropology and Archaeology”, combining disciplines in a single evaluation area under the broad area of “Humanities”. The American Academy of Arts and Sciences, for instance, considers archaeology to be part of the humanities, and anthropology to be a social science, despite recognising its humanistic perspective (AAAS, 2022). In other occasions, some proximity appears to exist, like in the case of “Architecture, Interior and Industrial Design”. However, a comparative evaluation here becomes harder to perform due to quite distinct citation practises in those disciplines.

The evaluation area system designed by CAPES evolved over time, in part following the “cognitive” approach described by Glänzel and Schubert (2003), where areas can be iteratively defined according to the experience of those involved, in this case the agency’s experts and committee members. However, CAPES (2020e) states that the area classification also has an eminently practical purpose, aiming to provide research units with a functional way to report their activities to the science and technology agencies in the country. As a consequence of the administrative component involved in the delimitation process, an unnatural delimitation of areas becomes evident in the literature, for instance:

i) Dias et al. (2017) reviews the process in which the area "Teaching of Science and Mathematics" was created from the existing "Education" area. According to the authors, the new area was the consequence of a long political movement within the original area, where a group of researchers could not find autonomy and recognition. Their work focused on applied research toward improving the training of human resources, for all levels, through the improvement of teaching methods. Aiming to strengthen the connections between science and society, CAPES supported the creation of the new area, leading to a clear division of applied research in "Teaching of Science and Mathematics" and the more conceptual and theoretical research in "Education". Two decades later, the areas have evolved towards better integration of academic and professional research, and the borders between the areas are no longer clear. As a consequence, their leaders have been calling for either a redesign of the areas or their unification.

ii) CAPES' ordinance n.º 83 (2011) renamed the "Teaching of Science and Mathematics" area to "Teaching".[2] The ordinance also created other areas such as "Environmental Sciences", with research programs migrating from existing areas. However, an analysis of CAPES' database of existing programs in the "Interdisciplinary" area (CAPES, 2021c), for example, reveals that there are several PPG in that area that did not migrate to the new area, despite obvious connection. Some PPG in the "Interdisciplinary" area are even named "Environmental Science".

iii) Stern (2019) describes how the areas of "Philosophy and Ethics" and "Religion and Theology" were created in 2016 from the division of a single area. The author reports that, despite the epistemological differences between the areas, it took more than a decade of negotiations to achieve the desired separation. Ultimately, the new areas were only created after a political crisis: during the election of the coordinator of the original combined area, all research programs in "Religion and Theology" unified to support a single candidate, while no consensus was found within the "Philosophy and Ethics" ones. The philosophers called for CAPES to annul the election, which was denied, but that gave traction for the separation to finally happen.

[2] This research translates the original Portuguese term "Ensino" as "Teaching & Learning" according to commonly used international terminology.

Different types of stories can be told about how new evaluation areas have been created, and others have been combined or restructured over time. Those stories show how the Brazilian classification of evaluation areas was created with a purpose, and that its development aimed to address issues such as the expansion of the country's research and graduate education system, and the evolution of science. However, the main challenge regarding the CAPES' classification can be described by Glänzel and Schubert (2003, p. 1), who said that "after many centuries of constructive but yet inconclusive search for a perfect classification scheme, the only sensible approach to the question appears to be the pragmatic one: what is the optimal scheme for a given practical purpose?"

From this perspective, the main purpose of the classification system adopted by CAPES has been the evaluation of graduate programs in the country. Linked to this primary goal is the allocation of funding in a comparative perspective within each area, relying on metrics which often fail to capture the variation of disciplinary practises. Furthermore, the classification is also relevant to analyse the evaluation of the Brazilian science system in the international scenario, which also determines funding distribution.

6.2 The Brazilian classification compared

Assessing Brazilian science from the CAPES classification is particularly challenging, as adjustments made in the model to address local peculiarities have led to a significant mismatch with other classification systems, such as the OECD Fields of Research and Development (FORD) and the UNESCO International Standard Classification of Education (ISCED). Some of these inconsistencies are visible in Figure 6.1 and Figure 6.2, where broad areas adopted by CAPES were matched with the broad classifications of FORD and ISCED. For that, a multilevel analysis was performed based on areas, subareas, and specialities for the three systems (CAPES, 2020e; OECD, 2015; UNESCO, 2015).

Figure 6.1 shows the nine broad areas in the CAPES classification system on the left, with numbers representing the evaluation areas. Fractional numbers can be seen in the FORD part of the Sankey chart, as the areas or sub-areas of the Brazilian system may be divided into different groups as defined by OECD (2015). For the broad group of Biological Sciences, for example, some

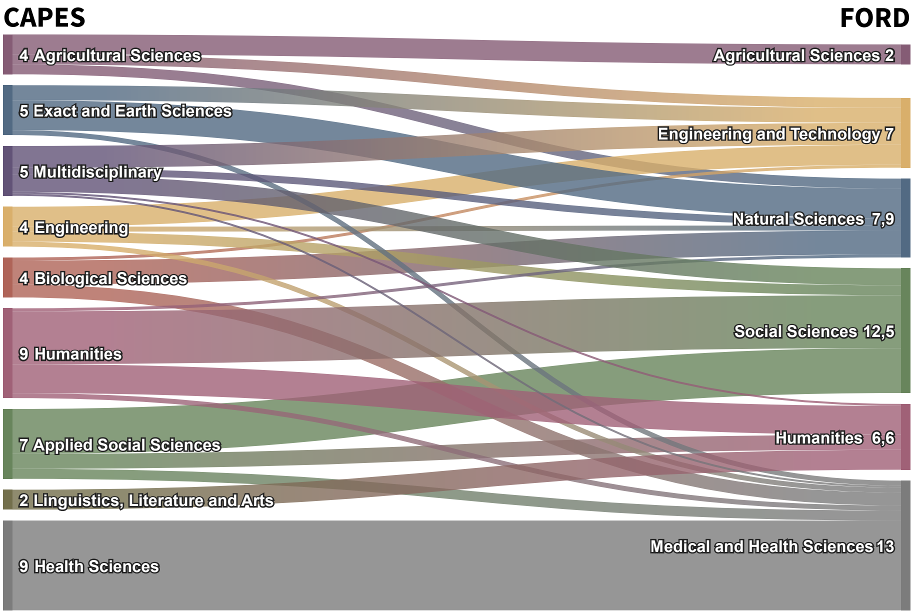


Figure 6.1.: CAPES broad area relations with FORD's broad classification.

subareas of "Biological Sciences III" fit into "Medical and Health Sciences" (for example, immunology and parasitology) and others belong to 'Engineering and Technology' in the FORD schema (e.g., cell & tissue engineering).

Another distinction between the two systems connected in Figure 6.1 is related to the social sciences and humanities, as inconsistencies can be seen in the distribution of groups among classifications. For instance, more than half of what CAPES considers part of the humanities is classified as social sciences by FORD (e.g., political science and psychology). One could argue that the observed conflicts may come from the design of the Brazilian system with graduate education in mind. However, such mismatches in SSH are also visible in Figure 6.2, where the broad areas of CAPES relate to the ISCED classification.

The ISCED groups are significantly different from the FORD ones, especially due to broad classifications such as "Services", "Education", and "ICTs". Once again, the connections between the SSH disciplines are very inconsistent. Additionally, the CAPES multidisciplinary broad area has a small connection to nearly all ISCED groups, as the system counts with a specific code in each group to include

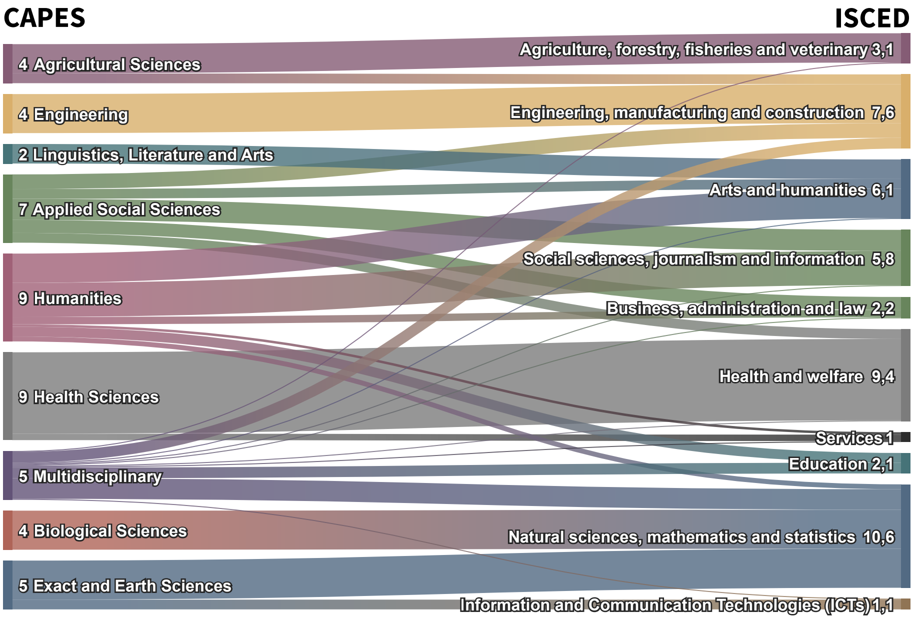


Figure 6.2.: CAPES broad area relations with ISCED's broad classification.

interdisciplinary programs and qualifications. Therefore, many of the different graduate programs within the CAPES "Interdisciplinary" evaluation area find a specific home within the ISCED classification.

6.3 Rethinking the Brazilian classification

The differences between the main classification system adopted in Brazil and alternatives such as FORD and ISCED are a problem for the country to conduct comparative studies on funding allocation, research dynamics in countries and disciplines, scientometrics. Although matching classification systems at their most granular levels – like what has been done for this chapter – can help conduct some of the types of study mentioned, it is unlikely that the time-consuming activity will be replicated widely and consistently. A solution would be to review the Brazilian classification to improve international equivalence, something also suggested by the special committee in charge of monitoring the Brazilian National Plan for Research and Graduate Education (PNPG).

Since the 1970s, Brazil has issued periodic PNPGs to help guide evaluation and science funding policies in the country (Brasil, 2020). The most recent plan covered the period 2011-2020 and the execution and results were monitored by a special committee. At the end of their term, the group prepared a report with many recommendations, including the need to rethink the current classification system, as the 49 areas do not reflect the modern panorama of science (PNPG Committee, 2020). Although the committee's recommendation for change is aligned with the findings of this study, there is a significant disagreement on the methods.

The PNPG Committee (2020) report suggests a substantial reduction in the number of evaluation areas, using the nine broad areas as a reference. However, we have seen significant discrepancies between the broad areas of CAPES and those of international classifications. Additionally, merging areas can represent a setback to a crucial achievement for research evaluation. After decades of area expansion, peer review committees achieved a level of freedom to customise evaluation criteria to suit their practises and value their principles. Moreover, the comparative perspective of the evaluation system has value when similar PPG exist within each area, but can be damaging in heterogeneous environments. Perhaps, the most adequate approach is not aiming for numbers, but for an adequate distribution of research that can be suitable for national evaluation and funding purposes, as well as international comparisons.

A possible method for reviewing the classification system may be supported by scientometrics. To demonstrate one possibility, microdata from the 2017-2018 papers in the three "Biological Sciences" areas (BioSci) have been collected from the CAPES Open Data System (CAPES, 2021a). Information such as DOI, ISSN, authorship, volume, page numbers, etc. was used to match the publications to the Web of Science.

Departing from the 15.375 documents matched to WoS, a term map of BioSci papers was produced using the VOSviewer software (van Eck and Waltman, 2009). For that, the title and abstracts of the articles were collected from WoS (Clarivate Analytics, 2022). Binary counting was used to extract more than 280 thousand noun phrases from the corpus, of which 8161 appeared in at least ten documents. A relevance score was calculated for each of these terms, with a threshold of 60%, and the resulting 4897 terms were used to produce the map seen in Figure 6.3.

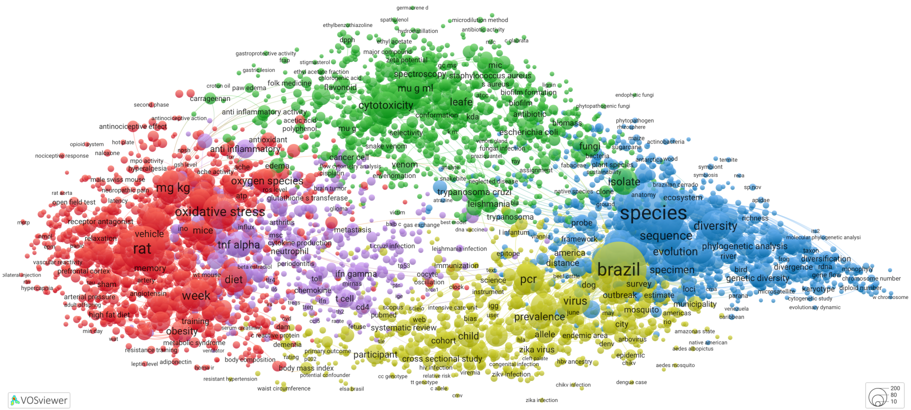


Figure 6.3.: Term map of papers from the Biological Sciences areas (2017-2018).

In Figure 6.3, the size of each circle represents the number of documents in which a term occurs. Proximity or distance between terms reflects co-occurrence, which also influences the creation of the five observed colour clusters.

With the term map representing the thematic publication profile of the three BioSci evaluation areas, Figure 6.4 adds a colour overlay to highlight publications of researchers affiliated with graduate programs in each of the areas. To improve comparability across the three areas, the scale is normalised by subtracting the mean from each variable and dividing the result by the standard deviation.

The bibliometric profiles in Figure 6.4 are revealing. First, we notice that BioSci I (a) and II (b) operate on opposite sides of the term map, showing that the areas focus the majority of their attention on specific research interests. Regarding BioSci III (c), the area operates towards the middle of the map, slightly overlapping BioSci I, but with greater attention to issues such as parasitology and immunology and with greater focus on issues of regional interest (observed in the 'Brazil' cluster). Although an expert committee could reach more robust conclusions from the maps provided, deciding whether the three areas need adjustment, the bibliometric perspective indicates that the research outputs of each area align with their associated subareas and specialities listed in the CAPES classification document (2020e).

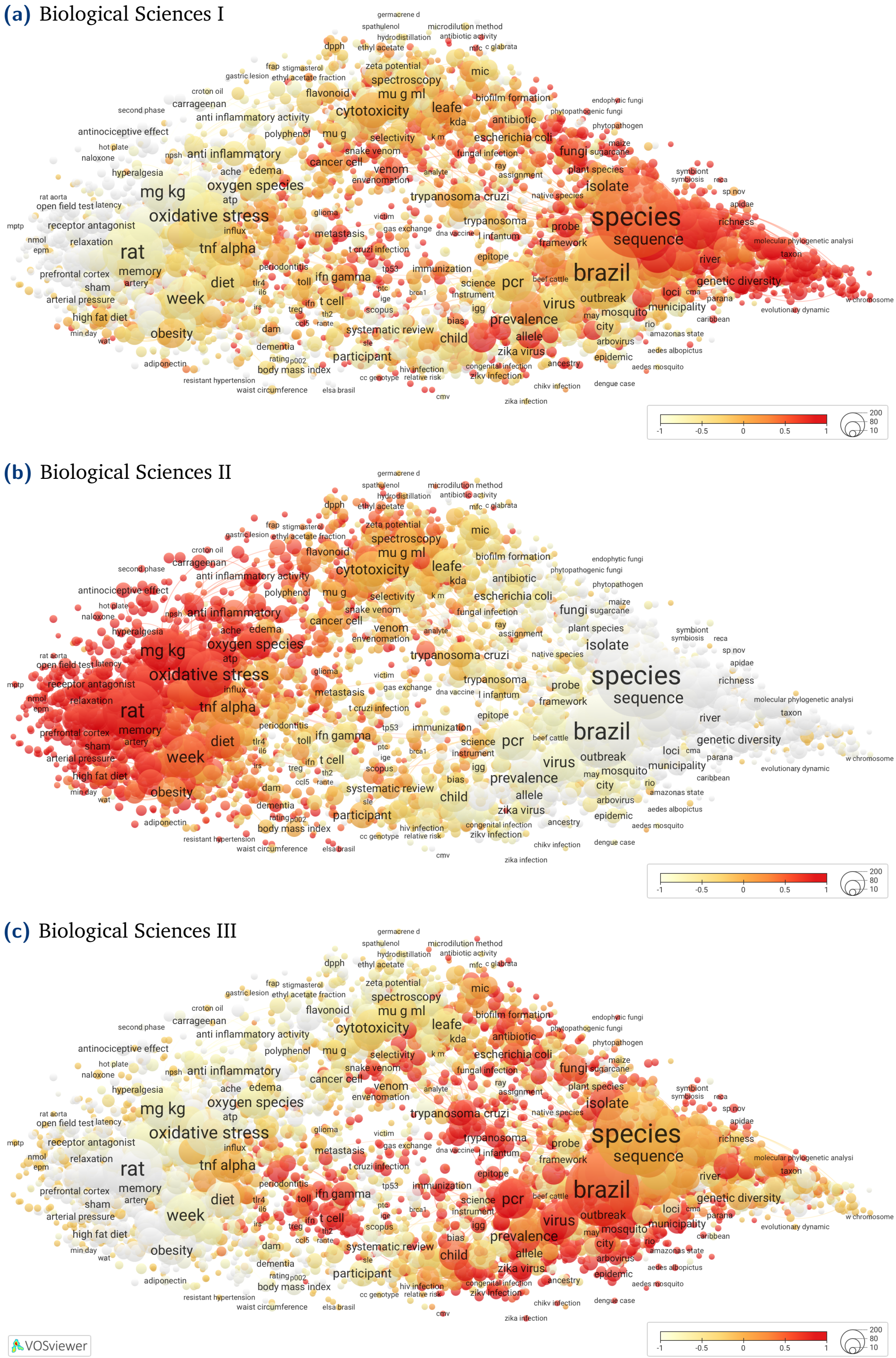


Figure 6.4.: BioSci term maps with publications highlighted per area (2017-2018).

Another application of term maps, as seen in Figure 6.5, is to focus on the profiles of individual graduate programs and how their research compares with the broader map of BioSci research.

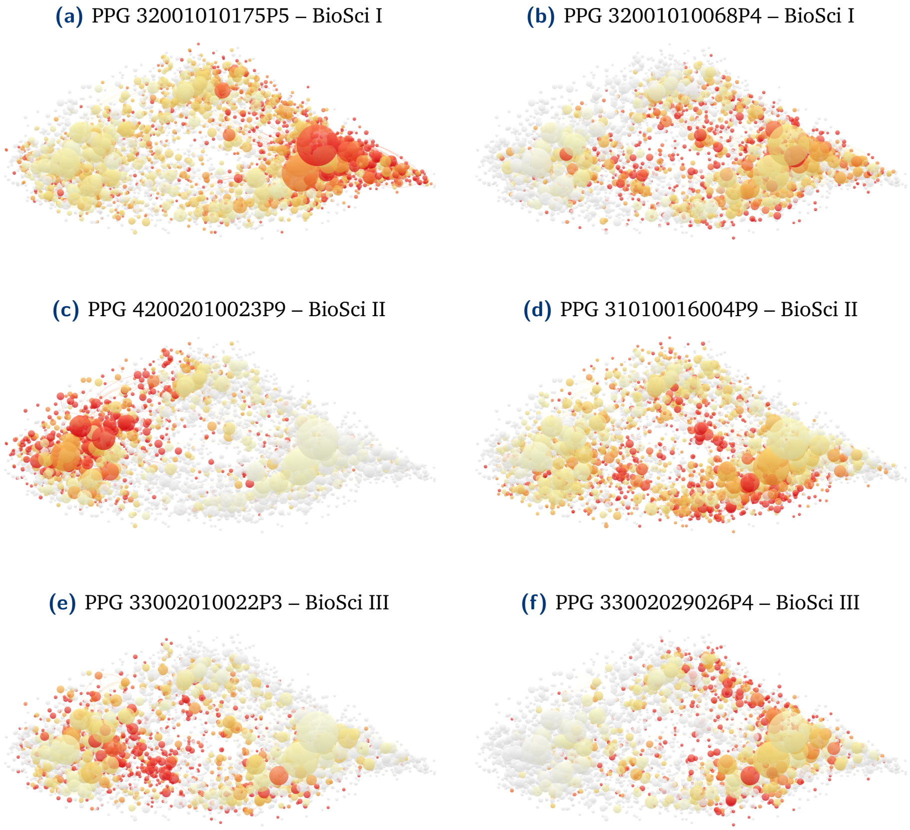


Figure 6.5.: Term maps of papers from the BioSci evaluation areas, highlighting the publication profiles of individual PPG (2017-2018).

Figure 6.5 displays publication profiles of two graduate programs in each of the Biological Sciences areas. The term maps shown on the left (a, c, and e) are from graduate programs whose profiles fit within the publication topics shown in Figure 6.4 for their respective areas. However, the maps shown on the right (b, d, and f) are examples of PPG profiles that may be more well suited for a different BioSci area.

It would be feasible to consider the profiles seen in Figure 6.5 as evidence to support the migration of some of these programs to different areas that would be more suited to their research profiles. However, the proposed approach should only be considered if it supports the work of disciplinary experts who have the required background to analyse the evidence and decide whether or not a migration would be recommended.

A complementary approach to help disciplinary committees in the further assessment of these publication profiles is the observation of how papers from selected areas are inserted into a broader map of science. To proceed with the analysis of the three areas of "Biological Sciences", the 2022 version of such a map was used as a starting point. The resulting visualisation seen in Figure 6.6 is built using the Leiden Algorithm, a method that performs cross-citation and semantic analyses of titles and abstracts between WoS-indexed publications since 2000 (Traag et al., 2019). The map in question displays a total of 4159 clusters, each of them composed of papers that have thematic relationships. Clusters are sized according to the total number of publications from 2017-2018, and the distances between them reflect the proximity of research subjects and citation-relations.

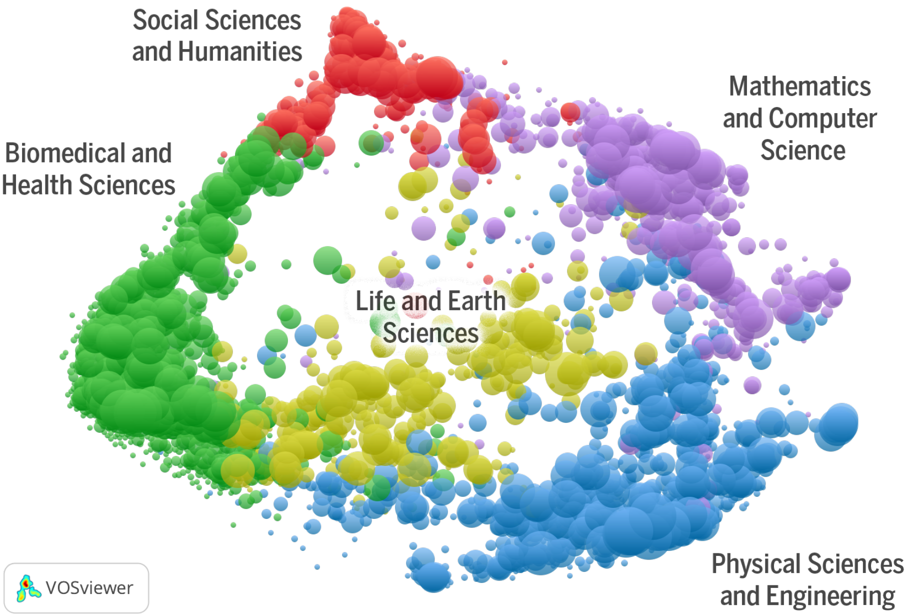


Figure 6.6.: Map of scientific publications indexed by the Web of Science (2017-2018)

Using Figure 6.6 as a canvas, it is possible to visualise publications from the three BioSci areas under analysis, recalculating the sizes of the respective clusters. The result, seen in Figure 6.7, shows the expected distribution of the papers mainly around clusters connected to the major fields of "Life and Earth Sciences" and "Biomedical and Health Sciences", which were highlighted in green and yellow on the previous map.

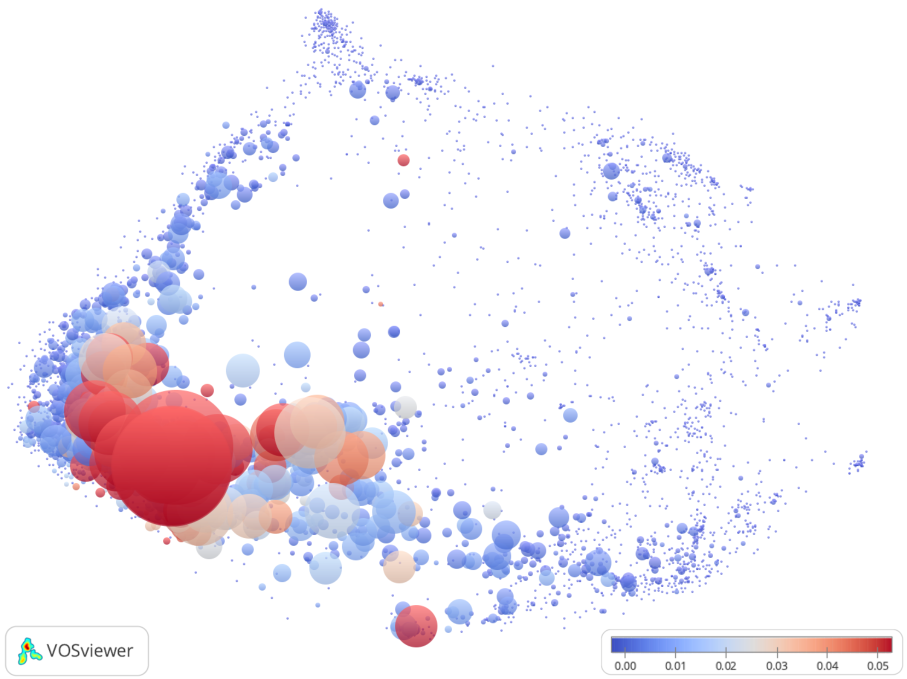


Figure 6.7.: Map of WoS-indexed scientific publications from Brazilian graduate programs in the BioSci evaluation areas (2017-2018)

Of the 4159 clusters on the displayed map of science, the BioSci graduate programs contribute to a total of 1580, 420 of those with more than 10 papers in the period. The colour overlay added to Figure 6.7 shows the percentage of those publications in relation to global production. Clusters displayed in vivid red are those in which the percentage of Brazilian papers is higher in relation to the total output. For instance, an analysis of the interactive version of the visualisation, available at https://bit.ly/xxxxxxxx, reveals that almost 20% of the world's publications in clusters related to tropical diseases such as Chagas and Leishmaniasis or in topics such as antivenom come from Brazil.

However, more than its contribution to global science, for the purpose of this study, the most relevant is understanding how Brazilian BioSci research is distributed in the three existing evaluation areas. For that, Figure 6.8 shows the previous map filtered for each of the BioSci areas. The visualisations are cropped to display the lower right of the original map, where most of the publications of those areas can be found.

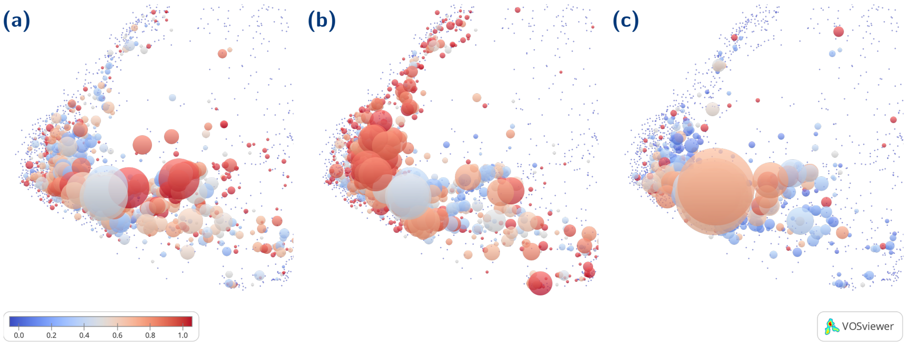


Figure 6.8.: Map of scientific publications indexed in WoS from Brazilian graduate programs in: (a) BioSci I; (b) BioSci II; and (c) BioSci III (2017-2018).

Figure 6.8 seems to confirm the conclusions derived from Figure 6.4, for instance, with respect to BioSci II (b) operating in its own research topics, while some overlap can be observed between BioSci I (a) and III (c). Such an overlap can be seen with the help of a new colour overlay, which applies the scale to the percentage of publications in each BioSci area in relation to the total of the three areas. Therefore, the predominance of reddish tones in many of the BioSci II clusters (b) indicates that 80 to 100% of the papers included there come from the area. However, while there are clusters highlighted in (a) where a majority of the papers belong to BioSci I, that is not the case for (c) where even clusters particularly large record only around 60% of the papers.

To better understand what the map reveals, Table 6.2 looks at the top 10 clusters for each of the three areas (as there is some overlap, the three top 10 are seen in 20 clusters). The table identifies the clusters with their unique id at the database and includes associated keywords to give some perspective of the topics included. For each of the three BioSci areas, the total number of papers (P) and their percentage in relation to the entire area are shown. The same is done for the combination of the three areas.

Table 6.2.: Top 10 clusters for each BioSci area, combined and sorted by total number of publications (2017-2018)

Id	Keywords	SciBio I		SciBio II		SciBio III		SciBio (all)	
		P	%	P	%	P	%	P	%
503	visceral leishmaniasis, psychodidae	127	10.0	141	11.1	238	18.7	364	13.0
521	chagas disease, reduviidae, hemiptera	175	13.8	172	13.6	197	15.5	353	12.6
53	zika virus, dengue, west nile virus, aedes	137	10.8	123	9.7	241	19.0	332	11.9
1190	phospholipase, snakebite, lipoprotein	88	6.9	72	5.7	67	5.3	150	5.4
1117	histoplasmosis, cryptococcal meningitis	62	4.9	42	3.3	112	8.8	144	5.1
7	microsatellite marker, genetic structure	127	10.0	2	0.2	5	0.4	132	4.7
1804	characiformes, teleostei, siluriformes	128	10.1	28	2.2	1	0.1	129	4.6
66	cerebral malaria, chloroquine, antibody	30	2.4	30	2.4	104	8.2	119	4.2
50	carvacrol, thymol, ocimum basilicum l	34	2.7	85	6.7	18	1.4	118	4.2
520	schistosomiasis, strongyloides stercorali	66	5.2	35	2.8	62	4.9	114	4.1
145	ixodidae, lyme disease, babesia	54	4.3	24	1.9	57	4.5	109	3.9
675	p2x, p2x7 receptor, extracellular atp	24	1.9	94	7.4	10	0.8	108	3.9
473	renin receptor, ace2, angiotensin ii	23	1.8	92	7.3	2	0.2	100	3.6
294	candida albican, candidemia	27	2.1	34	2.7	58	4.6	88	3.1
1707	galectin, tim, t cell immunoglobulin	56	4.4	59	4.7	26	2.0	84	3.0
272	urocortin, fever, cytokine, interleukin	17	1.3	69	5.4	9	0.7	80	2.9
82	morphine, ketamine, gabapentin, opioid	16	1.3	69	5.4	6	0.5	77	2.7
615	monogenea, acanthocephala, perciformes	14	1.1	14	1.1	54	4.2	73	2.6
45	tetrahydrobiopterin, arginase, nitroxyl	9	0.7	66	5.2	4	0.3	69	2.5
769	fabry disease, pompe disease	56	4.4	16	1.3	0	0.0	58	2.1

The first interesting observation from Table 6.2 is that the top 3 clusters represent more than 37% of the total number of publications in the combined areas. These are particularly relevant for BioSci III, as they concentrate more than half of the papers in the area, which are shown as the largest adjacent circles seen in Figure 6.8 (c). However, despite the high proportion of papers from the area in those clusters, the contributions from BioSci I and II are also significant. In fact, they reveal another interesting perspective: collaboration.

When the total number of papers of the BioSci combined is compared to those of the individual areas, they do not seem to add up. However, that happens because the same paper can be counted for more than one area, when a coauthorship led the publication to be reported in PPG from distinct areas. In the case of cluster 503, a total of 364 papers from all BioSci are mapped, 238 without crossarea collaborations (54, 57, 127 per area). Out of the 111 remaining papers from BioSci III, for instance, 42 were coauthored with BioSci I and 53 with BioSci II researchers, while 16 came from collaborations involving the three areas.

Evidently, the calculations used to build the maps of science and underlying clusters could consider fractional counting of publications to the proportion of the contribution of each area into account. However, the goal here is to map the research with which graduate programs are involved, making the full count approach appropriate, even because it helps identifying the crossarea collaborations.

Regarding the graduate programs, it is also possible to visualise and list their individual publications to the map of science and respective clusters, identifying those which are more or less aligned with the respective area profiles. The method, similar to what was shown in Figure 6.5, would be even more powerful, as the visual alignment would be complemented by a detailed list of publications in each cluster, complete with journals, collaborations and other resources that would be valuable for expert committees rethinking the classification system of evaluation areas in Brazil.

6.4 Conclusions

This study investigated the Brazilian classification system for research and graduate education. An analysis of the system's history has shown that the motivation behind its creation was a noble one: to guarantee that merit was a core element to the distribution of grants awarded by the chief funding agency in the country. Through the implementation of peer review committees, an evaluation model anchored by expert analysis was established. This is a model that is still current in Brazil.

Over time, the original committees multiplied towards the current 49 evaluation areas, organised into nine broad areas and three upper groups. This evolution was guided by the evaluation dynamics at CAPES, in part to follow the advancement of science, but also as a strategy to better manage the immense growth of the Brazilian National System of Graduate Education (SNPG). Furthermore, since the resulting classification played a central role in a high stakes national evaluation, its use beyond CAPES by other agencies and also by every higher education institution engaged in research in Brazil was inevitable.

Considering its evolution process, the Brazilian classification system under analysis became somewhat peculiar, especially when compared to international classification systems such as the OECD Fields of Research and Development (FORD) and the International Standard Classification of Education (ISCED). In particular, the misalignment among the evaluation and broad areas of the Brazilian system and their corresponding levels in the alternatives analysed is significant, especially in the SSH profiles.

One of the conclusions of this study is that the Brazilian classification system needs to be re-examined. Not only because of the misalignments identified, but because of other issues such as cryptical names of some evaluation areas, inadequacies in the allocation of graduate programs, the combination of sub-areas with significant epistemical differences, and the existence of areas which evolved to become apparent duplicates of each other.

Furthermore, the proposal for revision is aligned with recommendations from the special committee that was in charge of monitoring the Brazilian National Plan for Research and Graduate Education (2011-2020), which also highlighted the need for change. However, the PNPG Committee (2020) suggested the change to be one of significant reduction in the number of evaluation areas, reversing decades of efforts to build a system where the growing number of areas allowed for the comparative evaluation to be performed among graduate programs that were closer to each other.

This study proposes a different path. Instead of aiming for a reduction in the number of areas, the suggestion is to value and go beyond the "cognitive" approach described by Glänzel and Schubert (2003), which considers the input of different types of experts. For that, it is recommended to adopt the complementary "scientometric" approach to provide expert committees with evidence to suport their analysis.

The scientometric methods explored in this chapter demonstrate the potential of the approach, as the different analyses performed could be considered as starting points to help CAPES and the Brazilian academic community in the challenging task of promoting a sound and evolutionary review of the adopted classification system. With that, it is less important that the resulting classification decreases or increases the number of evaluation areas existing today. The important is that those new areas properly reflect the reality of the Brazilian science system and its international connections.

References

AAAS. (2022). The Scope of the "Humanities" for purposes of the Humanities Indicators. Retrieved April 20, 2022, from https://bit.ly/36uQCAH

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2021a). A national evaluation push towards increased societal impact: The Brazilian experience in valuing broader research outputs. In *Proceedings of the eu-spri conference 2021*. Extended abstract, Oslo, Norway: Eu-SPRI. Retrieved from https://www.euspri2021.no/abstracts/

Brasil, A., Trevisol, J. V., & Drooge, L. v. (2022). Research evaluation in Brazil and The Netherlands: A comparative study. In *Eu-spri 2022*, Utrecht, NL. Retrieved December 31, 2022, from https://bit.ly/3gRMNuo

CAPES. (2011). *Portaria nº 83, de 6 de Junho de 2011. Cria 4 áreas de avaliação: Biodiversidade, Ciências Ambientais, Ensino e Nutrição*. Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/383xcmX

CAPES. (2016a). *Portaria nº 141, de 14 de setembro de 2016. Define e disciplina as formas de colaboração e os procedimentos de escolha dos consultores científicos* (Ordinance No. 141/2016). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3v5NG7B

CAPES. (2020e). *Tabela de Áreas de Conhecimento/Avaliação*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3I7infy

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. (2021d). *Portaria nº 122, de 5 de Agosto de 2021. Consolida os parâmetros e os procedimentos gerais da Avaliação Quadrienal de Permanência da pós-graduação stricto sensu no Brasil*. Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/384xDgE

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

Castro, C. d. M., & Soares, G. A. D. (1983). Avaliando as avaliações da Capes. *Revista de Administração de Empresas*, *23*(3), 63–73. https://doi.org/10.1590/s0034-75901983000300007

Clarivate Analytics. (2022). Web of Science Core Collection. CWTS in-house version. Retrieved from https://bit.ly/3rHmCJT

CNPG. (1974). *Plano Nacional de Pós-Graduação - PNPG 1975-1979*. Brasília, DF: MEC.

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

Córdova, R. d. A. (2001). *CAPES: Origem, Realizações, Significações (1951-2001)*. Brasília, DF.

Dias, A. M. I., Therrien, J., & Farias, I. M. S. d. (2017). As áreas da educação e de ensino na Capes: identidade, tensões e diálogos. *Revista Educação e Emancipação*, *10*(1), 34–57. https://doi.org/10.18764/2358-4319.v10n1p34-57

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*(3), 357–367. https://doi.org/10.1023/a:1022378804087

Gouvêa, F. C. F. (2012). A institucionalização da pós-graduação no Brasil: O primeiro decênio da CAPES (1951-1961). *Revista Brasileira de Pós-Graduação*, *9*(17), 373–397. https://doi.org/10.21713/2358-2332.2012.v9.312

OECD. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. Retrieved from http://oe.cd/frascati

Stern, F. L. (2019). A criação da área de avaliação ciências da religião e teologia na Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). *ESPAÇOS - Revista de Teologia e Cultura*, *1*(26), 73–91. Retrieved from https://espacos.itespteologia.com.br/espacos/article/view/62

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, *9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z. eprint: 1810.08473

UNESCO. (2015). *ISCED fields of education and training 2013: Detailed field descriptions*. UNESCO Institute for Statistics. Montreal. Retrieved from http://dx.doi.org/10.15220/978-92-9189-179-5-en

van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

7

Between Bibliometrics and Peer Review

> *Qualis should not be considered an appropriate source for classifying the quality of scientific journals for purposes other than the evaluation of graduate programs.*
>
> **— Rita Barata**

Since the 1970s, Brazil has endeavoured to evaluate research and graduate education through a system that serves as a critical determinant of accreditation, permanence, and funding allocation (Brasil, 2021b; C. B. Martins, 2018). However, with the natural expansion of this system, evaluating scientific production qualitatively, a vital component of the process, became increasingly difficult (CAPES, 2003; Hortale, 2003). In response to these challenges, the Qualis ranking system was established in 1998 to assess the quality of academic journals as a proxy for the research contained therein (CAPES, 2003).

This chapter delves into the genesis of Qualis, examining its initial conception and its evolution over the years. Throughout its development, Qualis has shown adaptability to the dynamic nature of academic research and assessment, but cumulative improvements have led to or revealed fragilities in the system, some of which result from misuse of the ranking by external actors (M. F. B. Leite et al., 2010; Soma et al., 2016; Spagnolo and Souza, 2004). However, since its first

This chapter is an expanded version of the conference paper: Brasil, A. (2023). Between Bibliometrics and Peer Review: The Evolution and Challenges of Brazil's Qualis System. *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. Leiden, The Netherlands. https://doi.org/10.55835/6442f8d5389baf2145b264d9

use, Qualis has played an important role in the evaluation of research and graduate education in Brazil, adding to its foundation of continuous improvement (Hortale, 2003; C. B. Martins, 2018).

To better understand Qualis' current role, particular emphasis is placed on the two most recent national evaluations in which Qualis was adopted: The Quadrennial Evaluations of 2017 and 2022. The 2017 evaluation was crucial because the adopted assessment model was the culmination of two decades of minor adjustments, leading to an evident need for a more significant reform (Barata, 2019; CAPES, 2018b; PNPG Committee, 2018). This reform was planned and implemented in the following years, but put to the test in the evaluation of 2022 (Amado et al., 2020; CAPES, 2023b; PNPG Committee, 2020; Reategui et al., 2020; R. J. Ribeiro, 2022a; R. J. Ribeiro, 2022b). This latest iteration of Qualis reflects a more robust approach to the evaluation of journal publications, addressing some of the previous limitations while acknowledging its inherent strengths. However, the current Qualis system is still flawed, and while this study acknowledges its progress, it also pinpoints areas that warrant further improvement. Thus, building upon previous critiques (Barata, 2016; Barata, 2019; CAPES, 2018b; PNPG Committee, 2018; PNPG Committee, 2020), this study proposes alternative strategies to address the remaining weaknesses and improve the general reliability and efficacy of the Qualis system.

7.1 Qualis conception and early developments

Brasil (2023b) extensively examined the way in which several fundamental principles of the Brazilian evaluation system emerged as a result of resource constraints. For almost half a century, the primary unit of assessment within the national evaluation framework has been the graduate program (PPG). This was due to the expansion of the Brazilian National System of Graduate Education (SNPG), which experienced a growth in the number of PPGs, master's and doctoral students, and scholarships available. Despite this expansion, the personnel responsible for processing all concessions remained limited in number. Consequently, a decision was made to delegate the assessment of individuals to their respective graduate programs. CAPES would evaluate PPGs, award them with a specific number of scholarships, and then allow programs to carry out their own internal selection process to determine the allocation of available grants. This model persists today and was adopted similarly when the Qualis system was first developed in 1998.

During that time, the Brazilian evaluation went through a significant restructuring. For example, a standardised evaluation form was adopted to evaluate graduate programs (PPG) in various disciplines, resulting in more consistent and comparable outcomes (Monteiro et al., 2019). Abílio Baeta Neves, CAPES' president during the reform of the evaluation system, discussed part of the motivation behind the changes in an interview with Ferreira and Moreira (2002). Neves explained that Brazil's PPGs had already achieved a reasonably high quality level, and many of the adopted indicators made it easy for a large number of these programs to achieve the highest possible evaluation score. For example, he cited the number of faculty members with doctoral degrees as an indicator used in the 1996 evaluation. However, a simple analysis revealed that faculty without doctorates were already scarce, rendering the indicator obsolete. Consequently, if no new criteria were introduced, efforts to differentiate programs could be compromised. Therefore, a decision was made to improve the evaluation of the quality and international integration of graduate education. Attaining this goal required replacing indicators such as the mere number of published articles, used at the time, with more significant metrics.

Qualis was the answer to the pressing problem. Taking into account the lack of resources to qualitatively assess all the paper production in the country, CAPES opted to classify scientific publishing outlets, assuming that articles accepted by indexed journals with a peer review system would guarantee a certain level of quality (Barata, 2016; Glänzel and Moed, 2002). The aim of Qualis was never to become a journal ranking. The idea, as highlighted by Neves in Ferreira and Moreira (2002), was to identify journals that should count as having scientific relevance in the national or international scenario, distinguishing what circulates knowledge and matters in each field from what serves as an auxiliary tool in graduate training and qualification, even though such dedicated publication channels can be vital for the development of scientists-in-training.

Qualis was then created with the primary purpose of classifying the journals listed by graduate programs in the annual data collection system used by CAPES to map the work conducted by faculty members and graduate students. The first classification system assessed journals in two main dimensions: quality or relevance in a specific scientific field (A for high, B for average, or C for low);

and their circulation (1 for international, 2 for national, and 3 for local). Three additional rankings were added to the nine possible combinations: *SR* – which means that the areas did not have enough information to classify the journal; *IP* – improper, meaning that it was not considered a scientific journal by the area committees; and *NC* – not classified by the area (CAPES, 2003).

Therefore, CAPES would collect the complete list of publications from graduate programs throughout the country and make them available to each evaluation area committee for classification. The areas had some flexibility with respect to the evaluation methods adopted, but taking into account the general guidelines provided by a council of area and agency representatives (CTC-ES). As expected, evaluations would rely more on database indicators in the areas of science, technology, engineering, and mathematics (STEM), and qualitative methods for those in the social sciences and humanities (SSH) (Barata, 2016; CAPES, 2003; Soma et al., 2016). Some additional characteristics of the original Qualis are:

i) Qualis is not a comprehensive list of journals. It contains only those with publications reported by graduate programs during each evaluation cycle;

ii) Classification is *ex post*, so journals are ranked according to the assessment performed after the publications are reported to CAPES by graduate programs. No *ex ante* component is present in the classification, so no expectation of future performance can be derived from a Qualis result;

iii) Qualis is a temporary list, not a cumulative one. That means that classifications from one cycle are not transported to the following one, meaning that some journals will leave the list, others will be included, and those that remain may receive a different classification;

iv) Journals can have multiple classifications across evaluation areas, as the same journal can be used to publish papers from graduate programs in different areas, and each committee conducts an independent analysis.

Taking into account the premisses and characteristics listed for the original Qualis, its primary purpose is reinforced and it becomes clear that "Qualis should not be considered an appropriate source for classifying the quality of scientific journals for purposes other than the evaluation of graduate programs" (Barata, 2016, p. 17). As a consequence, PPG should not use Qualis to hire staff, as the candidates' publications may have taken place in journals that are not in the current Qualis. Qualis should also be used with caution when selecting journals to publish, as a journal classified in the top strata in one evaluation cycle might not be granted the same level in the next. However, some evaluation areas progressively incorporated other purposes for Qualis, such as making select journals more attractive for prospective authors by artificially inflating their ranking, or adding journals not reported in the data collection to the list, also aiming to stimulate publications in journals considered important in each field (CAPES, 2003).

With some of these distortions already being incorporated to the basic premisses defined in its beginnings, Qualis was used for a whole decade undergoing just minor evolutive adjustments after each new evaluation cycle. After the 2007 national evaluation, CAPES considered that it was time for a more significant change in the classification.

7.2 Reviewing Qualis for a new phase

The original Qualis was conceived around a major turning point for scientific publishing. The end of the 20^{th} century was a dynamic period that reshaped publishing and citation practises. CAPES, for example, launched its Portal of Journals in 2000, an online platform that provided graduate programs throughout the country with access to 1.419 digital journals, a number that would multiply to tens of thousands in the following years (Brasil, 2020). With the digitalisation of scientific publishing, classifying journal publications on a scale that considered local, national, or international circulation made little sense. Thus, the Qualis classification was restructured after extensive discussions within CAPES and the evaluation area committees. The main change was the replacement of the double scale for circulation and quality with a new single scale of seven strata: *A1*, *A2*, *B1*, *B2*, *B3*, *B4*, and *B5*, plus an additional stratum *C*[1] for publications that did not meet the minimum criteria established in each area (Soma et al., 2016). The following rules applied to the new scale, which continued to be independently attributed by each evaluation area.

[1] Some area committees made use of an additional stratum, *NP*, to classifly journals considered not to be of scientific nature. However, since both *C* and *NP* journals would be excluded from the calculation of any indicator used in the evaluation process, many areas did not care to make a distinction between the two strata, which have been used irregularly over the years. For this study, both strata will be unified as *C*, as it is the case in the CAPES Qualis reports available at https://qualis.capes.gov.br/.

i) Fewer journals should be classified as *A1* than as *A2*;

ii) The number of journals in *A1* + *A2* can account for a maximum of 25% of the journals listed in the area;

iii) Similarly, *A1* + *A2* + *B1* cannot add up to more than 50% of the journals;

iv) All strata must be populated, with only *C* as a possible exception.

The rules behind the new Qualis scale reveal that the previous model may have led to overpopulation of the upper stratum in some areas. This becomes clear, as the review is said to be primarily motivated by the need to recover the gradual loss of discriminatory power experienced over the years (Barata, 2016; Soma et al., 2016). Although every journal could be ranked *A* on the previous scale, the new top *A1* rank was limited to be around the 12% percentile. *A2* would include the remaining top-quartile journals and *B1* those above the median. Some discriminatory power was regained, even though it was kept at a level of detail sufficient for the evaluation of graduate programs, but broad enough so that Qualis would not become a competitive ranking.

Despite additional advances in data collection, little else changed in the new Qualis. Old rules remained in place and old problems persisted. For instance, an exception to the original Qualis purpose allowed areas to inflate rankings of journals they wanted to promote, and committees also used the mechanism to devalue journals due to issues such as pertinence, relevance, and adherence to the area. Qualis became a classification of more than quality within the specific purpose for which it was intended. This has led to inconsistencies, such as in the case of journals with completely distinct classifications between areas.

An example of this issue is shown in Table 7.1, which shows a multitude of different Qualis classifications attributed to "Evaluation: Journal of Higher Education Evaluation", a multilingual Brazilian journal founded in 1996. Operating under the Diamond Open Access model (free for authors to publish without any article processing charges), the journal is indexed in databases such as the Scientific Electronic Library Online (SciELO) but not in the more international ones such as Scopus and Web of Science. Although the journal has been valued as *A1* by the evaluation area of Education since the 2010 edition of Qualis, the same is not true for other areas. Table 7.1 lists the various strata attributed to the journal in the 2017 Qualis, together with the number of papers published by each area in the 2013–2016 evaluation period (CAPES, 2023d).

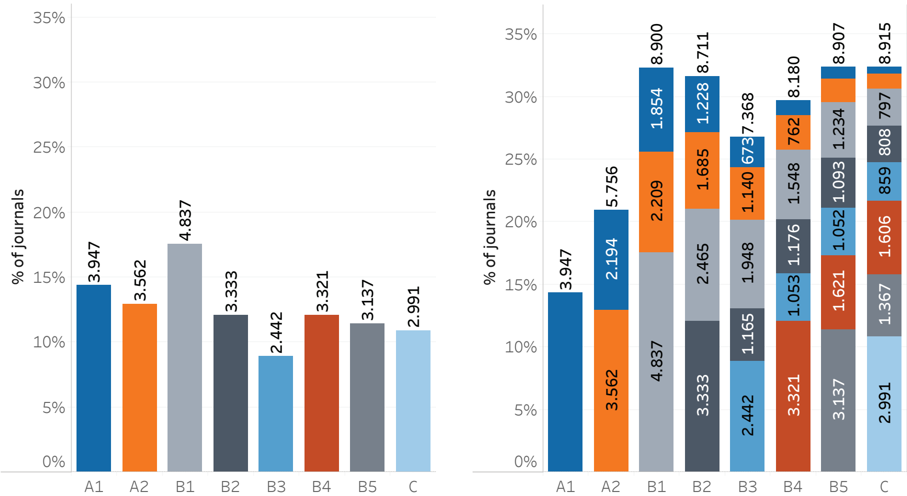
Table 7.1.: Example of a journal with multiple strata in the Qualis 2017 classification

Journal	Stratum	Evaluation area	Papers 2013-2016
Evaluation: Journal of higher education evaluation ISSN: 1414-4077	A1	Education	72
	A2	Literature and linguistics	3
		Interdisciplinary	32
		Teaching and learning	3
	B1	Arts	1
		Journalism and information	2
		Social work	1
		Sociology	1
		Agricultural sciences	3
		Environmental sciences	2
	B2	Physical education, therapy and rehabilitation	2
		Public health	1
		Business and administration, accounting and tourism	43
		Economics	1
		Psychology	3
	B3	Engineering I	1
		Dental studies	1
		Political Science and international relations	1
	B4	Engineering III	5
		Medicine II	1
		Nutritional science	1
	B5	Pharmacy	2
	C	Medicine I	1

Considering Qualis as a system that links strata to percentiles with its *A1*, *A2*, and *B1* limitations, it makes sense that areas such as “Business and administration...” would rather not give a ranking above *B2* to an Education journal, as the higher classification could be used for a journal with closer connections to the area. However, it is difficult to consider adequate that areas are allowed to challenge the quality assessment of a discipline about one of its own journals, to the extreme of ranking it as *C*, like Medicine I has done for "Evaluation...". With that, the area is saying that the journal is not considered good or relevant enough to even be considered in the evaluation process, as *C* journals do not add any value to the assessment of a PPG’s scientific production.

Furthermore, while the example in Table 7.1 refers to a locally relevant journal, the problem affects even internationally established journals. For instance, the official Qualis 2017 results reveal that while “Science” has been classified as *A1* in most areas, it was also classified as *A2* in Economics and in “Business...”, and *B1* in Law.

Although areas may be able to demonstrate a variety of methods used in their journal classification and they may also justify any posterior adjustments they have performed, the extreme variations across all areas weaken Qualis as a trustworthy instrument in the national evaluation performed by CAPES. The extent of these variations can be seen in Figure 7.1, with (a) showing the distribution of the best classification granted to each journal in the 2017 Qualis, and (b) showing all classifications obtained per journal.



(a) Best classification per journal

(b) All classifications per journal

Figure 7.1.: Strata distribution of the Qualis 2017 classification of journals

Figure 7.1a displays the highest rank achieved by academic journals in the 2017 Qualis. The figure presents a distribution of 27.570 unique ISSN numbers that combine the results of the 49 evaluation areas. However, since the number of journals varies by area, the distribution of the first three strata deviates slightly from the *A1*, *A2*, *B1* rules. For example, almost 15% of the journals were rated *A1* by at least one evaluation area. *A2* is close to what was expected, as is the large percentage of journals in *B1*, which could even be higher since it represents the entire second quartile of journals. Figure 7.1b shows the same data as Figure 7.1a, but with all other journal classifications stacked on top of the highest-ranking bars. For example, of the 3947 journals classified as *A1* in at least one area, 2194 were ranked as *A2* by one or more areas, while 1854 were also classified as *B1*, etc. Reaching the lowest stratum, *C*, are 158 journals. The same pattern can be observed for all other rankings.

Although CAPES (2023b) recognises that the multiplicity of strata for the same journal was the biggest challenge offered by Qualis, the agency defended the process, arguing that diversity in classifications was not an inconsistency, but a reflexion of how each journal was valued with respect to its relevance to each evaluation area. Technically, the argument is correct, and the criteria and indicators used by each of the 49 evaluation committees to build their Qualis lists are publicly available in detailed reports, published before the classification takes place and after discussion and approval by the council of area and agency representatives (CTC-ES) (CAPES, 2020d; Soma et al., 2016). However, even the evaluation director at the time of the 2017 Qualis classification states that "there is no reason why the same journal receives classifications that are so disparate across areas" (Barata, 2019, p. 5). Undoubtably, Figure 7.1b shows not only how disparate these classifications could be, but also that they are more a rule than an exception.

7.3 Towards the current Qualis system

In 2015, CAPES started a new effort to reform Qualis, appointing a working group to study the classification and propose changes to its methodology. The WG presented its diagnosis and a series of recommendations in the following year. Some of the main proposals, as reported by Pascutti (2016), were as follows:

i) Encourage dialogue towards more homogeneous evaluation criteria and classifications, so that the disparity of strata across areas would not be so extreme;

ii) Avoid excessive emphasis on journal adherence to the areas, limiting stratum variations that exceed one level (i.e., if the journal quality would grant it the *A1* rank, lack of adherence to the area would warrant an adjustment to *A2*, but not to further strata such as *B5* or *C*);

iii) Adopt reference areas for journals, when possible. That means the stratum attributed by an evaluation area to its main journals would serve as reference for the evaluation in other areas;

iv) Evaluations should be based on qualitative and quantitative criteria, rather than subjective ones;

v) Introduce bibliometric indicators in fields that do not yet adopt them. Even if as a secondary criterion, they could encourage PPG to publish in indexed journals and stimulate the indexing of national journals.

The changes proposed by the working group were not implemented at that time, as the national evaluation for the 2013-2016 quadrennium would be held in 2017, so no change in procedures would be adviseable near the end of the assessment period. However, the debate was resumed shortly after the 2017 evaluation results were made public, and a new working group was appointed, with the aim of continuing the effort and proposing changes to be implemented in the next Quadrennial Evaluation (CAPES, 2018d). The proposed new model would not only align with the main recommendations of the previous working group, but would take them further. The main principles of the new Qualis are described in a technical report (CAPES, 2023b), and can be summarised as follows.

i) Qualis becomes, as originally intended, an instrument for the evaluation of graduate programs through the classification of journals based on quality. Pertinence, relevance, and adherence are no longer part of the analysis, which should rely on objective indicators;

ii) Journals are no longer classified in multiple strata across evaluation areas. Now, each journal has a unique classification;

iii) A journal is classified by the area it aligns with most closely, denominated its "mother area". This designation is determined by the volume of articles published in the journals between 2013 and 2019, although with allowance for shifts as negotiated between evaluation areas;;

iv) A new scale was adopted, expecting a more balanced distribution of percentiles between strata: *A1*, *A1*, *A2*, *A3*, *A4*, *B1*, *B2*, *B3*, *B4*, and *C*, which remained for those journals with zero value for the evaluation.

v) The previous restrictions on the proportion of journals per category were lifted. This means that the evaluation areas were no longer bound by guidelines such as ensuring that the sum of journals rated *A1* and *A2* didn't exceed 25% of the total journals in the area.[2]

[2] As will be discussed later in this study, while the restriction on journals allocated to each strata was removed from the Qualis methodology, the change was somewhat offset by the introduction of CAPES adoption of a suggested classification derived from bibliometric indicators, which allocates 12,5% of the total number of journals to each of the eight strata, from *A1* to *B4*.

A first look at the Qualis classification released in early 2023 (CAPES, 2023d) reveals some positive and negative consequences of the new model. For example, Figure 7.2 shows a distribution matrix of the journals according to the mother areas established in the evaluation and published in CAPES (2023c). The numbers in rows and columns represent the codes of the evaluation areas, according to what has been reported in Brasil (2023b), and the areas are grouped according to the three broad areas in the CAPES classification.

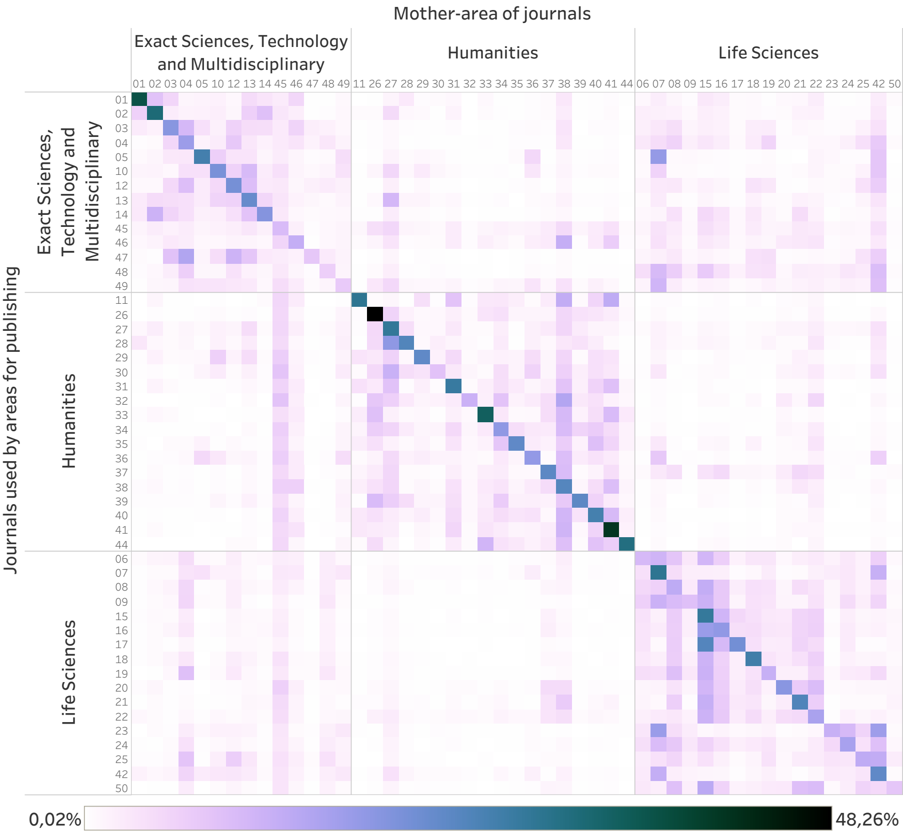


Figure 7.2.: Distribution of journals according to mother-area in the evaluation and publishing (2017-2020)

At the top of Figure 7.2, you see the mother areas, so the journals displayed in the columns represent those that were classified by these areas. On the left you see the areas publishing in journals that were classified by the areas on top. The darker diagonal, as expected, reflects the number of journals that each area used to publish and classified as the mother areas. For example, Law (26) classified a total of 1418 journals, which represent around 48% of the universe of journals used by its graduate programs during the 2016-2020 period. Law also published in many other journals, for instance, in 153 classified by Education (38), 113 by Sociology (34), 96 by Philosophy and Ethics (33), etc. Navigating on the Law column, it is possible to see that area 39, which is Political Science and International Relations, published in 209 of the journals that were classified by Law as their mother area.

Although Law is the area that classified the highest percentage of journals used by their graduate programs, Biotechnology (48) classified the least, being the mother area for only 6,4% of the journals used by its PPG in the period (267 journals). Evidently, as an interdisciplinary field, biotechnology research can fit many journals within areas such as Agricultural Sciences (42), Biodiversity (07), Chemistry (04), and others. The interactive matrix is available at https://andrebrasil.github.io/viz/qualis.html, where each intersection shown can be explored in detail. Furthermore, it is possible to change the visualisation from journal to published articles, for example, revealing that more than 88% of the articles published by Law were in the journals they classify.

Returning to the example presented in Table 7.1, the mother area of the journal "Evaluation: The Journal of Higher Education Evaluation" in the most recent Qualis was "Education". Thus, as a result of the change in approach that guarantees that a single stratum will be assigned to a journal by its mother area, the journal was granted the *A1* stratum. The classification was then used for the evaluation of the area, but was also applied to 18 other areas with published articles in the journal during the 2017-2020 quadrennium.

However, "mother areas" are not always absolute and independent in these decisions, in part because they can also have "sister areas". This happens if the area with the most publications in a journal does not include at least 50% of the journal's publications in the period of analysis. In that case, the area with the most publications is still the mother, but any decisions regarding the classification of the journal must be agreed with the sisters, which can be up to three other areas with a significant number of publications in the journal.

Furthermore, although some key concepts of the new Qualis were not polemical, such as the new scale, the same did not happen with changes like the single classification for journals and the methodology, or even the existence, of mother areas. Even during preliminary discussions, CAPES received many manifestations criticising ideas being explored to review Qualis. For example, a letter from the Brazilian Association of Collective Health (ABRASCO, 2019) mentioned that the number of publications was not sufficient to determine the motherhood of a journal and highlighted how interdisciplinary areas would lose much control over the classification of journals relevant to them. After the publication of the classification used during the Quadrennial Evaluation of 2022, numerous groups complained about the changes and results, and most of the criticism focused on the adoption of a single strata for a journal in all areas (Brigatti, 2023; Ferrari, 2023; Yamashita, 2023).

Some manifestations may be important for the evolutionary dynamics of evaluation. For instance, the letter by ABRASCO (2019) pointed out that the initial proposal to introduce "mother areas" considered only two years of publications for the decision on how journals should be distributed, but that a longer period should be adopted. As a result of constructive criticism, CAPES (2023c) described the final methodology that considered seven years of data for distribution. However, some other manifestations should be mainly ignored, for instance, the complaints of some Economics researchers who cannot accept that journals such as "Religion Studies" and "Experimental Dermathology" could be granted the same *A1* that is used for elite Economics journals (Brigatti, 2023). Why would it not be possible to accept that quality exists beyond the borders of one's discipline?

7.3.1 The challenge of an indicator-based Qualis

Right after the conclusion of the Quadrennial Evaluation of 2017, CAPES instituted many working groups to reflect on ways to improve the evaluation model for the next cycle. A seminar series branded "Rethinking Evalution" was organised to broaden the discussion and allow the groups to present preliminary findings and debate them with the evaluation community in Brazil. The second seminar of the series focused on the evaluation of scholarly production, including books, events, technical and technological production, and more (CAPES, 2018b). The presentation by the Working Group on Qualis Journals presented their main findings and the proposal for a new Qualis, which was very close to the final format that was implemented.

Regarding the use of bibliometrics, the group showed that indicators such as Scopus CiteScore and the Journal Impact Factor had a very high correlation with previous evaluations in most areas in "Life Sciences" and "Exact Sciences, Technology, and Multidisciplinary". As described in Brasil (2021b), these areas are better covered in databases such as the Web of Science. Therefore, well-established indicators should indeed correlate with the classifications performed by the areas that traditionally mix metrics with peer review in their analysis.

Based on the findings of the working group, a proposal was presented for a methodology based on the adoption of a common set of indicators for all areas. No agreement could be found as the classifications in a significant share of the areas, particularly within the Social Sciences and Humanities (SSH), did not correlate with the proposed metrics. After months of debate, a new working group was established to deal with the specificities of journal classifications in the SSH (CAPES, 2019c), and the original group was reinstated with its scope limited to "Life Sciences" and "Exact Sciences, Technology, and Multidisciplinary" (CAPES, 2018e).

As the new working groups worked on their proposals for a Qualis classification that included the use of open metrics that supported the evaluation of different areas, CAPES organised midterm seminars with the participation of graduate program directors (GPD) of more than 4500 PPG in the country. During a three-month period, each of the 49 evaluation areas met with their respective GPDs to diagnose the status of their areas and programs midway through the four years of the evaluation cycle (CAPES, 2019b). For that, a provisional Qualis was prepared, using for the first time the concepts of "mother area", unique journal classification, and the use of the mentioned indicators to serve as a reference for area committees in a peer review phase of the process.

The process worked as follows: CAPES would rely on indicators to suggest a stratum for the classification of each journal in the analysis. The evaluation area committees would receive the recommendation, and they could either accept or reject each proposed classification, reclassifying the journals under their purview according to the methods of preference in each area. Figure 7.3 shows how the area committees dealt with the suggested classification, displaying the indicator-recommended classifications at the top of the graph, and bars showing the percentage of recommendations that were validated (highlighted in orange) or reclassified during the peer review phase.

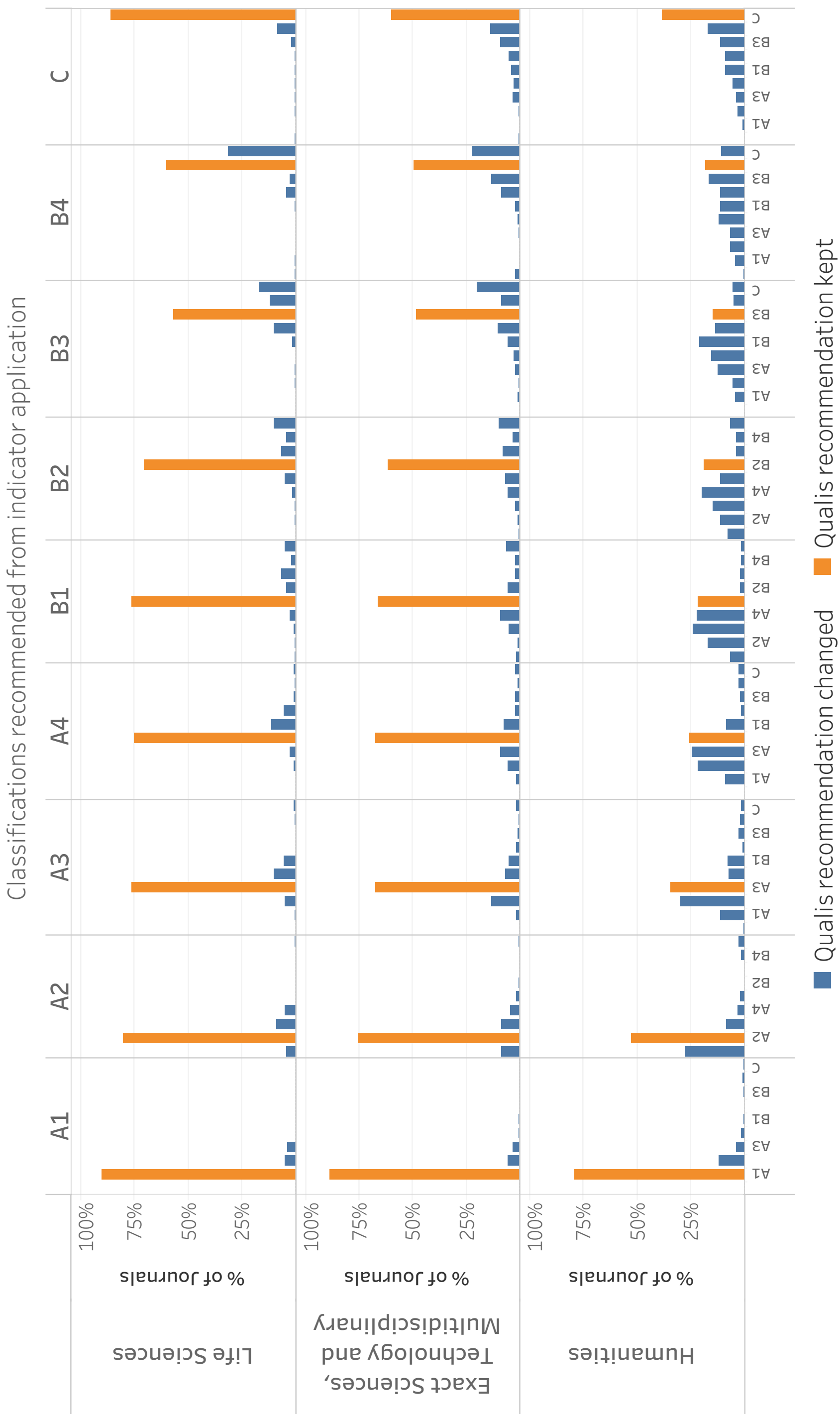


Figure 7.3.: Mid-term reclassification of journals from the indicator-based suggestions to the peer review results, grouped by broad areas

Figure 7.3 shows how the indicators used were able to capture what the committees in "Life Sciences" and "Exact Sciences, Technology, and Multidisciplinary" consider as a measure of the quality of the journals, particularly in the upper strata. Around 90% of the journals with a suggested classification of *A1* were confirmed by the committees as such, and a minority was reclassified mainly as *A2* or *A3*. The correlation continues high in the following strata, slowly decreasing as it reaches the bottom percentiles. Even then, for the B4 and B5 classifications, around 60% of the recommendations are kept in "Life Sciences", and 50% in "Exact Sciences...". In both cases, even the journals listed as *C* were considered to be mostly correct by the evaluation committees.

The panel for the "Humanities", including all of SSH, is quite distinct. Although there seems to be a good correlation with indicators in the top *A1* stratum, with nearly 80% of the indicator-based suggestions kept by the committees, the percentage drops drastically to 52% in *A2*, 34% in *A3*, and then is completely lost as a guide in the next strata. Regarding the stratum *C*, where journals were assigned when no indicators were available, more than 60% of them were reclassified by the "Humanities", some even reaching the stratum *A1*.

What is shown at Figure 7.3 can be considered a good result. Indicators indicate; they do not determine final scores, which are a prerogative of peer review committees. Through the experience and knowledge of the evaluators within their disciplines, the ability to challenge the indicators when necessary is a crucial benefit of the Qualis model as implemented during the mid-term evaluation. However, reclassification within "Humanities" may be considered excessive if two of the core recommendations of the 2015 working group are taken into account: Evaluations should apply qualitative and quantitative criteria, rather than subjective ones; Bibliometric indicators should be used in fields that do not yet adopt them. The results suggest that indicators may be mostly ignored in the process, and subjectivity may also be an issue of concern.

The mid-term Qualis was an exercise to test the methodology under development. With the ongoing work of both Qualis working groups, the exercise could contribute to the improvement of the methods, plus the addition of others useful for the SSH. Together, these approaches could make the final reclassification look more like the one in the "Exact Sciences..." for all three groups.

The final Qualis methodology was approved by CAPES in September 2020, one year before the Quadrennial Evaluation of graduate programs was scheduled to occur. The new process is detailed in CAPES (2020a), and the model reflected the recommendations of the working groups in the "Humanities" (Amado et al., 2020) and "Life Sciences" and "Exact Sciences..." (Santos et al., 2020). Some key results detailed in the final document relate to the calculation methods, which can be summarised as follows:

i) Journal percentiles are collected from Scopus CiteScore and Clarivate's Journal Citation Reports (JCR). Journals are ranked according to percentiles by comparison, after normalisation of their impact factors within 334 subsubjects (Scopus) or 235 categories (WoS).

ii) Google Scholar's h5 is used to calculate percentiles equivalent to those from Scopus and WoS, in case the journal is not covered by any of those. The imputations are calculated using regression models described in the documentation. Sometimes, a cap is applied to the results to avoid overvaluing journals considered not to be international enough, meaning that they are not indexed by the two preferred databases.

iii) The classification of "Humanities" journals considered:
- Existence of subareas for normalisation;
- Indexing by discipline-specific databases (i.e., "Business..." used indicators from the Scientific Periodic Electronic Library – SPELL);
- Optional use of the impact factor or reputation indicators;
- Consideration of publication languages in normalisation procedures;
- Use of h5 or h10, to better reflect citation practises in each area.

iv) The stratum *C* will include all journals that do not have any of the indicators adopted by the model, or that do not meet good publishing practises.

The presented selection of indicators is far from groundbreaking, going against principles from responsible research movements represented by the San Francisco Declaration on Research Assessment – DORA (ASCB, 2012) and the Leiden Manifesto for research metrics (Hicks et al., 2015). Furthermore, the proposal also takes Brazil in the opposite direction of evaluation systems like the one existing in The Netherlands, where indicators such as the Journal Impact Factor and the h-index are banned or restricted (VSNU et al., 2020).

Nonetheless, CAPES implemented the calculations to provide the classification suggestions used for the 2016-2020 Qualis. Figure 7.4 shows how the evaluation committees dealt with the suggestions.

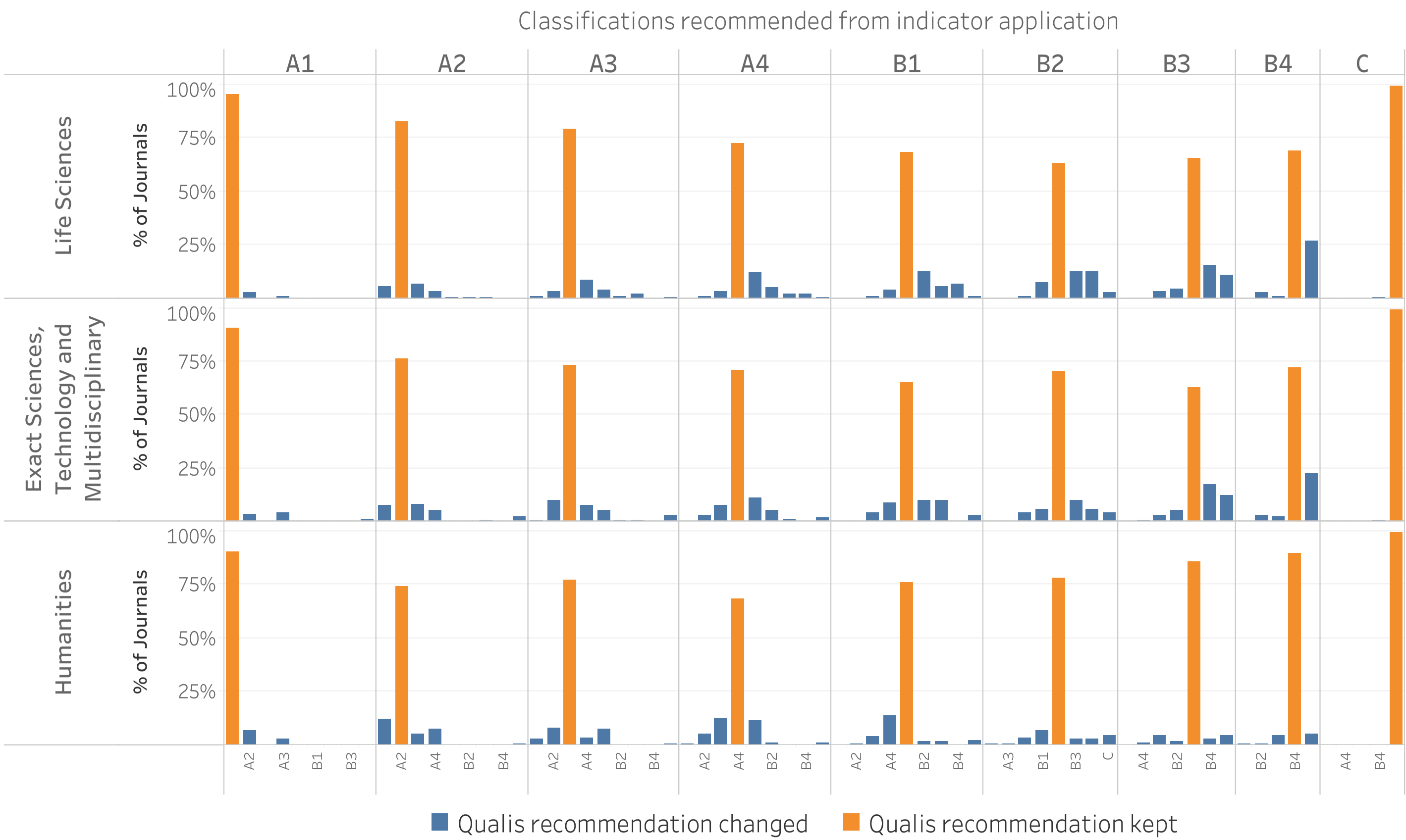


Figure 7.4.: Final reclassification of journals from the indicator-based suggestions to the peer review results, grouped by broad areas

Figure 7.4 is surprising at first glance, especially when compared to Figure 7.3. The chart covering the final 2016-2020 Qualis shows that committees in the three broad areas accepted most of the indicator-generated journal classifications offered by CAPES. Although little has changed for "Life Sciences" and "Exact Sciences...", the "Humanities" went from questioning and reviewing most indicator classifications to a level of agreement even higher than that of the other two groups. Could CAPES have found the perfect journal ranking system that makes expert review nearly irrelevant? A deeper investigation of the methodology described in CAPES (2020a) reveals that this is not the case.

According to CAPES (2020a, p. 10), "Each area can adjust its classification, observing a maximum of 30% of changes to the strata, of which 20% can be made in up to 1 level and 10% in up to 2 levels". Therefore, the apparent alignment between indicator classification and expert analysis seen in Figure 7.4 is more a representation that areas are not free to reclassify journals as needed. The limitations imposed prevent us from seeing the level of agreement between indicators and evaluators in the final Qualis, as was possible in its mid-term version shown in Figure 7.3.

7.4 The persistent need to evolve

There is no doubt that the latest iteration of Qualis has been an advancement. Historical problems derived from multiple and extremely diverse classifications of the same journals have been solved. With the mother area approach, evaluation can focus on quality and the classification ceases to represent a one-stop system where a magic stratum should capture quality, relevance, and pertinence. Finally, the adoption of an indicator-based reference that can be reviewed by experts has the potential to be a step forward, despite the evident flaws observed; that is, better indicators should be considered in future iterations of the classification, and the limited room committees have to question such indicators should be revisited.

However, while the discussion on seeking further evolution is relevant, it is also important to reflect if there is still a need for a Qualis at all. Arguments for the end of Qualis find support from many Brazilian academics, and the idea has been registered by a special committee in charge of monitoring the most recent National Plan for Research and Graduate Education (PNPG) in Brazil. In their final report, with reflexions on Qualis as part of research evaluation in the future, the committee stated:

> Qualis has played an important role in the qualification of graduate production. Given the new resources and features in IT tools and the consolidation of databases that have emerged in the time between the last triennial evaluation and this moment that we approach the second quadrennial evaluation and the maturity achieved in the SNPG, we have a new context. In this context, the evaluation of production in each evaluation area should be based on established international metrics widely accepted by the community. In this sense, the Committee proposes the extinction of Qualis after the next evaluation cycle (2021-2024) (PNPG Committee, 2020, p. 25).

There are three important points in the recommendation of the committee. The first relates to the available technology and the advances in tools. Indeed, while Qualis evaluation at the journal level has been a manifestation of a lack of resources and data to directly evaluate published papers, now there are not only resources, but also sophisticated indicators that can be used for a quality assessment that is a certain improvement over the limitations of journal evaluations (Polonsky and Whitelaw, 2005).

The second point relates to what is forgotten by the committe when they recommend using established international metrics, and to what they mean when talking about databases. As highlighted by (Brasil, 2021b), databases such as the Web of Science have uneven coverage of Brazilian publishing in evaluation areas, failing to include a significant part of the papers published in Portuguese, which are important for communicating with local audiences on topics of local relevance. It is wrong to believe that only science with an international impact can be of value, and evaluation must be responsible to not push researchers away from local impact, forcing them to chase metrics under penalty to see their work unrecognised and unrewarded. Databases do not need to be limited by Web of Science or Scopus, as there are alternatives with a better coverage, such as Dimensions and OpenAlex, that can be used for more detailed and relevant publishing analyses.

The third point is whether Qualis should be extinguished. If Qualis is seen as a ranking that reduces the complexity of scholarly publishing to a set of simple strata, constantly misused beyond its intended application of helping conduct a complex, multidimensional evaluation of graduate courses; then, yes, it may be time for Qualis to disappear. However, Qualis is a broader strategy to evaluate scientific production. Today there is a Qualis for books, another for events, for artistic and for technical/technological production. As with the Qualis journals, they are all based on applying a combination of quantitative and qualitative methods, combined with expert analysis, to help separate the good from the bad in the challenging process of evaluating graduate programs in Brazil.

Brazil cannot abandon a system that allows the valuation of journal production of national and international relevance, and also of a broader set of products that fit within DORA recommendations for responsible evaluation (ASCB, 2012). Qualis should remain as a means of combining the best available expert review with methods and indicators that can be applied to Brazilian reality, considering the scale of the SNPG and the limitations it imposes on an evaluation system. To abandon that in exchange of the suggested international metrics will probably hurt the country's efforts towards a science system that aims both to expand the frontiers of knowledge and produce societal impact.

7.5 Conclusion

This chapter has delved into the intricacies of the Qualis system, analysing its historical development and implications as a crucial part of evaluating research and graduate education in Brazil. Qualis was established around 25 years ago to incorporate a qualitative dimension into the assessment of the country's expanding scientific production. Considering that resources were limited, an article-level analysis of scholarly production was not feasible, so Qualis has been an imperfect but necessary solution to classify scientific journals as a proxy of the expected quality of articles published by Brazilian researchers.

Although Qualis has experienced numerous changes since its inception, it continues to grapple with the challenge of balancing objectivity and precision with the unique characteristics of scholarly work across various disciplines. In its latest iteration, Qualis addressed some of its notable historical shortcomings.

For instance, different evaluation areas used to classify journals into distinct strata, factoring in not just quality, but also pertinence and relevance as ranking adjustments. While these factors are essential in an evaluation, it is improper to artificially elevate a journal's rank to incentivise more submissions, despite its diminished quality. Similarly, it is unjust for areas to downgrade a journal's ranking merely because it is from a different discipline.

The new Qualis now empowers evaluation areas with the responsibility to classify the journals that belong to them, and the resulting ranking is unique, being adopted in the assessment of all areas. The advance created a new challenge, as the methods adopted for classification in different areas must be comparable and valid across all disciplines. For that, quantitative methods are now applied, relying on indicators such as CiteScore, JCR and Google Scholar's h-index to suggest classifications to expert committees, which can make the necessary adjustments through a peer review process. However, the quantitative approach seems to have stripped the committees of their voice, since only 30% of the calculated rankings can be modified in the peer review phase, with any modifications also being limited in range.

However, the importance of having a framework that enables the valuation of scientific production should not be underestimated. While some call for the extinction of Qualis and subsequent replacement by international metrics from which many developed countries try to escape, the system remains vital to foster an inclusive and equitable academic landscape in Brazil. Although established international metrics offer greater objectivity and acceptance for the evaluation of scientific production, they may fail to capture the unique context and challenges faced by Brazilian researchers and graduate programs. Furthermore, an overemphasis on international metrics can discourage researchers from focussing on locally relevant research, often produced in the local language, which is crucial to address the needs and concerns of the Brazilian society.

Moving forward, it is essential not to abandon, but to refine and enhance the Qualis system by drawing on the best available methods and indicators tailored to the Brazilian context, as well as recognising the importance of expert review. By doing so, we can work toward creating a more inclusive, equitable, and comprehensive evaluation framework that supports and values the diverse forms of scientific production that contribute to the growth and development of the Brazilian graduate system and the broader scientific community.

References

ABRASCO. (2019). Apreciação da proposta de Qualis Periódico Referência. Retrieved March 2023, from https://bit.ly/3mSLm2a

Amado, A. M., Tomanari, G. A. Y., Fernandes, L. M. R., Dalmonte, E. F., Magalhães, J. S. d., Siqueira, V. B. C., ... Oliveira, T. M. (2020). *Relatório do GT do Qualis Periódicos Humanidades*. CAPES. Brasília, DF. Retrieved April 12, 2023, from https://bit.ly/3KSUzAF

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

Barata, R. B. (2016). Dez coisas que você deveria saber sobre o Qualis. *Revista Brasileira de Pós-Graduação*, *13*(30), 13–40. https://doi.org/10.21713/2358-2332.0.947

Barata, R. B. (2019). Necessary changes in the evaluation of graduate programs in Brazil. *Interface - Comunicação, Saúde, Educação*, *23*. https://doi.org/10.1590/interface.180635

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2021b). Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases. In *Proceedings of the 18th international conference on scientometrics & informetrics* (pp. 193–204). Leuven, Belgium: KU Leuven. Retrieved from https://bit.ly/3jOHQBR

Brasil, A. (2023b). Rethinking a national classification of research and graduate education. *Profesional de la Información*, *32*(2), e320224. https://doi.org/10.3145/epi.2023.mar.24

Brigatti, F. (2023). Economistas se rebelam contra mudança em ranking de revistas científicas. *Folha de São Paulo*. Retrieved from https://bit.ly/3AesptT

CAPES. (2003). *QUALIS: Concepção e diretrizes básicas*. Brasília. Retrieved July 24, 2020, from https://bit.ly/3u77Iuy

CAPES. (2018b). Avaliação da produção intelectual de programas de pós-graduação. Seminários: Repensando a Avaliação. Retrieved February 2022, from https://bit.ly/3AYGFqb

CAPES. (2018d). *Portaria nº 150, de 4 de julho de 2018. Institui o Grupo de Trabalho (GT) do Qualis Periódicos*. Diário Oficial da União. Brasília, DF.

CAPES. (2018e). *Portaria nº 271, de 14 de dezembro de 2018. Institui o GT do Qualis Periódicos das Humanidades* (Ordinance No. 271/2018). Diário Oficial da União. Brasília, DF. Retrieved April 21, 2023, from https://bit.ly/3mW7bhh

CAPES. (2019b). Meio-Termo: refletir para avançar. Retrieved April 21, 2023, from https://bit.ly/41rui2k

CAPES. (2019c). *Ordinance nº 8, de 31 de janeiro de 2019. Institui o GT do Qualis Periódicos das Ciências da Vida e Ciências Exatas, Tecnológicas e Multidisciplinar* (Ordinance No. 8/2019). Diário Oficial da União. Brasília, DF. Retrieved April 21, 2023, from https://bit.ly/3mUUvXY

CAPES. (2020a). *Metodologia do Qualis Referência: Documento base aprovado no CTC 198, de setembro de 2020*. CAPES. Brasília, DF. Retrieved April 20, 2023, from https://bit.ly/41w96rW

CAPES. (2020d). Sobre as áreas de avaliação. Retrieved April 9, 2021, from https://bit.ly/3s0XB9f

CAPES. (2023b). *Documento técnico do Qualis Periódicos*. CAPES. Brasília, DF. Retrieved April 12, 2023, from https://bit.ly/3KTmUGY

CAPES. (2023c). Metodologia do Qualis Referência: Quadriênio 2017-2020. Retrieved April 12, 2023, from https://bit.ly/3my156n

CAPES. (2023d). Qualis Periódicos. Brasília, DF: CAPES. Retrieved from http://qualis.capes.gov.br

Comissão Especial de Acompanhamento do PNPG 2011-2020. (2018). *Proposta de aprimoramento do modelo de avaliação da PG*. CAPES. Brasília, DF. Retrieved from 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*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3wyR097

Ferrari, L. (2023). Pesquisadores criticam novo 'ranking' de revistas científicas da Capes, o Qualis. *Estadão*. Retrieved from https://bit.ly/41EGuMY

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193. https://doi.org/10.1023/a:1014848323806

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

Hortale, V. A. (2003). Modelo de avaliação CAPES: desejável e necessário, porém, incompleto. *Cadernos de Saúde Pública*, *19*(6), 1837–1840. https://doi.org/10.1590/s0102-311x2003000600027

Leite, M. F. B., Viana, A. B. N., & Pereira, G. G. (2010). Métodos quantitativos na avaliação da CAPES: Uma pesquisa bibliográfica. *FACEF Pesquisa - Desenvolvimento e Gestão*, *9*(2), 166–174. Retrieved from https://bit.ly/2ZEYg3X

Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). *Revista Brasileira de Sociologia*, *6*(13), 9–26. https://doi.org/10.20336/rbs.256

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*. Brasília, DF: CAPES. Retrieved from https://bit.ly/2PmAf0q

Pascutti, P. (2016). Entendendo o Qualis Periódicos CAPES. Campinas, SP. Retrieved April 2022, from https://bit.ly/40leSuZ

Polonsky, M. J., & Whitelaw, P. (2005). What Are We Measuring When We Evaluate Journals? *Journal of Marketing Education*, *27*(2), 189–201. https://doi.org/10.1177/0273475305276842

Reategui, E., Pires, A. d. S., Carniato, M., & Franco, S. R. K. (2020). Evaluation of Brazilian research output in education: confronting international and national contexts. *Scientometrics*, *125*(1), 427–444. https://doi.org/10.1007/s11192-020-03617-z

Ribeiro, R. J. (2022a). Carta da SBPC à CAPES sobre audiência pública para debater a Avaliação Quadrienal. Retrieved from https://bit.ly/3Ki9QYx

Ribeiro, R. J. (2022b). Depoimento do presidente da SBPC sobre a Avaliação Quadrienal da CAPES. Rio de Janeiro. Retrieved from https://bit.ly/3tp0FyM

Santos, P. J. P. d., Oliveira, T. M. d., Paixão, F. d. A. J., Pascutti, P., Amado, A. M., Oliveira, D. d. H., . . . Guterres, S. S. (2020). *Relatório do GT Qualis Periódicos*. CAPES. Brasília, DF.

Soma, N. Y., Alves, A. D., & Yanasse, H. H. (2016). O Qualis Periódicos e sua utilização nas avaliações. *Revista Brasileira de Pós-Graduação*, *13*(30), 45–61. https://doi.org/10.21713/2358-2332.2016.v13.1128

Spagnolo, F., & Souza, V. C. (2004). O que mudar na avaliação da Capes? *Revista Brasileira de Pós-Graduação*, *1*(2), 8–34. https://doi.org/10.21713/2358-2332.2004.v1.38

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

Yamashita, M. (2023). Lista Qualis e a duvidosa qualidade dos critérios de qualidade. *Questão de Ciência*. Retrieved from https://bit.ly/3UShMq1

8 Beyond the Web of Science

> *Many researchers in the developing world feel trapped in a vicious circle of neglect and – some say – prejudice by publishing barriers they claim doom good science to oblivion.*
>
> — **W. Wayt Gibbs**

Brazil employs a national evaluation system which is conducted by the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES). As the name of the agency suggests, such evaluation focuses on masters and doctoral courses, but the Brazilian Society for the Advancement of Science (SBPC) indicates over 95% of all Science & Technology research conducted in the country comes from these graduate courses. Thus, it is possible to state that evaluating graduate education is nearly the same as assessing the whole science system in Brazil (Nobre and de Freitas, 2017; SBPC and ABC, 2020b).

First conceived in the 1970s, the evaluation of graduate programs was initially intended to better allocate funding for the development of science in the country. Over time, the model grew in size and complexity, becoming a high-stakes assessment of almost 8000 master's and doctoral courses organised into nearly 5000 research units or programs (PPG). Every four years, all research programs in the country go through a compulsory, government-funded evaluation that grades them on a scale from "1" through "7"; the last representing the highest level of excellence. Positive results not only guarantee status for the PPG and the institutions that promote them, but also implicate in an increase in funding,

This chapter has been published as: Brasil, A. (2021). Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases. In *Proceedings of the 18th International Conference on Scientometrics & Informetrics* (pp. 193–204). Leuven, Belgium: KU Leuven.

a higher number of available scholarships, and access to a wider range of grants. In comparison, a low performance will not only lead to funding cuts, but it can also threaten the programs' existence. For instance, research units graded "1" or "2" are no longer allowed to enrol new graduate students and must suspend their activities completely as soon as the last of the currently enrolled students graduate (CAPES, 2017b; Nobre and de Freitas, 2017).

While an additional study is being conducted to explore the complexity of the national evaluation system in Brazil, for this paper it is relevant to know that the actual evaluation relies heavily on mixed methods. Disciplinary and interdisciplinary panels from 49 distinct evaluation fields perform the assessment of PPG with qualitative and quantitative data collected yearly from every research unit in the country. From the available evidence, several dimensions are assessed by the committees in charge, from infrastructure to educational results and societal impact of the PPG. Scientific production is, as expected, one of the most valued dimensions in the process, including specific evaluation of distinct types of output such as books, technical and artistic products, conference proceedings, and papers, among others.

The assessment of papers published by graduate programs in the country follows the principles of the broader evaluation system: each of the 49 research areas receives a comprehensive list of each PPG's publications, together with a series of indicators gathered and calculated from national and international databases. Panels composed by established researchers in each area interpret the available information – combined with their inherent knowledge in the field – and generate what is known as Qualis Journals: a nine-level classification system for all journals used by Brazilian researchers during the evaluation cycle (Barata, 2016; CAPES, 2021c).

Qualis has been an essential element in Brazilian evaluation since 1998. The classification system evolves over time to reflect advancements in scientific publishing as well as in information systems (Barata, 2016). To promote changes in the current evaluation cycle (2017–2020), CAPES established a working group to review Qualis procedures (CAPES, 2018d). In a preliminary report, the group states that the new assessment model must induce internationalisation of both publishing and journal indexing. The initiative proposes to reduce the qualitative perspective, attributing to journal-level indicators – such as *jif* and *h5* – most of the responsibility for the classification (dos Santos et al., 2019).

Considering the evaluation literature strongly discourages practices which rely heavily on indicators to assess scientific productivity, especially those dependent on journal-level metrics (Hicks et al., 2015), it is not surprising that a consensus has not been found among the research-area representatives working with CAPES. Several disciplines, especially those in the Social Sciences and Humanities (SSH), are concerned with the proposal for a 'New Qualis', especially because the current focus is on adding weight to the value of publications indexed by databases such as the Web of Science (WoS) and Scopus, while undervaluing those indexed by regional databases, mostly in local language. As a consequence, even though ordinance no. 150 (CAPES, 2018d) established a three-month deadline for publication of a report by the Qualis Journals working group, almost three years have passed, and the document remains unpublished.

The main goal of this research is to contribute to this current debate around the 'New Qualis', primarily by investigating topics that may shed light over concerns such as: i) How is the coverage of Brazilian papers in international databases? ii) How representative is the local language for the country's publications, and how different is the coverage from regional to international databases? iii) In case significant differences are found, how extreme are the disciplinary variations? iv) From a thematic standpoint, are there research topics that have more space in regional databases?

8.1 Methods and data

The central data set that supports the current research comes from all Brazilian PPG. It is a virtually complete list of all publications produced in these programs, by both faculty members and student body alike. There are two main reasons why this data set covers the majority of the actual PPG output. The first one lies on the mentioned high stakes of evaluation. Since performance relates to funding and to the continued existence of research programs, their directors, university pro-rectors, and researchers are concerned with the quality of the data provided. The second reason is that, for many years, CAPES has counted on information systems to collect PPG data to perform the evaluation. Distinct systems have been developed over time and the data collection is now conducted through the Sucupira Platform: an integrated system that is robust enough to deal with the size of the Brazilian National System of Graduate Education. The platform is open continuously for data submission and it grants the general public direct and real-time access to all of it. This means all PPG researchers and stakeholders become part of a relevant auditing system. As such data is subsequently audited by the committees in charge of the evaluation, there is also an authenticity layer of control (CAPES, 2021c; Siqueira, 2019).

Integrated data sets from graduate education in Brazil are made available through CAPES ' Open Data platform (CAPES, 2021a). In this research, the R language (R Core Team, 2021) was used to combine and clean available data sets which relate to four categories in the database: i) general information from graduate programs; ii) scientific output from graduate programs; iii) detailing of bibliographic production from PPG; iv) authorship details of papers and reviews. The most recent Qualis rankings were subsequently gathered from the Sucupira Platform (CAPES, 2021c) and combined with the broader data set.

The resulting data include all papers published from 2013 to 2018, totalling more than 1,3 million records of around 750 thousand individual publications. The reason for the difference is that CAPES' records are PPG-based, which means publications co-authored by researchers from distinct research programs are recorded for each individual PPG. Publications from 2019/2020 were not yet included in the study because data for those years are only partially available, with relevant details such as DOI and ISSN still missing (CAPES, 2021a).

Once the work on the core data was completed, a series of complementary databases were consulted to enrich the resulting data set. For that, key fields like DOI, ISSN, and additional information on author last names, journal volumes/issues, page numbers and titles of publications were all used to match the consolidated data with the following additional sources:

i) SciELO – bibliographic database structured on a cooperative model of Open Access (OA) journals, which are selected based on a set of quality and operational criteria. The network focuses on the scientific communication needs of developing countries, especially in Latin America and the Caribbean. It was established in Brazil and now extends to 15 other countries, including Portugal, Spain, and South Africa (SciELO, 2021).

ii) RedALyC – similar to SciELO, it is a network of OA journals focused on scientific output from Latin America and the Caribbean, while also extending to other Portuguese and Spanish speaking countries. It reaches 26 countries and establishes a series of criteria to include journals in its database (RedALyC, 2021).

iii) Latindex – another network focused on OA publishing. It reaches 23 countries, mostly Spanish or Portuguese speaking, counting with a directory of 29.192 indexed journals. Even though the broader coverage may be useful to map Brazilian output, only a small percentage of the journals have been through the compliance process with the Latindex methodology, which includes criteria similar to those adopted by the other networks mentioned. These quality-assured journals form a Latindex subset of 2265 journals, known as Catalogue 2.0 (Latindex, 2021).

The enriched data set created allowed a series of analyses on the coverage of the regional databases, publication languages, research areas, and more. Nevertheless, at least one international database was also needed, and the Web of Science (WoS) was chosen to complement the research. According to Chavarro et al. (2018), this database is one of the main data sources used to obtain bibliographic data for many quantitative research assessments, and this includes the impact factors which inform the Brazilian Qualis classification.

8.2 Findings and discussion

From 2013 to 2018, CAPES' database of scientific production lists a total output of 752.453 papers and reviews, published in 29.679 distinct journals (merging ISSN for electronic and print versions). All of them were classified under the Qualis criteria which, in the current scale, consists of grades A1, A2, A3, A4, B1, B2, B3, B4, and C/NP (A1 being the top grade). Each disciplinary committee determines the score attributed to each of the possible grades, and individual publications receive corresponding points during the evaluation process, except C/NP. This rank is reserved for low quality or predatory journals, as well as those which the commit-tees consider lacking scientific rigour (e.g., non-peer-reviewed publications) (CAPES, 2021c).

To best represent the universe of valid journals, all results in this paper exclude those which were ranked C/NP, as well as any others which may have been disqualified by the evaluators during the data auditing process. As a consequence, the number of considered papers drops to 585.945, published in 23.508 journals. The filtered CAPES database was then matched with the Web of Science, following the guidelines described by Visser et al. (2021). A total of 9.597 journals (40,8%) were found in the database, with 298.170 papers published (49,4%). Figure 8.1 shows the results of this analysis.

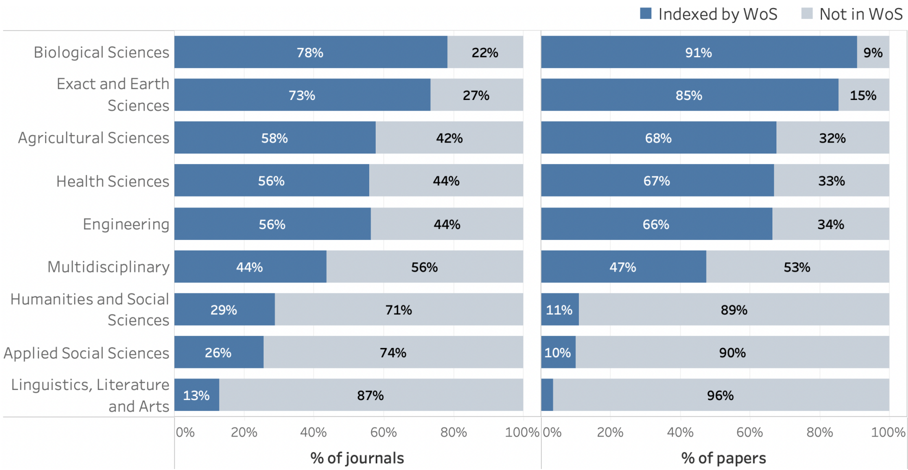


Figure 8.1.: Percentage of journals and papers indexed by WoS (2013-2018), according to research areas adopted by CAPES

As Figure 8.1 shows, there is a considerable difference in WoS coverage when results are grouped according to the nine research areas defined by CAPES: a meso-level aggregation of the 49 evaluation fields conducting the evaluation process. As it can be seen, 78% of the journals used by Biological Sciences PPG to publish their papers (2013-2018) are indexed by WoS. In contrast, the same is true for only 13% of those from Linguistics, Literature, and Arts.

From the papers' perspective, it can be noted that the number of publications in research areas with broader WoS coverage tends to be even higher. For instance, 73% of Exact and Earth Sciences journals are indexed by WoS, and this number rises to 85% when looking at the paper level. Opposite to that are all Social Sciences and Humanities (SSH), where the percentage of papers in the Web of Science is even smaller than their journal coverage, indicating a lower-than-average number of publications in such journals. Of course, caution should be used to interpret disciplinary differences observed in Figure 8.1. Scholars such as Tijssen et al. (2006) and Chavarro et al. (2018) have already highlighted the dangers in analysing differences in database coverage across research areas, as large variations are often influenced by internal biases in the database's own coverage. Nevertheless, for the trained eye, Figure 8.1 seems to paint a very extreme picture.

To help analyse if the disciplinary variation in the Web of Science coverage is more prominent for Brazilian scientific publications, Figure 8.2 was designed from data collected from the bibliometric version of the WoS core collection hosted at the Centre for Science and Technology Studies (CWTS). All papers and reviews from the 2013-2018 period were included in the visualisation designed using the VOSviewer software (van Eck and Waltman, 2009), and the publications were grouped according to the micro-level field classification (Waltman and van Eck, 2012).

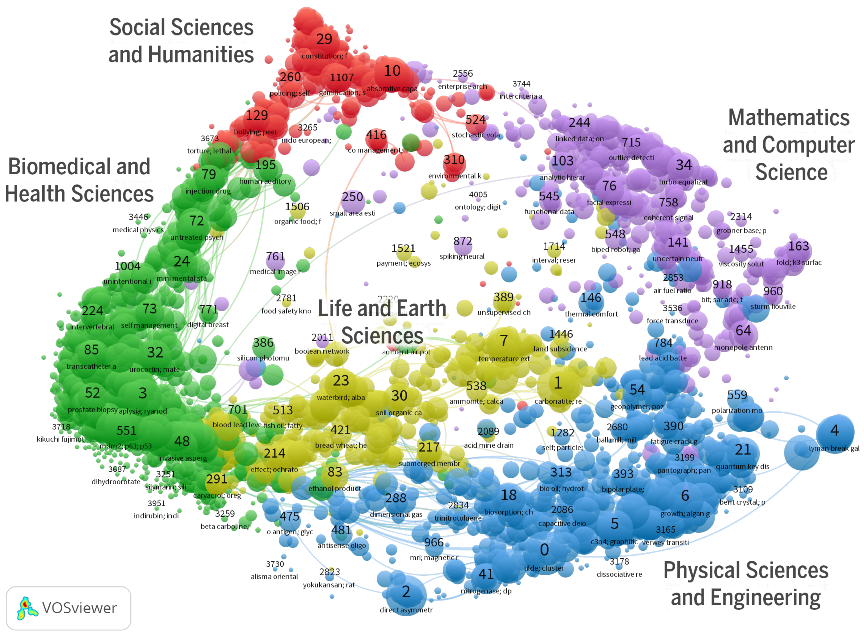


Figure 8.2.: Scientific production indexed by the Web of Science (2013-2018) according to CWTS micro-level field classification

Figure 8.2 shows 4013 clusters of publications in the WoS database. The size of the circles represents the number of publications in each cluster and their positions are calculated from the citation traffic between all papers. Five main fields of science are shown, and they confirm that there is a proportionally smaller number of SSH clusters and publications in the database. To see how Brazilian scientific publishing relates to the broader WoS coverage, Figure 8.3 maintains the configuration from the previous map but including only Brazilian papers and reviews, with resized circles according to the proportion of the country's available publications.

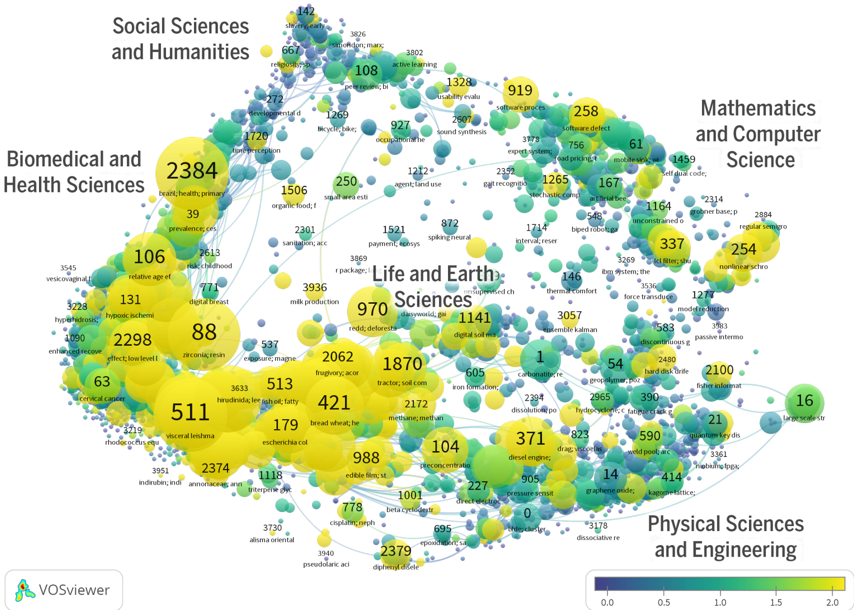


Figure 8.3.: Brazilian scientific production (2013-2018) mapped according to CWTS micro-level field classification

The size variation of clusters in Figure 8.3 is evidently more pronounced than in the previous map, and the colour overlay helps understand the significance of that. Cluster colours pointing to 1.0 in the included scale imply the proportion of Brazilian publications is around the same as for the global database. Lighter colours mean Brazil produces a higher proportion in these fields, and the darker colours mean the exact opposite. The resulting visualisation shows Brazilian publishing is proportionally higher across the Biomedical and Health Sciences, and in some Life and Earth Sciences clusters. Isolated points of higher relative productivity can be found in other research areas, but most other clusters are underrepresented.

The main message from Figure 8.3 seems to be that Brazilian publishing behaviour within WoS amplifies the coverage biases already observed in the database. To confirm this perception, the internal coverage calculation described by Moed and Visser (2007) is applied. This indicator analyses the extent to which the articles included in the WoS cite other articles which may or may not be found in that same database. From the number of missing references, it is possible to infer the percentage of scientific publications that are not indexed by the Web of Science. Figure 8.4 shows the results of the analysis, but grouping the CWTS field classification systems at their macro-levels, for clarity purposes.

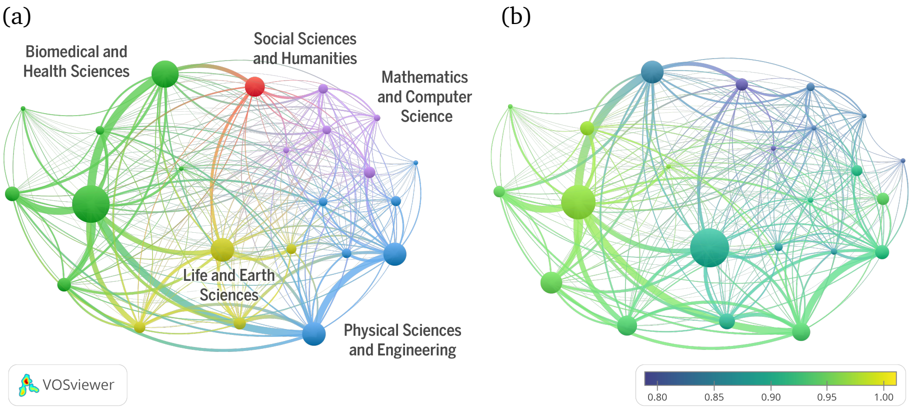


Figure 8.4.: Scientific production mapped according to CWTS macro-level field classification for (a) complete WoS database (2013-2018), and (b) only Brazilian publications, with relative internal coverage overlay

Figure 8.4a shows the 23 broad disciplines of CWTS macro-level field classification. The colour clustering follows the scheme from Figure 8.3, and the size of each circle indicates the number of papers in the WoS database (2013-2018). To design Figure 8.4b, the data was filtered to show only Brazilian publications in proportionately resized clusters. As in the previous maps included in this paper, it is possible to see differences in the publication distribution across fields from (a) to (b). For instance, Brazil produces relatively more in most Life and Earth Sciences clusters, but much less in the single SSH cluster.

To design the colour overlay on Figure 8.4b, the average internal coverage (Moed and Visser, 2007) was calculated for the 23 broad disciplines in both data sets: the complete one in (a) and the Brazilian one in (b). The results show that the internal coverage in (a) for the largest Biomedical and Health Sciences cluster is of 91%, for the most prominent Life and Earth Sciences is of 73%, while for the single SSH cluster is of only 34%. The numbers for Brazil are slightly lower in (b), at 88%, 65%, and 27%, respectively. This means Brazilian publications are citing a larger percentage of works not indexed by WoS.

A second step for the colour overlay in Figure 8.4b was to calculate the ratio between the broader database internal coverage and the Brazilian one. For SSH, for instance, this would mean $0,27/0,34$, indicating Brazil's relative coverage is of 79% of the outcome seen for the unfiltered database. The results displayed in the overlay show Brazilian internal coverage is lower across all 23 broad disciplines, and the largest variations are seen exactly on clusters that already registered low coverage in the broader database. This analysis confirms the original impression that coverage biases in the Web of Science are exacerbated in Brazilian output. Is this a result of publishing barriers or a matter of choice to publish in journals indexed elsewhere?

Maybe the answer can come partially from scholars such as Gibbs (1995) and Tijssen et al. (2006), who already discussed challenges that researchers in the developing world need to face to overcome publishing barriers. Among them are both disciplinary and language obstacles, as English journals dominate databases such as the Web of Science, with much less room for Portuguese or Spanish publications. This is a significant problem for Brazil, as the most recent Education First Proficiency Index (2020) – which ranks 100 regions according to English Skills – places the country in the low proficiency group (53rd position). Besides that, the British Council (2014) reports only 5% of Brazilians have some knowledge of English, while only $^1/_5$ of those present some level of fluency. The numbers are also not good, yet less abysmal, when researcher population is considered. According to de Vasconcelos (2008), who analysed data from the *curricula vitae* of over 50 thousand Brazilian scientists, around 33% of them declared to be proficient in English. Even disregarding biases from the self-declaration of proficiency seen in the CVs, still only a third of Brazilian researchers would not find language an obstacle to publish in English.

Considering these limitations, Figure 8.5 shows all Brazilian papers from 2013-2018, isolating publication percentages according to WoS coverage and the main language reported for the individual papers. Results are also shown according to the research areas adopted by CAPES' classification system.

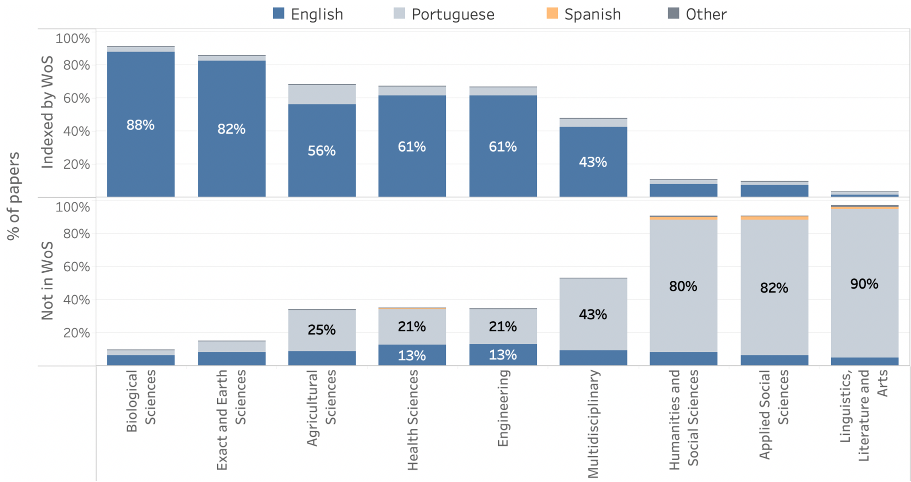


Figure 8.5.: Comparison of language profile of Brazilian papers (2013-2018), in and out of WoS, according to research areas adopted by CAPES

Disciplinary differences are, once again, strikingly evident in Figure 8.5. Most of the Brazilian papers indexed by the Web of Science are in English, and publications not covered are mostly in Portuguese. SSH papers in local languages account for more than 80% of publications in all three research areas. The Multidisciplinary field is also impressive as publications are nearly split in the middle regarding WoS coverage, with language participation mirroring each other. Figure 8.5 also shows that there is only a very small share of papers in Spanish or other languages, which add to less than eight thousand publications (a little more than 1% of the total).

After identifying that most papers not indexed by the Web of Science are in Portuguese, even in research fields such as Health Sciences, we wonder if such publications would be found in regionally relevant databases. Figure 8.6 shows the results of such analysis by mapping the whole set of Brazilian publications (2013-2018) with the Latindex, SciELO, and RedALyC databases.

The two polar charts included in Figure 8.6 show the percentage of publications in each of the CAPES-adopted research areas that can be found in Latindex, SciELO, and RedALyC. The left chart shows non-English publications, which we have seen are mostly in Portuguese. The broader Latindex directory covers a significant percentage of papers not found in WoS, reaching nearly 80% in the

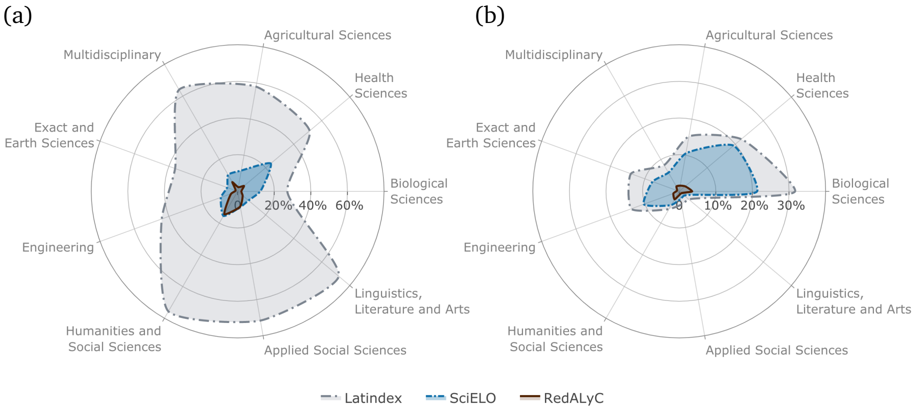


Figure 8.6.: Brazilian papers (2013-2018) not indexed by WoS but available at Latindex, SciELO or RedALyC: (a) Non-English and (b) English

Humanities and Social Sciences, with its lowest coverage in Biological Sciences, where it is around 30%. SciELO and RedALyC, counting with more curated collections, cover a smaller share, but over 20% of the 'missing' Health Sciences papers are included in SciELO, for example.

For English publications not included in the WoS collection, Figure 8.6b shows that there is very little to find in RedALyC, but Biological and Health Sciences have a significant share of papers in SciELO and Latindex. The much larger Latindex coverage that was seen in Figure 8.6a is no longer present, which may indicate the predominance of Spanish and Portuguese journals in this bibliographic database.

By the investigation of only three established Latin American databases, it becomes evident that the local language scientific publishing is finding alternative outputs for publication. From this initial inspection, other databases might be explored, and a study by Alonso-Gamboa and Russell (2012) highlighted some interesting alternatives for the future. Among them are LILACS, a Health Sciences database covering 901 journals from 26 countries: and PERIÓDICA, a data-base of around 1500 Science & Technology journals from Latin America and the Caribbean.

After that, there is still a pertinent question to answer: Are Brazilian researchers investigating significant research topics which are only covered by regionally relevant databases? To answer that, all titles from the 585.945 papers and reviews published in Brazil in 2013-2018 were exported to a corpus file which went through Neural Machine Translation into English. This corpus was used to generate a term map using VOSviewer software (van Eck and Waltman, 2009), which can be seen in Figure 8.7.

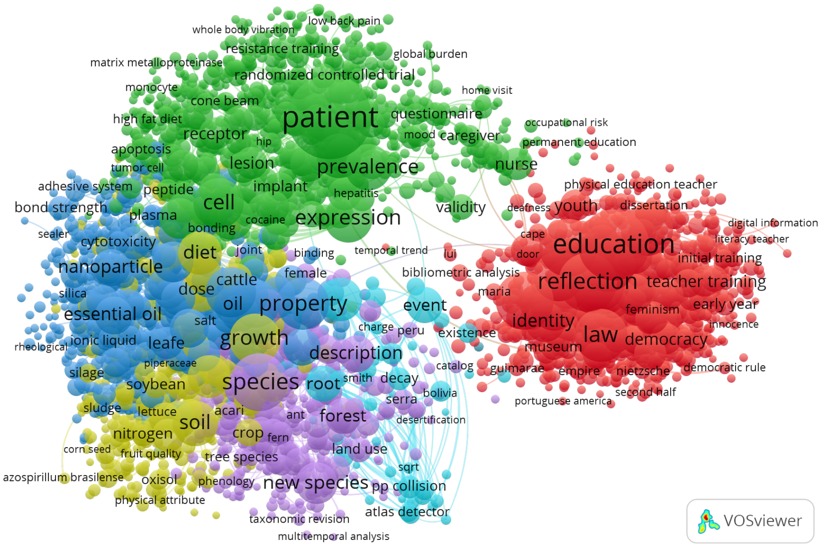


Figure 8.7.: Term map generated from titles of Brazilian papers (2013-2018)

Figure 8.7 shows six different colour clusters with an evident relation to some of the main fields of science. Biomedical and Health Sciences appears in green and it is quite a significant cluster, in line with what has been seen about the Brazilian publishing profile. Other clusters seem to cover Life and Earth Sciences, with relevant topics related to Agricultural and Environmental Sciences, and a minor light-blue cluster goes into Physical Sciences and Engineering. Another large cluster is detached to the right of Figure 8.7, showing in red what evidently represents Social Sciences and Humanities. The fact that SSH research topics seem to be isolated from the rest becomes clearer with a colour overlay applied to the term map, as displayed in Figure 8.8.

The overlay on Figure 8.8 was possible because, while exporting the publication titles to create the corpus, a score file was also produced, indicating if each individual paper was indexed by the Web of Science or not. From that, the SSH

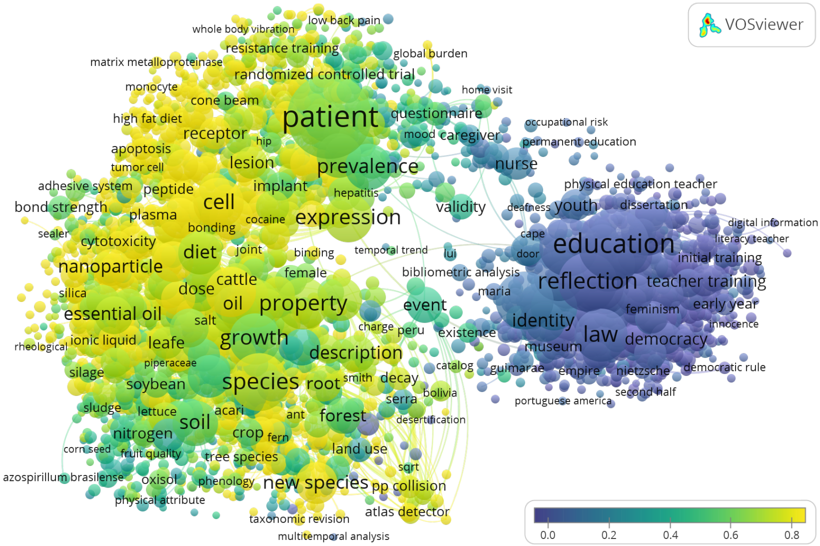


Figure 8.8.: Term map generated from titles of Brazilian papers (2013-2018), colour-coded by Web of Science coverage

island seems even more detached, as most clusters are close to zero in the scale, meaning nearly all publications on the subject are not indexed by WoS. A small group of clusters in the left border of SSH reaches around 20% coverage, being weak links to other main fields of science.

Considering the left group of research topics displayed in Figure 8.8, some of the Biomedical, Health, and Life Sciences clusters shown at the top left, in light colours, are quite noticeable, as nearly 80% of the occurrence of such topics are in WoS indexed papers. Several other clusters are shown in shades of green, which mean their incidence is relatively balanced regarding WoS coverage, fluctuating between 40% to 60% in the presented scale.

8.3 Conclusion

A recent study conducted by Chavarro et al. (2018) investigated the extent to which the inclusion in the Web of Science would be an indication of journal quality. Among their findings was that journals with similar characteristics and editorial standards often receive unequal treatment due to their place of publication, discipline, and language. The authors warn research evaluators against the assumption that WoS and similar databases would assess journals without bias, joining several other scholars such as Alperin (2014), Garfield (1995), and Mounier (2018) in a call for caution in developing countries, with particular emphasis on Latin America.

In Brazil, policymakers and evaluators are engaged in ongoing discussions over a proposal for changes in the assessment of journals and publications that would trade a significant share of the qualitative methods currently adopted in favour of more journal-level indicators from WoS-like databases. This study adds to the warnings about qualitative dimensions of such databases and shows further science can be found in regionally relevant databases as well, where topics overlooked by the Web of Science are covered.

The data presented here show that WoS indexes around 41% of the qualified journals Brazilian researchers use to publish their work, accounting for just under 50% of the country's papers. Disciplinary distribution is hugely unbalanced, and the analysis of internal coverage for the broader WoS versus the Brazilian set of publications shows the existing biases in the database are taken to extremes in Brazil's case. That means disciplines such as the Social Sciences and Humanities, with already low coverage in the Web of Science, are even more underrepresented in Brazilian publications.

When it comes to using WoS indicators for evaluation, Moed and Visser (2007) describe four possible types of bibliometric studies from an internal coverage inquiry. The first type is known as a 'pure' WoS analysis, as it would be possible to rely solely on WoS source journals to analyse citation impact when internal coverage is over 80%. Types 2 and 3 – with coverage-ranges at 60–80 and 40–60, respectively – would require different levels of expansion in the sources covered, whether by including target articles not published in WoS source journals (2) or adding articles in proceedings volumes from a range of subsequent years (3). In case WoS coverage in a field is below 40, the mere value of a citation analysis based on WoS data should be questioned, even if target or source universes are expanded (Type 4 study).

The internal coverage analysis at the micro-field level classification for Brazil reveals that the bibliometric studies recommendations for the 4013 clusters would be: Type 1 – 43%; Type 2 – 29%; Type 3 – 14%; Type 4 – 14%. The problem is that of the 1741 clusters eligible for a 'pure' WoS analysis, 979 are in the Biomedical and Health Sciences, and 626 are in the Physical Sciences and Engineering, while only 19 are in Mathematics and Computer Science and eight in Social Sciences and Humanities. Even though for Mathematics, 48% of the clusters suggest a Type 2 study, 64% of those in SSH are ranked as Type 4. Consequently, we should question the mere value of the provided indicators for any quality assessment.

While these findings should be enough for policymakers designing Brazilian evaluation to step back from expanding their reliance on indicators from WoS-like databases, this research has also shown that most of the Brazilian scientific output not indexed by the Web of Science is written in Portuguese. That is particularly important for Brazil, as only a third of the researcher population states to have achieved some level of English fluency. Local language publications mean more people can produce and consume science, and societal impact may be even more relevant than citation metrics.

Since a significant share of local language papers could be found in Latindex, SciELO and RedALyC, future studies on identifying other high-coverage regional databases are recommended. Such databases cover various research subjects, some of them widely overlooked or just absent from WoS indexed publications. Besides that, they usually have a core concern to promote the expansion of OA journals.

For instance, the Directory of Open Access Journals (DOAJ, 2021) – a community-curated online service with freely available data on OA details for 15.874 journals – reveals very interesting results when crossing its database with this project's data set. Around 62% of the Brazilian papers indexed by Latindex, SciELO or RedALyC were published in DOAJ listed OA journals. These are nearly 170.000 papers, of which almost 140.000 were in diamond OA, with no costs for authors. In contrast, only 25% of the set of articles indexed by WoS were published in DOAJ listed journals, and 63% of those were made open only through APC charges.

The presented numbers are consistent with a comprehensive study by Pavan and Barbosa (2018), which analysed OA publishing for Brazilian authored articles in WoS. Using data from the 2012-2016 period, the authors estimated 59% of OA papers in the database were in APC journals, at an average cost of USD 1492,27 per article. For a country like Brazil, where only 50% of PhD candidates receive monthly stipends (with values under USD 450), the high cost for OA publishing is a considerable problem. It's not that researchers would pay for publishing themselves, but the fact is the average APC cost of three papers could fund a whole year of stipends for a PhD candidate.

In conclusion, this research has shown that while adopting international indicators from established databases might seem like a good simple solution to improve local science, the reality is that databases such as WoS tell only half of the whole story. Brazilian scholarship is not only about state-of-the-art publications in top journals; it is also about regionally relevant topics, often destined to a Portuguese speaking audience; it is about access to publish and consume science, given the socioeconomical reality of the country. Brazilian science goes beyond the Web of Science, and a sound and comprehensive evaluation system should strive to capture its complexity, instead of trading it for restraining, short-sighted simplicity.

References

Alonso-Gamboa, J. O., & Russell, J. M. (2012). Latin American scholarly journal databases: a look back to the way forward. *Aslib Proceedings*, *64*(1), 32–45. https://doi.org/10.1108/00012531211196693

Alperin, J. P. (2014). Citation databases omit local journals. *Nature*, *511*(7508), 155. https://doi.org/10.1038/511155c

Barata, R. B. (2016). Dez coisas que você deveria saber sobre o Qualis. *Revista Brasileira de Pós-Graduação*, *13*(30), 13–40. https://doi.org/10.21713/2358-2332.0.947

British Council. (2014). *Learning English in Brazil: Understanding the aims and expectations of the Brazilian emerging middle classes*. São Paulo, Brazil. Retrieved from https://bit.ly/3b3lbeZ

CAPES. (2017b). *Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal*. Diário Oficial da União. Brasília, DF.

CAPES. (2018d). *Portaria nº 150, de 4 de julho de 2018. Institui o Grupo de Trabalho (GT) do Qualis Periódicos*. Diário Oficial da União. Brasília, DF.

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/

Chavarro, D., Ràfols, I., & Tang, P. (2018). To what extent is inclusion in the Web of Science an indicator of journal 'quality'? *Research Evaluation*, *27*(2), 106–118. https://doi.org/10.1093/reseval/rvy001

de Vasconcelos, S. M. R. (2008). *Ciência no Brasil: Uma abordagem cienciométrica e lingüística* (Doctoral dissertation, Universidade Federal do Rio de Janeiro).

DOAJ. (2021). Public data dump. Retrieved from https://doaj.org/docs/public-data-dump/

dos Santos, P. J. P., de Oliveira, T. M., Paixão, F. d. A. J., Pascutti, P., Amado, A. M., Oliveira, D. d. H., ... Mamiya, E. N. (2019). *Proposta do GT Qualis Periódicos*. Brasília, DF.

Education First. (2020). *EF English proficiency index: A ranking of 100 countries and regions by English skills*. Retrieved from https://www.ef.nl/epi/

Garfield, E. (1995). Quantitative analysis of the scientific literature and its implications for science policymaking in Latin America and the Caribbean. *Bulletin of the Pan American Health Organization (PAHO)*, *29*(1), 87–95.

Gibbs, W. W. (1995). Lost science in the third world. *Scientific American*, *273*(2), 92–99. https://doi.org/10.1038/scientificamerican0895-92

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

Latindex. (2021). Latindex Directory. Retrieved from https://www.latindex.org/

Moed, H. F., & Visser, M. S. (2007). *Developing bibliometric indicators of research performance in computer science: An exploratory study*. Leiden.

Mounier, P. (2018). 'Publication favela' or bibliodiversity? Open access publishing viewed from a European perspective. *Learned Publishing*, *31*(S1), 299–305. https://doi.org/10.1002/leap.1194

Nobre, L. N., & de Freitas, R. R. (2017). A evolução da pós-graduação no Brasil: Histórico, políticas e avaliação. *Brazilian Journal of Production Engineering*, *3*(2), 18–30. Retrieved from https://bit.ly/38sAuMO

Pavan, C., & Barbosa, M. C. (2018). Article processing charge (APC) for publishing open access articles: the Brazilian scenario. *Scientometrics*, *117*(2), 805–823. https://doi.org/10.1007/s11192-018-2896-2

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org

RedALyC. (2021). Sistema de Información Científica RedALyC: Colección de Revistas. Retrieved from https://www.redalyc.org

SBPC, & ABC. (2020b). Qual o papel da CAPES na construção da ciência e da pesquisa no Brasil? [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

SciELO. (2021). SciELO Analytics. Retrieved from https://analytics.scielo.org

Siqueira, M. B. (2019). Sucupira - A Platform for the Evaluation of Graduate Education in Brazil. *Procedia Computer Science*, *146*(2019), 247–255. https://doi.org/10.1016/j.procs.2019.01.081

Tijssen, R. J. W., Mouton, J., van Leeuwen, T. N., & Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation*, *15*(3), 163–174. https://doi.org/10.3152/147154406781775904

van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *arXiv preprint arXiv:2005.10732v2*.

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

9 The unseen costs of article processing charges

> *As long as scholars rely on journal impact measures in order to boost their career perspectives, OA publishers can benefit from higher APCs by improving the standing of the journals they run.*
>
> — **Sergio Copiello**

Some 25 years after the start of the open science movement, a strong development in the transition towards opening practises can be seen in parts of academia, ranging from scholarly publishing access to the first steps taken towards open governance. While many in the science system applaud the movement, it is necessary to follow such developments critically, making sure the intended goals are reached. One of such initial goals of the Open Science (OS) effort was to improve accessibility of publicly funded research to everyone in the science system, creating a global level playing field. Thus, Open Access became one of the most recognisable aspects of OS (Fecher and Friesike, 2014).

This study investigates particular dimensions of Open Access publishing, especially those connected to the payment of Article Processing Charges (APCs) and the less visible consequences of such practises. Research was carried out at the country level, with a special focus on Brazil and the Netherlands, as both played a pioneering role in the development of OS in their own contexts. Brazil clearly was a forerunner in adopting Open Science, as can be seen by the 1998

This chapter has been published as: Brasil, A., Van Leeuwen, T. (2022). The unseen costs of article processing charges: The different realities of Brazil and the Netherlands. In *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*. Granada, Spain. https://doi.org/10.5281/zenodo.6966707

launch of the SciELO database – introducing a cooperative publishing model for OA journals (Packer, 2010) – or by the country's approach to make closed publications democratically accessible with a nationally maintained Portal of Journals (de Almeida et al., 2010). From the Netherlands side, the country already implemented a national Gold Open Access mandate in 2014, followed by national transformative agreements with large publishing houses.

The country selection for this study was also motivated, as will become clear in discussing the results, because Brazil and the Netherlands are in opposite extremes of a spectrum when APC payments and resulting publication impact are taken into consideration. From the initial focus on these two cases, a scenario representative of the Global South and North will become evident, as will the complexity of Open Access practises leading not only to blessings, but also to unintended effects that may force countries to adapt in creative ways.

9.1 Methods and data

This study has been conducted primarily through the analysis of data originating from the Web of Science core collection, considering the in-house version available at the Centre for Science and Technology Studies (Clarivate Analytics, 2022). Data were used to identify journals and articles considering affiliations to assign publications by country. The Digital Object Identifier (DOI) of the publications was used to collect OA information from Unpaywall (Else, 2018). The Directory of Open Access Journals (DOAJ) also provided data on APC. The financial data and the corresponding exchange rates were obtained from the Organisation for Economic Co-operation and Development (OECD), including the Purchasing Power Parity (PPP) (OECD, 2022). For the analyses, only articles published between 2015 and 2018 were considered, allowing the calculation of the Mean Normalised Citation Score (MNCS) based on a four-year window.

For a detailed perspective on Brazil, microdata of the scientific output of the country was collected from the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES, 2021a). The agency's information system offers details on published articles that allow data to be matched with WoS, using a procedure described in a previous study (Brasil, 2021b). The data set was also enriched with information from Latindex, RedALyC, SciELO, and DOAJ.

9.2 Analysing impact and APCs

In the debate around the different types of Open Access, the cost of APCs usually plays a central role (Bosman et al., 2021; Raju et al., 2020; Siriwardhana, 2015). Most scholars recognise that the prices practised by many publishers are not reasonable, and the potential of such practises to increase asymmetry in the science system is often considered a consequence (Siler et al., 2018). Despite that, the movement toward opening science is stronger every day, and those who can afford publishing in high-impact journals, despite costs, often pursue that path to make their findings available to everyone (Jubb et al., 2017). The reality of different countries varies, as do the available resources for publication and the consequential choice of journals (Copiello, 2020). Evidently, this may also cause problems with the impact of research by these countries, and Figure 9.1 shows a known relation between APC costs and impact, in this case represented by the Mean Normalised Citation Score (Waltman, 2016).

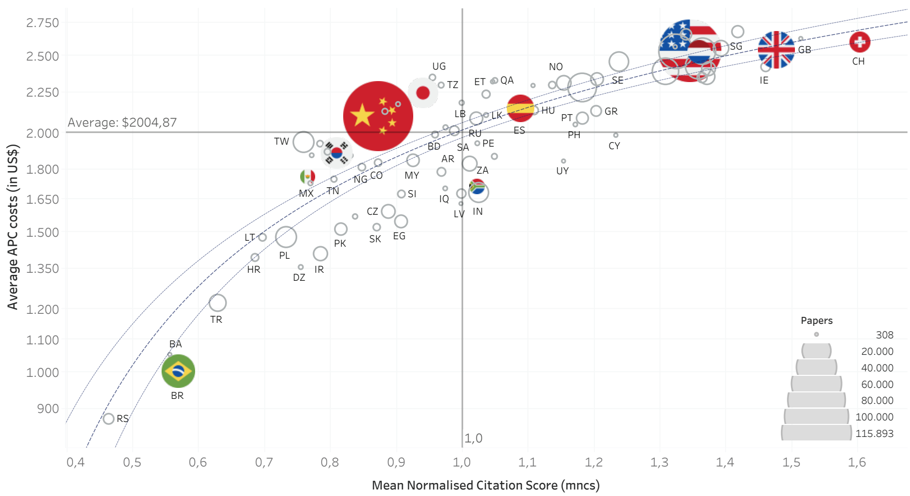


Figure 9.1.: MNCS of OA papers indexed by Web of Science, by country, in relation to average APC costs (2015 – 2018). All figures shown in this chapter are available in interactive format at https://tabsoft.co/3F2mc5y

Figure 9.1 displays countries according to the average APC costs paid for OA publications and the related mean normalised citation score (mncs). The size of each marker represents the number of Gold OA papers found in WoS, with corresponding matches in Unpaywall, as connected by the digital object identifier (DOI). The visualisation suggests that the more a country pays for its open publications, the higher the impact measured by mncs. For instance, it is possible to see a high concentration of countries paying around average value APCs with an impact close to the mncs database average. Some countries take the lead with higher average APC costs and mncs, and there are a few countries with low APCs expenditures and low observed impact.

One could argue that the seeming correlation between impact and high APCs may be a result of the supply-demand mechanism, as high-impact journals would receive many submissions of high-quality papers, being able to raise their charges without reducing demand to a quality-threatening level. However, this study is interested in investigating another economic dimension related to APC costs for countries in distinct social-economic realities. In this sense, Figure 9.1 shows the Netherlands among the leading countries both in average investment and resulting impact, with an mncs 36% above the average of the database for its 12,5 thousand papers. At the other extreme, towards the tail of the logarithmic trend line, Brazil is seen with more than 25 thousand articles that match the adopted criteria, but with an mncs 43% under the reference and an APC value of US$ 1004, half of the average measured for all countries.

If the theory of Purchasing Power Parity (PPP) is considered, one cannot convert currency values to understand how much an APC would cost for any given country. It is necessary to equalise currencies by considering costs of predetermined baskets of products and services in those countries and currencies. This process reveals the US$ 1004 average APC cost for Brazilian OA papers would be equivalent to US$ 2139 in the United States of America. Therefore, Figure 9.2 replaces the average APC of the previous scatterplot with the exchange rate corrected according to the Purchasing Power Parity index available from the Organisation for Economic Co-operation and Development (OECD, 2022).

Although the PPP-corrected expense of APCs for Brazilian publications more than doubles the nominal investment, the situation is quite different for most countries with high mncs averages. For example, the Netherlands spends an average of US$ 2527 on APCs, and the corrected value is only slightly higher, at US$ 2818. Figure 9.2 also shows that most countries with higher impact measures spend proportionally less on APCs, while several countries with around average mncs spend much more if power purchasing parity is considered.

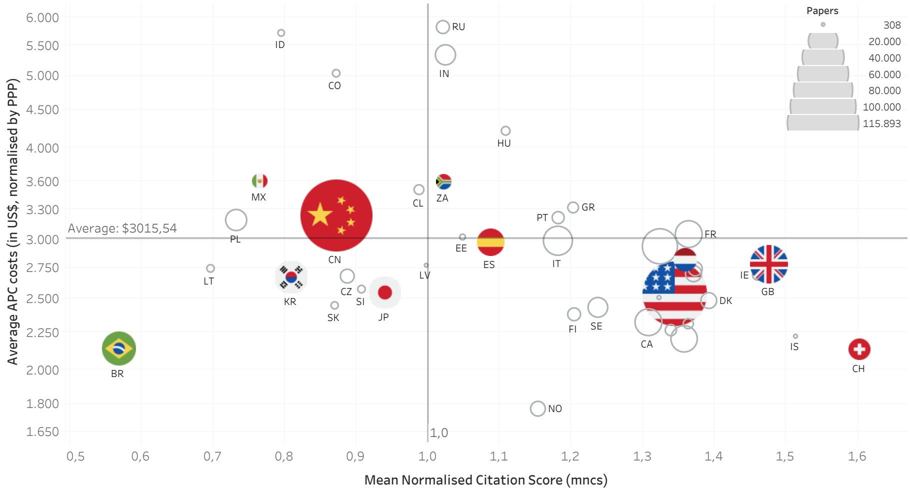


Figure 9.2.: MNCS of OA papers indexed by Web of Science, by country, in relation to average APC costs adjusted by Purchase Power Parity (2015 – 2018).

Although Brazil's investment in golden open-access publishing is small compared to the Netherlands, the reality is that the costs are relatively high if the social-economic situation of the country is considered. This perspective suggests much injustice behind some scholarly communication practises, which intensifies the Matthew effect in science. While everyone is expected to pay exorbitant prices for high-impact publishing, the investment is significantly more expensive for numerous countries, especially those in the Global South. However, this study found a second layer of complexity with respect to publishing practises beyond the cost of APCs and what they represent for different countries.

9.3 The Brazilian practises of OA publishing

A previous study has shown that the Web of Science indexes only around 50% of all Brazilian papers, and most of the country's scientific output finds space in databases such as SciELO, Latindex, and RedALyC (Brasil, 2021b). Conducting a similar analysis of OA publishing in Brazil, we see even more extreme results, since 6429 of the 8010 journals in this category are not indexed by WoS. Most of those are Diamond OA, with no APCs charged for publication (6226 journals, representing 77,7% of the total). Of the 1581 WoS-indexed journals, only 668 are APC-free, representing 8,3% of the total. The distribution of OA journals by indexation in WoS and the price range of APCs can be seen in Figure 9.3.

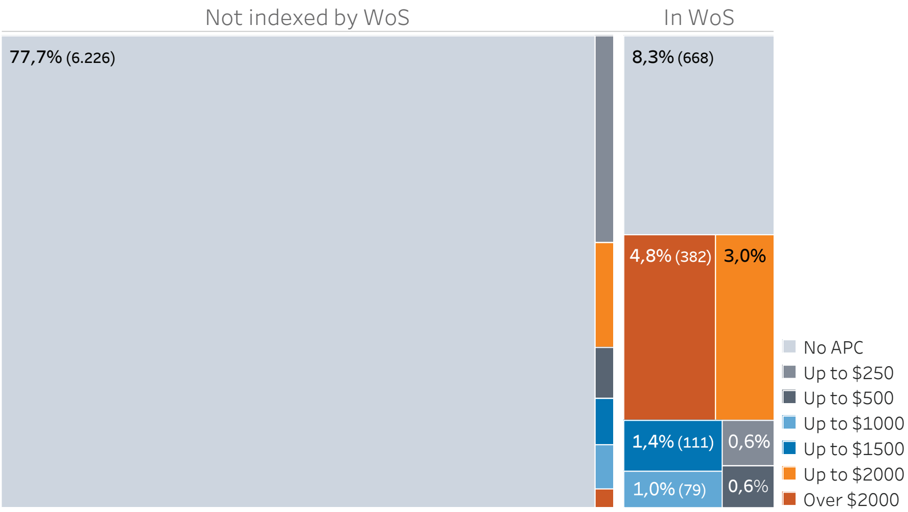


Figure 9.3.: Journals with Brazilian OA output according to APC costs and whether they are indexed by WoS (2015 – 2018).

Figure 9.3 reveals that it is not possible to measure the impact of Brazilian OA publications by only looking at the Web of Science, as the database indexes less than 20% of OA journals used by the country's researchers. Coverage is always an issue to evaluate research performance, especially when economic inequalities are mostly neglected in scholarly publishing practises. For Brazil, that means finding different ways of financing the publication system, so that open publishing alternatives can be provided without costs for authors.

9.4 Publication quality and accessibility

As described in a previous study, indicators are not always sufficient to capture the complexity of some national science systems (Brasil, 2021e). In Brazil, part of the solution to improve evaluation practises came from the implementation, in 1998, of a journal classification system known as Qualis. Such a system combines quantitative evidence from multiple data sources with a qualitative perspective provided by peer review. Therefore, while indicators such as h-index, impact factor, Scopus percentile, and others may be used to calculate provisional classifications, disciplinary peer review committees can challenge these indications and reclassify journals according to what each discipline considers to be of quality. Table 9.1 shows the distribution of the 8010 journals analysed in this study according to the percentile of the Qualis classification.

Table 9.1.: Journals with Brazilian OA output and articles published, according to the APC costs and percentile of the national classification of journals (2015 – 2018).

	Percentile of the Brazilian Classification (Qualis)									
	Top 12,5		12,5–25		25–50		50–75		Bottom 25	
	J	P	J	P	J	P	J	P	J	P
No APC	340	36.689	557	61.191	1221	82.536	1978	84.131	2798	48.521
Up to $250	8	1292	29	16.060	56	16.353	31	3502	17	237
Up to $500	6	2180	16	6260	29	1511	15	88	5	8
Up to $1000	18	2043	18	4276	36	1682	17	68	9	24
Up to $1500	25	1496	38	1801	40	554	17	88	11	33
Up to $2000	102	9010	84	1579	60	972	24	145	15	90
Over $2000	210	11.009	111	2984	44	529	19	88	6	32
Total	709	63.719	853	94.151	1486	104.137	2101	88.110	2861	48.945

Table 9.1 shows the number of journals (J) and articles (P) published from 2015 to 2018, grouped by the percentile ranges of the Qualis classification results and the APC cost categories adopted earlier in the study. The data displayed confirm a general perception in the country's academic community that there is a high volume of lower quality journals within the Diamond Open Access system. This becomes clear from the 4776 APC free journals classified in the bottom 50 percentile of Qualis, accounting for almost 60% of all OA journals used by Brazilian researchers in the period of analysis. Furthermore, more than 130.000 papers have been published in such journals in the period, around 33% of the open access papers in the period.

A whole study could be produced around the journals considered to be of lower quality and about the criteria used by peer review committees to classify them as such. For example, a valuable analysis could depart from the work of Van Noorden (2017), who argues that the usefulness of scientific work is not always captured by citations, and Sugimoto and Larivière (2018), who highlight that most bibliometric databases focus mainly on papers, thus being unable to capture the impact of papers in other types of output, such as policy documents.

However, Table 9.1 also allows for the investigation of the top percentiles of Qualis classification. As can be seen, 340 APC-free journals are ranked among the top 12,5% best, a similar number to the 312 journals with APCs above US$ 1500 at the same level. However, the number of articles published in Diamond OA journals is almost twice that of paid counterparts. Expanding the analysis to the top 25% or 50% journals in the classification, the difference becomes even more significant, as highly priced APCs seem to be paid primarily for those journals considered to have the highest impact.

Diamond OA journals contribute to reduce asymmetry in scholarly communication, as opening research results becomes possible for those unable to pay for exorbitant APCs. But being open is about more than access, and Brazilian Open Access Journals are also a space to publish in Portuguese, the local language. As described by Brasil (2021b), English literacy rates in the country are low even among academics, and publishing in English often restricts the possibilities to produce and consume science. To understand the role of Diamond and low-cost OA journals in this dimension, Figure 9.4 displays the language of publication of the articles listed in Table 9.1.

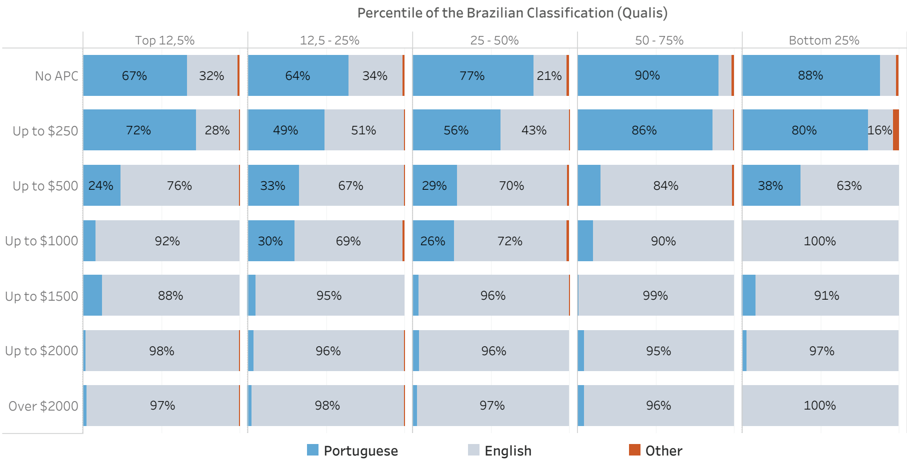


Figure 9.4.: Percentage of Brazilian OA papers, by language, according to APC costs and percentile of the national journal classification (2015 – 2018).

Figure 9.4 shows most Brazilian publications are in Portuguese or English, as the percentage of other languages is negligible. For high-APC journals, English publication dominates, as would be expected for international journals. However, moving towards the top of the chart with low-APC or Diamond OA journals, the number of papers in Portuguese increases significantly. Higher percentages are seen in the lower percentiles of the classification, but are progressively replaced by English in the top percentiles. Considering the nearly 100.000 papers published in Diamond OA by the top 25% of journals, around 65% are in Portuguese, considered to be of as much value as the English publications in expensive journals.

9.5 Conclusion

The analysis of APCs and Gold OA publishing suggests that there may be a significant correlation, at the country-level, in APC expenses and the citation impact of resulting publications. High APC values are then a disadvantage for low- and middle-income countries, and they are even worse than what their nominal value reveals, as exchange rates based on purchase power parity indicators show how much more costly it can be to publish in US dollars. Thus, while OA publishing was intended to further equality, we now witness a higher chance of scholars being excluded from publishing.

As the results of this study show, countries like Brazil are not able to make the same level of investment seen in the Global North, developing alternative publication paths that are not only more coherent with its economic reality, but are also able to value high-quality output even in local language. The problem, as argued by (Brasil and Waltman, 2022), is that metrics from regional publication databases are often invisible in the international context, and unless research analytics become globally inclusive, part of that work remains invisible in evaluation processes and hence is less appreciated.

Although this research has already generated additional results – some available interactively at https://tabsoft.co/3F2mc5y – the format of the short article is a limiting factor. Therefore, an extended version of the study is already underway, for example, to include a historical analysis of the development of Open Science both in Brazil and in the Netherlands, highlighting milestones that helped shape the different approaches of the countries to Open Access. The study will also replicate, for the Dutch system, some of the data exploration performed for Brazil, although recognising that information systems are not as developed or transparent in the Netherlands.

References

Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P.-C., & Proudman, V. (2021). *OA Diamond Journals Study. Part 1: Findings*. Zenodo. Retrieved from https://bit.ly/3vibnrL

Brasil, A. (2021b). Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases. In *Proceedings of the 18th international conference on scientometrics & informetrics* (pp. 193–204). Leuven, Belgium: KU Leuven. Retrieved from https://bit.ly/3jOHQBR

Brasil, A. (2021e). When indicators are not enough: The Brazilian Qualis experience in combining metrics with peer-review. In *Proceedings of the nordic workshop of bibliometrics and research policy*, Odense, DK: University of Southern Denmark.

Brasil, A., & Waltman, L. (2022). Make analytics inclusive: Publication databases still miss too much of the world's research [Research Europe]. Retrieved February 20, 2022, from http://bit.ly/123412

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

Clarivate Analytics. (2022). Web of Science Core Collection. CWTS in-house version. Retrieved from https://bit.ly/3rHmCJT

Copiello, S. (2020). Business as Usual with Article Processing Charges in the Transition towards OA Publishing: A Case Study Based on Elsevier. *Publications*, *8*(1), 3. https://doi.org/10.3390/publications8010003

de Almeida, E. C. E., Guimarães, J. A., & Alves, I. T. G. (2010). Dez anos do Portal de Periódicos da Capes: Histórico, evolução e utilização. *Revista Brasileira de Pós-Graduação*, *7*(13), 218–246. https://doi.org/10.21713/2358-2332.2010.v7.194

Else, H. (2018). How Unpaywall is transforming open science. *Nature*, *560*(7718), 290–291. https://doi.org/10.1038/d41586-018-05968-3

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In S. Bartling & S. Friesike (Eds.), *Opening science* (pp. 17–47). Springer, Cham.

Jubb, M., Plume, A., Oeben, S., Brammer, L., Johnson, R., Bütün, C., & Pinfield, S. (2017). *Monitoring the transition to Open Access*. Universities UK. Retrieved from https://bit.ly/3vqUHPY

OECD. (2022). Purchasing power parities (PPP) (indicator). Retrieved from http://doi.org/10.1787/1290ee5a-en

Packer, A. L. (2010). The SciELO Open Access: A Gold Way from the South. *Canadian Journal of Higher Education*, *39*(3), 111–126. https://doi.org/10.47678/cjhe.v39i3.479

Raju, R., Claassen, J., Pietersen, J., & Abrahamse, D. (2020). An authentic flip subscription model for Africa: library as publisher service. *Library Management*, *41*(6/7), 369–381. https://doi.org/10.1108/lm-03-2020-0054

Siler, K., Haustein, S., Smith, E., Larivière, V., & Alperin, J. P. (2018). Authorial and institutional stratification in open access publishing: the case of global health research. *PeerJ*, *6*, e4269. https://doi.org/10.7717/peerj.4269

Siriwardhana, C. (2015). Promotion and Reporting of Research from Resource-Limited Settings. *Infectious diseases*, *8*, 25–9. https://doi.org/10.4137/idrt.s16195

Sugimoto, C. R., & Larivière, V. (2018). *Measuring Research: What everyone needs to know*. https://doi.org/10.1093/wentk/9780190640118.001.0001

Van Noorden, R. (2017). The science that's never been cited. *Nature*, *552*(7684), 162–164. https://doi.org/10.1038/d41586-017-08404-0

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007

10

A national evaluation push towards increased societal impact

> *It is only by analysing the processes that induce social impact that we have a chance of recognising potential research impacts and the contributions made by research that might otherwise not be evident.*
>
> **— Jack Spaapen & Leonie van Drooge**

The Brazilian National System of Graduate Education (SNPG) constitutes the locus of the majority of scientific and technological research within the country (SBPC and ABC, 2020b). This system, described at length in a previous study (Brasil, 2020), exhibits two primary features pertinent to the current investigation: i) The SNPG emerged as a consequence of consistent public policy implemented over various regimes and governments for many decades, with an applied perspective at its core, emphasising the role of science in driving the country's development; ii) An obligatory evaluation system has been established since the 1970s, impacting the operation, funding, and autonomy of graduate programs (PPG) in the SNPG, which operate only after being accredited by the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES), with reaccreditation required in four-year evaluation cycles.

This chapter has been submitted for review and it is available as a preprint: Brasil, A. (2023). A national evaluation push towards increased societal impact: The Brazilian experience in valuing broader research outputs. SocArXiv. https://doi.org/10.31235/osf.io/s98x3

Despite the SNPG's developmental impetus, by the 1990s, policymakers perceived a growing disconnect between the system and societal needs, deeming it excessively academic. In an attempt to enhance science engagement, professional graduate programs were introduced in the country (Brasil, 2020). These programs were not novel, having been sparsely implemented in countries like Canada, the United States, Australia, and the United Kingdom throughout the 20th century (Maxwell and Kupczyk-Romanczuk, 2009; Taylor, 2008). In Brazil, they started operating in 1998, but only at the master's level (CAPES, 1998). After nearly 20 years of what has been considered a successful experience, the modality was authorised for the doctoral level in 2017 (CAPES, 2017b).

Several studies have documented the successes and failures of professional master's in Brazil (Barata, 2020; Brasil, 2018; Brasil, 2020; Paixão et al., 2014). However, many of the main critics of the modality say that the programs have mostly become "different but equal", a phenomenon Maxwell and Kupczyk-Romanczuk (2009) believe to be an issue for most of the first generation of professional programs around the world. Until the 1980s or 1990s, despite having a different mission, they offered the same types of coursework and thesis model seen in academic programs, and producing equivalent outputs: primarily scientific papers, conference proceedings, and books.

Prior research has revealed distinctions between academic and professional program outputs in Brazil, with the latter showing some increase in technical and technological production (Brasil, 2018). However, this difference remains negligible in several disciplines. Consequently, this study aims to explore recent efforts to foster a more applied output profile for professional programs in Brazil, including the creation of two working groups tasked with devising classification and evaluation systems for technical and technological production (Winter et al., 2019) and to examine the socioeconomic relevance and impact of research (Martinelli et al., 2019). The results of these initiatives could have the potential to steer the Brazilian science system to generate more diverse and impactful research outputs, as new official methods and principles to assess and value a broader universe of products would be introduced.

This investigation seeks to assess the actual and potential ramifications of these initiatives by addressing the following research questions: Have the policies promoting technical and technological production been implemented in graduate programs and if so, how? Were the proposals of the working groups accepted and adapted by the 49 evaluation areas that constitute the SNPG? Can discernible effects of the policies implemented be observed?

To address these questions, a mixture of qualitative and quantitative methodologies was employed. The initial step involved analysing the policy decisions made in Brazil to improve research engagement with societal needs, achieved by examining relevant legislation and policy documents. Subsequently, a literature and policy review was conducted to gauge the success of the country's initiatives in recent decades. By identifying any weaknesses, the study proceeds to scrutinise the reports of the working groups and a series of over 100 documents incorporating their recommendations in relation to disciplinary attributes. Lastly, the investigation culminates in a large-scale analysis of all production outputs from Brazilian PPGs between 2013 and 2020, with the aim of detecting any trends attributable to the proposed policies.

The results of this research will not only improve the understanding of the policy landscape in Brazil, but will also provide valuable insights into the effectiveness of such policies in fostering a more applied production profile for applied research programs. Additionally, the findings will contribute to an ongoing dialogue among researchers, policymakers, and stakeholders in the field of research evaluation, facilitating the identification of best practises and potential areas of improvement.

10.1 On the evaluation of societal impact

In recent decades, universities have become increasingly orientated towards economic and industrial needs. Moed (2006) identified such trends in the university system of most of the member states of the Organisation for Economic Co-operation and Development (OECD). The study also highlighted changes in the funding of Research & Development with fewer resources from the government, partially compensated by an increase in funding from the private sector. Since universities are expected to show increased economic relevance and to make more substantial contributions to national innovation systems, government funding for academic research became more mission-orientated and contract-based. According to Leydesdorff et al. (2017), this represents a shift towards an entrepreneurial university capable of translating academic research into outputs other than traditional scientific research and teaching. The consequence is the existence of growing expectations that universities will become engines of regional, national, and international innovation.

In the Brazilian context, the complex funding models and the overall structure of the National System of Graduate Education (SNPG) inadvertently hindered the promotion of applied research, as argued by Brasil (2020). Consequently, efforts to foster research with increased contributions to innovation systems and increased emphasis on social issues depended on the alternative approach to implement the professional modality of graduate courses in the country (CAPES, 1998). Initially available solely at the master's level, these courses aimed to generate not only high-quality scientific research but also applied knowledge readily accessible to the productive sector. Since the inception of this modality in Brazil, both the academic quality and the professional dimensions of the anticipated outputs have been subject to evaluation by CAPES, in a manner similar to the academic evaluation conducted in the country since the 1970s.

According to Spaapen and Drooge (2011), the assessment of societal impact of research is much more complex than that of scientific impact, as reliable and accepted indicators are widely available only for the latter. Different types of output are expected from more applied research programs, as a variety of potential audiences have different needs and expectations from the research. Furthermore, the lack of databases such as WoS and Scopus is a problem, and there are also time and linkage issues: impact usually takes longer than citations to manifest, and it is harder to link to specific research efforts.

In Brazil, the challenge of assessing societal impact became evident in every evaluation cycle in this century, when even newly created professional graduate programs, which in theory produced a wider variety of outputs, had a large share of their evaluation based on more traditional scientific production (Paixão et al., 2014). This was not a problem in design, as diversity in output types has always been a strongly presented goal, from the first reports and legislation around professional courses (CAPES, 1995a; CAPES, 1995b; CAPES, 1998) to more recent evolutions of the modality (BRASIL. Ministério da Educação, 2009; MEC, 2017; CAPES, 2017c). One of the most comprehensive ordinances on the topic to date stated that the evaluation process would consider intellectual and technical production, including technological publications; patents and other intellectual property assets; software development; technical reports; manuals and protocols; and many others that were flexible to the reality of each discipline (BRASIL. Ministério da Educação, 2009). More recently, the professional modality was extended to the doctoral level through legislation that reinforced the purpose of programs to transfer knowledge to society, according to specific demands orientated towards national or regional development (MEC, 2017).

Unfortunately, actual evaluation practises did not appear to be capable of reflecting the proposed comprehensive assessment. Despite the adequate design, the analysis of the evaluation documents shows that, in practise, there was little distinction between the procedures and indicators applied to assess professional and academic research in the first decade of their division. Paixão et al. (2014) proposed that the biggest flaw in the evaluation was the lack of value given to technical and technological products and the fact that only researchers were involved in the process, which lacked the participation of societal stakeholders. To reach that conclusion, the authors conducted an extensive survey with 229 participants involved in professional PPG, including program directors (GPD), disciplinary representatives appointed to lead evaluations in the national evaluation system, and expert schollars. One of their most relevant conclusions is that conducting evaluations mostly based on strict academic metrics forces graduate programs that are professional in name to become academic in practise.

According to Fink (2006), academic programs are expected to be process driven, motivated by discovery, and targeted at fresh researchers seeking academic training. Knowledge dissemination is a central goal, and bibliometric evaluations often steer publishing practises toward impactful peer-reviewed journals, mainly in English (Hammarfelt and de Rijcke, 2014). For professional courses, at least in the Brazilian setting, the focus is on experienced practitioners seeking additional qualifications and seeking to address societal demands through the scientific method and state-of-the-art knowledge (CAPES, 2017c). With that in mind, one may conclude that evaluating these programs with the same metrics and rules as academic ones limits the potential of outcome-driven research.

10.2 Towards valuing broader research outputs

The national evaluation that occurred in Brazil in 2013 took the first step of proposing specific sets of criteria for the evaluation of professional programs.

At that time, the distinction was still superficial (Paixão et al., 2014). The following evaluation, which took place in 2017 went a step further, as separate assessment committees were nominated for the first time, even allowing limited participation of stakeholders from non-academic backgrounds (CAPES, 2017b). However, even as the quadrennial evaluation occurred, it became clear that the initiative was not yet enough, as the data on nontraditional output from research programs were not suitably structured, and the appropriate assessment criteria were also imperfect.

By the time the quadrennial evaluation was completed, CAPES had already started a debate with Brazilian academics to reflect on the advances of the assessment methods up to that point and plan further improvements for the new cycle (SBPC and ABC, 2020b). Among the pressing issues was the evaluation of societal impact and the assessment of technical and technological production, which were addressed by appointing two working groups to propose alternatives for the ongoing evaluation cycle and the next. Members of both groups were nominated from disciplinary representatives involved in the evaluation system, covering the Humanities and Social Sciences; Exact and Earth and Natural Sciences; and Life Sciences. Evaluation officers from CAPES also integrated the groups (Martinelli et al., 2019; Winter et al., 2019).

The group on technical and technological products focused their work on classification, validation, detail for data collection, and strategies for product valuation. The impact group took a more qualitative approach to identify how different outputs from a research program may impact society. Considering not only scientific, technical, or technological products but also graduates from master's and doctoral programs, the group proposed a template of an impact declaration capable of bringing a qualitative dimension to the listed output of a research program. Taking into account the lack of consensus on the concept of impact in the literature, the study adopted the definition of HEFCE (2015) which defines impact as a measure of how much the output of graduate programs can produce positive effects on society.

Both groups produced very rich and detailed reports that resulted in well-structured recommendations for an improved evaluation. Their efforts were the target of only minor adjustments from discussions between CAPES and disciplinary representatives involved in the national evaluation, leading to action plans to implement the proposals over two evaluation cycles.

Among the recommendations of the working group that led to immediate action was the classification of technical and technological products that could be valued by evaluation committees in different fields. Before the proposal, a total of 62 types of products had been mapped in previous efforts. According to Winter et al. (2019), the classification was too broad, including a series of individual researcher activities that did not relate or represent the activities of the research unit. A new list of 21 types of products was proposed after a survey was conducted with representatives from all fields within the national evaluation system. Table 10.1 shows the product list proposed by the working group, including examples of subtypes suggested by the report or disciplinary documents made available by CAPES (2020d).

Table 10.1.: Types of technical and technological production to be adopted by the Brazilian evaluation system from 2020

	Product type	**Subtype (examples)**
1	Bibliographic product	Article published in technical magazine / Newspaper or magazine article / Art review or critique / Text in exhibition catalogue, playbill, or similar.
2	Intellectual property asset	Patent deposited, granted or licenced / Industrial design / Branding / Integrated circuit topography
3	Social technology	Product, process, method or technique developed as a solution to the demands of segments of society, relevant to the achievement of measurable returns of socio- and / or economic character
4	Professional training course	Training activities developed, organised or carried out at different levels.
5	Publishing product	Edited or organised book, catalogue, collection, encyclopaedia, artistic catalogue, proceedings, etc. Website production.
6	Development of teaching material	
7	Software development	Designing or collaborating on the development of software, systems, apps, games.
8	Event organising	Organising relevant events, seminars, symposiums and others (national or international coverage)
9	Regulatory standard or framework	Developing of regulatory norm, framework, or studies, / Preparation of draught rules or regulatory framework changes / Studies presented at public hearing / Arbitral awards, case studies, case law studies and procedural documents

Continue...

. . . Table 10.1 continued

	Product type	Subtype (examples)
10	Conclusive technical report	Conclusive technical report per se, as well as development of management processes, market research, simulations, scenarios, applied games, business models, managerial tools, etc.
11	Manual or Protocol	Development or adaptation of experimental technological protocol / application (e.g., SOP - Standard Operating Procedure) / Elaboration of technical operation manual.
12	Translation	Published translation or technical review.
13	Collection	Curatorship of exhibitions / Collections produced / Curated biological collections
14	Technical-scientific database	Developed from productive sector demand, which may or may not lead to transfer of information to the partner. Database developed by research unit, with deposit in an open access environment.
15	Cultivar	
16	Communication product	Organising, presenting, participating in radio or television programs, as well as similar media.
17	Chart, map or similar	
18	Protected products or proceedings	Declared impact of technical or technological production / Declared interest of the business sector in production under protection / Development of technology transfer instruments.
19	Taxonomy, ontology and thesaurus	
20	Innovative company or social organisation	Start-up companies originated from research unit. Companies associated/owned by research unit.
21	Non-patentable process, technology or product	

Source: Translated and adapted by the author from Winter et al. (2019) with additional information from disciplinary documents made available by CAPES (2020d).

The 49 disciplinary committees of the national evaluation system were invited to select around ten of the 21 types of technical and technological products shown in Table 10.1 as the most significant in their fields. From the choice made, each committee prepared evaluation guidelines containing instructions for all graduate programs in the country, some even highlighting a selection of different sets of products of relevance for academic and professional programs. All documents were made publicly available on CAPES (2020d), including customised assessment forms with respective indicators and scoring methods. Information on Figure 10.1 reflects the selection of the product for each evaluation area, highlighting the broad areas into which they are classified.

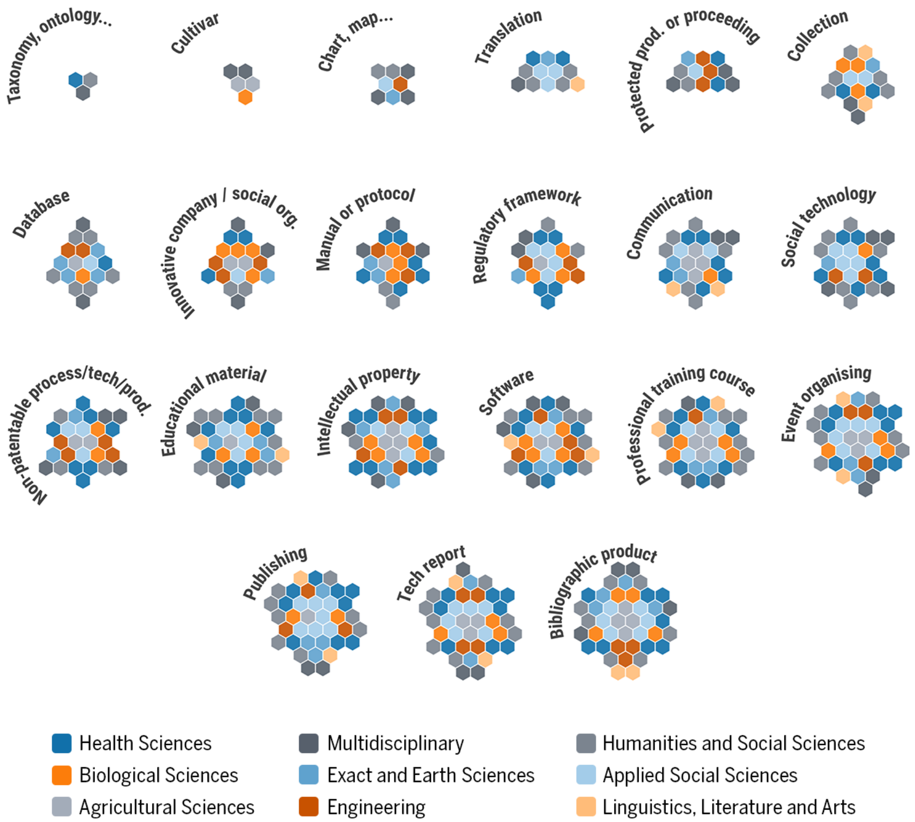


Figure 10.1.: Selection of product types per disciplinary committee, coloured according to CAPES' broad area classification

Figure 10.1, which can be interactively explored at https://andrebrasil.github.io/viz/techprod.html, displays a general overview of the areas that have reported each of the 21 types of products of Table 10.1 as relevant to their evaluation. The overview displayed aligns with expectations, with some products valued in most areas and others restricted to a few. For example, bibliographic and publishing products were valued by most of the committees. On the other hand, cultivars are significant for only five of them (e.g., agricultural sciences and biotechnology). These variations are very representative of the framework Brazilian evaluation is trying to achieve; one in which disciplines are free to determine what matters most for them, so it can recognised and rewarded, thus stimulating growth where desireable.

Further exploration of the interactive figure associated with Figure 10.1 reveals a varying range of selected products in different evaluation areas, with the majority opting to stay around the recommended number and considering something between 8 and 12 products. A notable exception is Astronomy/-Physics, which chose only four products as relevant to their evaluation. In contrast, 11 evaluation areas selected over 12 products. Among them is the expected Interdisciplinary, for which all products were deemed relevant. Political Science and International Relations was also very broad, encompassing 19 products (only cultivar and intellectual property assets were left out).

Regarding the evaluation methods adopted by the areas, it is interesting to see that 19 of them stated that they would employ only qualitative methods in their analysis. This means that they are not concerned with the amount of products listed as outputs from each of the graduate programs under evaluation. They would focus their efforts to qualitatively assess a portfolio of products selected by the PPG as representative of their work as a whole. The remaining 30 evaluation areas informed that they would adopt a mixed methods approach, also performing the portfolio assessment together with a quantitative analysis of the products from each program. Interestingly enough, no area chose a purely quantitative approach to their evaluation, meaning that all of them could value the academic quality and practical applicability of the research outputs.

10.3 Observable changes in the system

The 2018 initiative to rethink Brazil's national evaluation system began immediately after the publication of the 2017 quadrennial evaluation results, which analysed data from graduate programs spanning 2013-2016. The subsequent evaluation would encompass the 2017-2020 period, suggesting that any reform efforts would occur late in the four-year cycle. However, considering the high stakes nature of the Brazilian evaluation discussed by Brasil (2022), higher education institutions and their graduate programs closely monitored the debate to anticipate any changes in the upcoming evaluation. In the case of the assessment of technical and technological production, even though the report of the working group and the subsequent design of criteria within the evaluation areas would materialise only in 2019, the mere appointment of a working group showed that the issue would gain relevance in the evaluations to come. Thus, while substantial effects of valueying alternative types of scholarly output may only be seen in the 2021-2024 cycle and beyond, the proactive engagement of the SNPG already generates anticipation for some increase in the proportion of technical and technological products within the 2022 evaluation.

To verify whether an increase in technical and technological production has already manifested in the 2017-2020 period, all data sets on graduate program output available on the CAPES open data platform were collected (CAPES, 2021a), and the results of the analysis can be seen in Figure 10.2, which can be explored in detail using the interactive dashboard available at http://andrebrasil.github.io/techprod.html

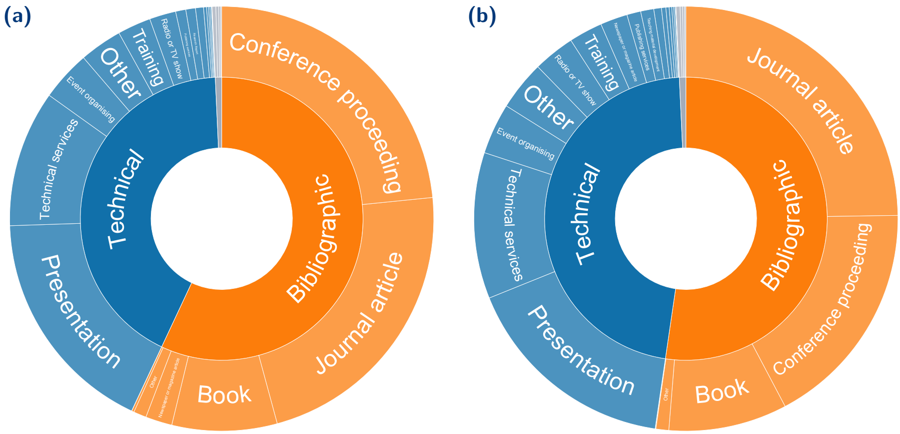


Figure 10.2.: Bibliographic, technical, and artistic production of graduate programs in Brazil, in: (a) 2013-2016; and (b) 2017-2020.

Figure 10.2a shows that nearly 57% of the graduate program output in the 2013-2016 period consisted of bibliographic products, especially books, journal articles, and conference proceedings. Technical and technological products accounted for around 42% of the total output, and artistic production appears as a thin grey slice at the top of the figure. The expected profile change started to materialise in 2017-2020, and Figure 10.2b shows that the share of bibliographic products decreased to around 52% of the total output in the period, with a corresponding increase in technical and technological production. The sunburst charts also reveal some changes in proportion for the subtypes of products, for instance with journal articles and conference proceedings exchanging places. Figure 10.3 shows the percentage variations of the different subtypes available in the database for bibliographic and technical & technological products.

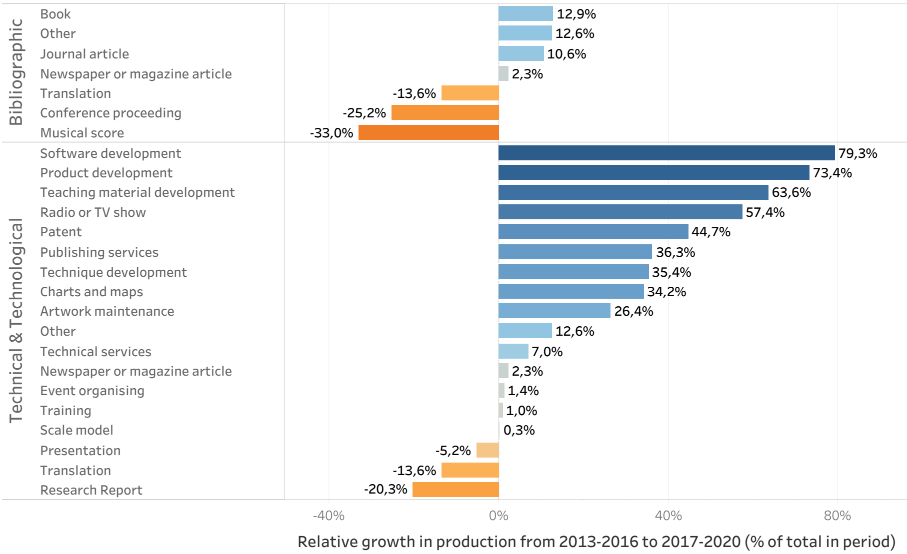


Figure 10.3.: Percent variation in the share of product subtypes in relation to total graduate program output: from 2013-2016 to 2017-2019

As the comparison shown in Figure 10.3 considers two evaluation cycles, the possible influence of the COVID-19 pandemic is not extreme in the most recent production profile of graduate programs, but can be observed in the slight reduction in the percentage of presentations and the more significant drop in conference proceedings. Relevant increases in the representation of products such as patents and the development of software, products, and techniques suggest that an evaluation process that values a broader range of scientific outputs may encourage greater diversity in these outputs. However, careful scrutiny reveals a challenge in accurately analysing the current status of product subtypes: the production database has not been updated to reflect the product classification adopted for the 2017-2020 evaluation, as detailed in Table 10.1.

Changes leading to an increased valuation of technical and technological products, as well as to a new selection and classification of desired products, occurred toward the end of a four-year evaluation cycle. Therefore, the complexity of changing the data collection system to reflect the new situation would be high, especially considering that graduate programs report their production on an annual basis. The impossibility of updating the system led to challenges for data analysis of the products and also made the quantitative assessment proposed by the 30 evaluation areas adopting mixed methods weaker. If products are not recorded as expected, with the appropriate metadata for each product type, the evaluation committees cannot count on the information for their analyses. They either do a parallel data collection, not recorded in official evaluation databases, or their assessment changes to a mostly qualitative one.

Limitations in the data may restrict further advances in the analysis of subtypes, but it is possible to investigate changes in the overall production in each evaluation area. For this, Figure 10.4 shows the percentage of artistic, technical and technological production per area, with the academic and professional modalities of PPG separated to identify possible differences in the production profile. For each case, data can be seen for 2013-2016 and 2017-2020.

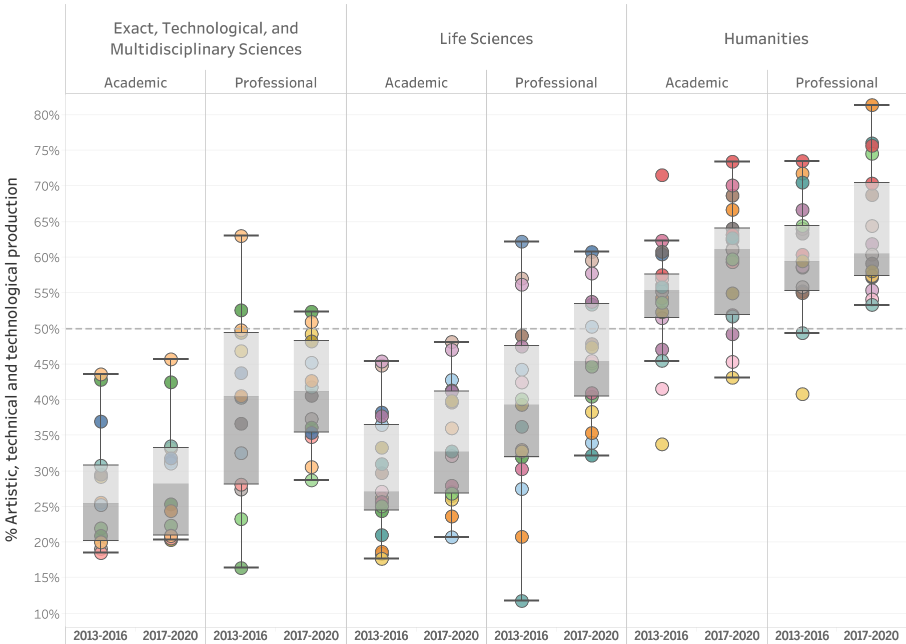


Figure 10.4.: Percent of artistic, technical and technological production per evaluation area from 2013-2016 to 2017-2020, grouped by PPG modality

Each of the circles shown in Figure 10.4 represents the percentage of artistic, technical and technological production in an evaluation area. This percentage can be as low as 12%, considering production from professional graduate programs in the area of Biological Sciences III, and as high as 81%, in the PPG of the same modality in Religion and Theology. The dashboard available at https://andrebrasil.github.io/viz/techprod.html allows each circle to be identified and explored in more detail. However, the collection of 49 evaluation areas reveals a series of interesting production behaviours within their broad area groups. For instance, there is a striking difference in behaviour across broad areas, with a very high percentage of artistic, technical and technological production in the Humanities. Even in more academic programs, the share of bibliographic production is below 50% in most areas.

Productivity variations between academic and professional programs are also evident, with the first more focused in the bibliographic production, more inward orientated as it is mostly destined to academic audiences. Professional programs would, as expected, be more outward orientated, with a higher percentage of artistic, technical and technological production, mostly destined to non-academic audiences. Whether or not that profile difference is under or above the desired, and if a more balanced behaviour should be an aim across the distinct evaluation areas, are questions for future research.

Finally, what is evident in most groups is the growth in outward proportion from 2013-2016 to 2017-2020. As expected, the variation is a bit more evident in the academic PPG, as the professional ones already had this sort of production a little more established in their primary objectives. However, even if little variation can be seen in the median of professional programs in “Exact...” and “Humanities”, the overall variation in the area distribution is quite significant.

10.4 Looking for impact

Despite the late adoption of the new approach that values technical and technological production in the Brazilian evaluation system, the National System of Graduate Education (SNPG) has already shown promising results. Although it is difficult to attribute the increase in outward production from graduate programs directly to the evaluation push, the numbers are growing significantly and may become even more impressive by the end of the 2021-2024 cycle. However, the future of impact evaluation, the focus of the second working group approached in this study, remains uncertain.

It is nearly impossible to find consensus on a single definition of the impact of research on society. That was one of the preliminary conclusions of Martinelli et al. (2019), who recognised impact evaluation as a more recent concern for researchers, and also a new focus in Brazil after the introduction of citation metrics in the graduate evaluation system. To serve as a guiding concept for Brazilian evaluation, the definition of impact proposed by the Research Excellence Framework (REF) was adopted, declaring impact to be a measure of the positive effects generated by graduate program outputs for a community (HEFCE, 2015).

According to Martinelli et al. (2019) graduate programs traditionally aimed to generate knowledge to address future problems, rather than immediate demands solvable through existing knowledge. Thus, scientific impact takes precedence, as the generation of new knowledge forms the basis for academic or business groups to develop future solutions to humanity's challenges. That has been the focus of Brazilian evaluation for decades, but researchers are increasingly expected to provide solutions to everyday problems facing the population. Since there is no tradition and little practise in assessing the economic and societal impact of projects and products, many researchers assume their role is to provide foundational knowledge and technology for public or private organisations to implement practical solutions, rather than developing directly applicable products and activities.

With this in mind, the Impact Working Group proposed that the first efforts to map impact in Brazilian evaluation should start with the assessment of the technical and technological products catalogued by Winter et al. (2019). This type of output helps translate scientific impact into potential impact, meaning researchers would show stakeholders how the research conducted could be applied to yield significant results and generate societal impact. The primary indicators the group suggested for application in impact evaluation include causality, impact type, anticipated duration, beneficiary sector, link to institutional strategic planning, application type, availability, beneficiary sector, and territorial scope. This comprehensive approach aims to accurately assess the various aspects of the impact of research within the Brazilian evaluation system.

Three relevant points derive from the indicator list. The first is that the working group proposed the indicators as the basis for an assessment form to be added to the information system cataloguing technical and technological production. As the definition of new product types came too late in the evaluation cycle to be implemented in time, the same happened for the proposed form. Thus, any application of indicators may have been used to inspire qualitative evaluation within areas, but no consistent documentation of that has been found in published evaluation reports (CAPES, 2020d). The next two points relate to the future of impact evaluation, as seen by the WG: time and autonomy.

Scientific impact can be converted to potential impact by the effort of researchers in translating to stakeholders. However, real impact occurs only after the target audience uses research products or services effectively. Accordingly, impact evaluation must consider a temporal scale (short-, medium-, and long-term), since some research benefits may only materialise after many years. With this perspective in mind, Martinelli et al. (2019) recommends a cumulative consideration of up to 12 years of graduate programme output to capture the real effect of research output.

Finnaly, with respect to autonomy, one of the proposed indicators looks at the relation of the research product with institutional strategic planning. For one, that intends to separate intentional from accidental impact, and an eventual distinction of value among these is a matter of complex discussion in itself. However, that also relates to the working group's ideas on the future of impact evaluation, one that relies on a self-assessment strategy where institutions can evaluate their impact based on their mission, objetives, actions, and output.

10.5 Conclusions

This study has presented an analysis of initiatives to further the impact of Brazil's National System of Graduate Education (SNPG). Covering from the historical design of the system and the implementation of a professional modality of graduate program, to recent efforts of an evaluation push to promote growth in technical and technological production, many practical steps have been taken towards the desired impact, several only possible due to the normative aspect of the SNPG, as well as the evaluation system in place.

One of the strengths of the SNPG in promoting any type of change is its high-stakes evaluation. It has direct effects on funding, determines accreditation and the continued existence of graduate programs, the base research unit in the country. Evaluation is compulsory, top-down, centralised, and comparative; thus, any new priority for evaluation can be implemented with a potential steering effect rarely seen elsewhere. Since one of the important recommendations after the 2017 national evaluation was that more attention should be paid to the impact of graduate programs, two working groups have embodied the efforts in that direction. At the end of 2019, they made recommendations to improve the classification and evaluation of technical and technological products and to develop a qualitative approach to assess the impact of these various research program outputs.

As the most recent national evaluation in Brazil assessed the performance of graduate programs from 2017 to 2020, the results of the working groups were a bit late to be fully implemented. However, the first steps taken have signalled academia that a more holistic approach to evaluation has arrived and that it values not only the more traditional bibliographic production but also technical and technological outputs and their relation with impact. Since the 49 evaluation areas in the Brazilian system have incorporated the new perspective in their guidance documents, the process has already started to foster a shift in the orientation of graduate programs, represented by a significant change in their publishing profiles to include a higher proportion of technical and technological products.

It is evident that a surge in the reporting of specific types of products does not necessarily indicate a corresponding increase in production. During the past two decades, the emphasis on bibliographic production within the evaluation system implemented by CAPES has left indelible marks, resulting in pervasive mindsets among faculty, students, and staff that insufficiently document other activities. These deeply embedded perspectives are now undergoing a reversal. For instance, the number of reported patents experienced a relative growth of approximately 45% between the 2013-2016 and 2017-2020 cycles. Although this increase in reporting may be attributable to the 2019 evaluation push, that does not imply that additional patents were produced as a direct consequence. It is more plausible that patents were among the numerous product categories underreported in the annual data collection from graduate programs, given their less pronounced impact in evaluations compared to the new model. However, as researchers become aware that their work will receive due recognition, reporting is expected to improve, consequently inspiring the production of more diverse research outputs.

In conclusion, the shift toward a novel evaluation model presents both challenges and opportunities for Brazilian graduate programs. Although the path to a more impact-orientated evaluation methodology is laden with complexities and uncertainties, it is imperative to recognise the constraints imposed by the previous focus on bibliographic production and to adopt a holistic approach that takes into account the whole of the programs' performance and research products. The potential benefits of implementing such a system are considerable, encompassing enhanced engagement with non-academic audiences and increased emphasis on the societal and economic advantages of research. By persistently refining and advancing the evaluation system and by cultivating a collaborative and transparent environment within the academic community, Brazil's national evaluation has the capacity to exert a pivotal influence on the future trajectory of graduate education and research, thereby ensuring that the country's science system may impact society in broader ways.

References

Barata, R. B. (2020). Programas de pós-graduação profissionais: Por que precisamos deles? *International Journal of Business & Marketing*, *5*(2), 30–34.

Brasil, A. (2018). Exploring differences in scientific production originated from professional and academic graduate courses in Brazil. In *Proceedings of the 23rd international conference on science and technology indicators* (pp. 27–40). https://doi.org/10.6084/m9.figshare.9443693.v1

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2022). Multidimensionality through self-evaluation: from theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation*, (53). https://doi.org/10.22163/fteval.2022.546

BRASIL. Ministério da Educação. (2009). *Portaria Normativa nº 17, de 28 de dezembro de 2009. Dispõe sobre o mestrado profissional no âmbito da CAPES (Revogada pela Portaria nº 389/2017)* (Ordinance No. 17/2009). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3H4FEyg

BRASIL. Ministério da Educação. (2017). *Portaria nº 389, de 23 de março de 2017. Dispõe sobre o mestrado e doutorado profissional no âmbito da pós-graduação stricto sensu*. Diário Oficial da União. Brasília, DF.

CAPES. (1995a). *Portaria nº 47, de 17 de outubro de 1995. Determina a implantação, na CAPES, de procedimentos apropriados à recomendação, acompanhamento e avaliação de cursos de mestrado dirigidos à formação profissional* (Ordinance No. 47/1995). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/33CLf0V

CAPES. (1995b). *Programa de flexibilização do modelo de pós-graduação senso estrito em nível de mestrado*. Brasília, DF. Retrieved from https://bit.ly/3sp5Aic

CAPES. (1998). *Portaria nº 80, de 16 de dezembro de 1998. Dispõe sobre o reconhecimento dos mestrados profissionais e dá outras providências*. Diário Oficial da União. Brasília, DF.

CAPES. (2017b). *Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal*. Diário Oficial da União. Brasília, DF.

CAPES. (2017c). *Portaria nº 60, de 20 de março de 2019. Dispõe sobre o mestrado e doutorado profissionais* (Ordinance No. 60/2019). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3BGiXzo

CAPES. (2020d). Sobre as áreas de avaliação. Retrieved April 9, 2021, from https://bit.ly/3s0XB9f

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

Fink, D. (2006). The professional doctorate: Its relativity to the Ph.D. and relevance for the knowledge economy. *International Journal of Doctoral Studies*, *1*(1), 35–44. https://doi.org/10.28945/59

Hammarfelt, B., & de Rijcke, S. (2014). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, *24*(1), 63–77. https://doi.org/10.1093/reseval/rvu029

HEFCE. (2015). *Annual report and accounts 2014-15*. Higher Education Funding Council for England. Bristol. Retrieved from https://bit.ly/3c9sfbC

Leydesdorff, L., Etzkowitz, H., Ivanova, I., & Meyer, M. S. (2017). The Measurement of Synergy in Innovation Systems: Redundancy Generation in a Triple Helix of University-Industry-Government Relations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2937647

Martinelli, A. E., Lima, A. C. S., Costa, F. N., Fernandes, L. M. R., Brito, M. V. H., Possamai, O., . . . Avellar, S. O. d. C. (2019). *Impacto e relevância econômica e social*. Retrieved from https://bit.ly/3vTjxFO

Maxwell, T. W., & Kupczyk-Romanczuk, G. (2009). Producing the professional doctorate: the portfolio as a legitimate alternative to the dissertation. *Innovations in Education and Teaching International*, *46*(2), 135–145. https://doi.org/10.1080/14703290902843760

Moed, H. F. (2006). *Citation analysis in research evaluation*. Dordrecht: Springer.

Paixão, R. B., Bruni, A. L., Becker, J. L., & Tenório, R. M. (2014). Avaliação de mestrados profissionais: construção e análise de indicadores à luz da multidimensionalidade. *Ensaio: Avaliação e Políticas Públicas em Educação*, *22*(83), 505–532. https://doi.org/10.1590/s0104-40362014000200010

SBPC, & ABC. (2020b). Qual o papel da CAPES na construção da ciência e da pesquisa no Brasil? [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

Spaapen, J., & Drooge, L. v. (2011). Introducing 'productive interactions' in social impact assessment. *Research Evaluation*, *20*(3), 211–218. https://doi.org/10.3152/095820211x12941371876742

Taylor, J. (2008). Quality and standards: The challenge of the professional doctorate. *Higher Education in Europe*, *33*(1), 65–87. https://doi.org/10.1080/03797720802228191

Winter, E., Brasil, A., Moraes, C. L. d., Granjeiro, J. M., Filho, J. R. d. S. V., Nichiata, L. Y. I., . . . Pascutti, P. G. (2019). *Produção técnica*. Brasília. Retrieved from https://bit.ly/3t06SiJ

Part IV

Towards a multidimensional model

11 Multidimensionality through self-evaluation

> *One should not present a series of separate rankings in parallel but rather a dataset and tools to observe patterns in multi-faceted data.*
>
> — **Henk Moed**

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 & technology research is conducted within graduate programs, both at the master's and doctoral levels (Balbachevsky, 2005; Brasil, 2020; CFE, 1965; C. B. 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

This chapter has been published as: Brasil, A. (2022). Multidimensionality through self-evaluation: From theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation* (53), pp. 97–106. https://doi.org/10.22163/fteval.2022.546

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, 2018c).

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, 2019a; 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 chapter 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 chapter 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.

11.1 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).

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

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]

11.1.1 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 generate 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

[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 multidimentional 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).

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 to distinct stakeholder groups. For some, reports about a particular field may be desirable (e.g., potential students); for others, the institutional-level ranking might be more relevant (e.g., HEI managers). Thus, UMR organises its data and indicators to allow 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 collections 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 11.1 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 indicators[3] for the University of São Paulo (USP), one of the largest HEI in Latin America.

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**	

Figure 11.1.: U-Multirank research dimension for the University of São Paulo (USP)

The "research" dimension displayed on Figure 11.1 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.

[3] The U-Multirank website refers to indicators as "measures".

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 11.2, once again with data from USP, according to the ranking's 2021 edition.

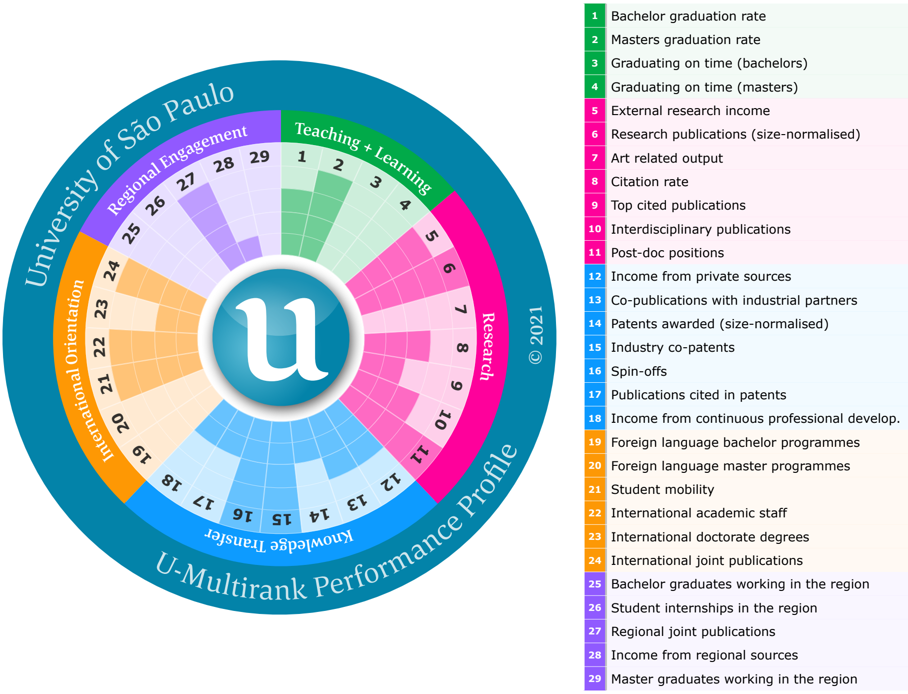


Figure 11.2.: U-Multirank university performance chart for the University of São Paulo (USP), also listing the 29 indicators displayed in the sunburst

The sunburst seen on Figure 11.2 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 11.2 with Figure 11.1, 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).

11.1.2 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 11.2.

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 11.3 displays the charts for the top three universities.

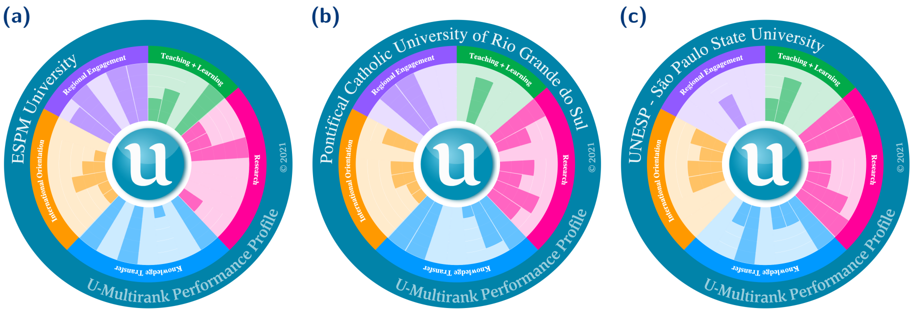


Figure 11.3.: U-Multirank performance charts for the three top scoring Brazilian HEI offering graduate programs. Interactive version at https://bit.ly/3fRH30m

The best-ranked institutions are shown from Figures 11.3a to 11.3c, 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 doctorates 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, 2021e; INEP, 2020).

As a direct comparison, the institution shown in Figure 11.3c – São Paulo State University (UNESP) – includes 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 "excellent", 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, 2021e).

This very superficial comparison tells one main story: ESPM and UNESP 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.

11.1.3 The proposed model

The concept of a multidimensional assessment for the SNPG matured over the course of more than two years, also building on 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. As a complement, 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.

11.2 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.

11.2.1 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 dimensions: 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 evaluation in 2017 consisted of two levels: 18 items organised into five dimensions. Each dimension is evaluated 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 this, Figure 11.4 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).

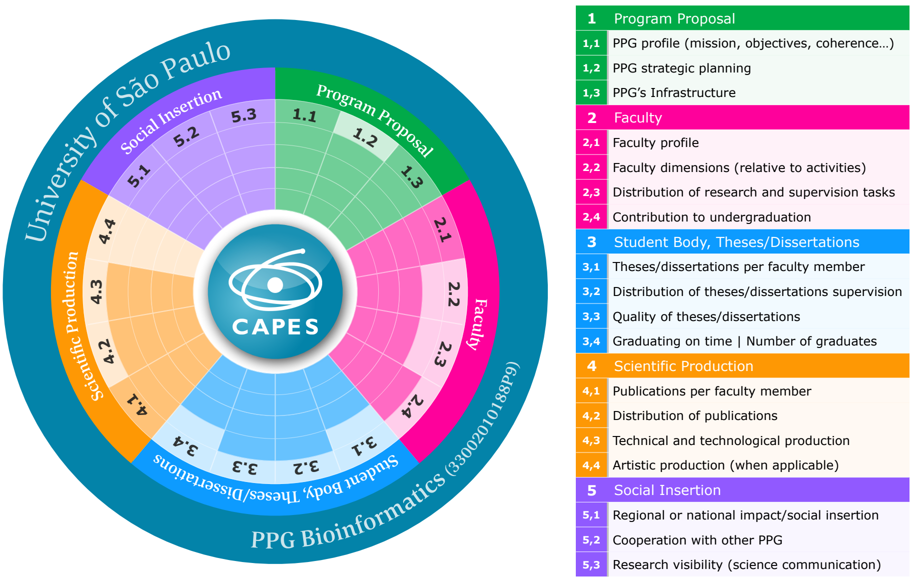


Figure 11.4.: Results from CAPES’ 2017 evaluation of the PPG on Bioinformatics (USP), transposed to U-Multirank’s sunburst

The Bioinformatics program shown in Figure 11.4 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 11.4, the PPG had seven items 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 aplicable (4.4, regarding artistic production). The sunburst shows the distribution of strenghts 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 dimensions, 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 dimensions 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 dimension 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 assessment form applied for the 2017 evaluation, especially since it has already been revised by Monteiro et al. (2019) for the quadrennial evaluation planned for 2021. The goal here is to show that the evaluation process in Brazil has already been multidimensional, although this is mostly lost when the evaluation is reduced to a single grade.

11.2.2 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 items) 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 multi-faceted 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 dimensions in the final grade was unbalanced: usually of 10% for the first and 35% for the later (CAPES, 2017b). 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.

11.3 Self-evaluation

As stated early in the chapter, 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 findings on self-evaluation, proposing strategies for its adoption in the assessment of PPG. The central concept was that each graduate program would implement a custom-made 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 (D. 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 items to the new assessment form proposed by Monteiro et al. (2019). Each item – "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?

11.4 Conclusion

D. 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 existing dimensions and items 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 items; 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 complement each other and must be articulated in such a way as to heighten their potential individual effects.

References

Audy, J. L. N. (2020). O novo modelo de avaliação da CAPES [webinar]. YouTube: Unesp Oficial. Retrieved August 13, 2021, from https://youtu.be/TyNrLIilnHU

Balbachevsky, 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). Rio de Janeiro: 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. https://doi.org/10.33871/nupem.2020.12.27.222-253

BRASIL. Ministério da Educação. (2021). Cadastro e-MEC de Instituições e Cursos de Educação Superior. Brasília, DF: MEC. Retrieved from 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. https://doi.org/10.3152/095820208x280907

CAPES. (2010). *Plano Nacional de Pós-Graduação - PNPG 2011-2020*. Brasília, DF: CAPES.

CAPES. (2017b). *Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal*. Diário Oficial da União. Brasília, DF.

CAPES. (2018c). *Avaliação Quadrienal em números*. Brasília, DF. Retrieved from https://bit.ly/2Y2rPvO

CAPES. (2019a). *Aprimoramento do processo de avaliação da pós-graduação*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3wDRgDU

CAPES. (2021b). Avaliação Quadrienal. Retrieved June 18, 2021, from https://bit.ly/35wv16S

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. (2018). *Proposta de aprimoramento do modelo de avaliação da PG*. CAPES. Brasília, DF. Retrieved from 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*. CAPES. Brasília, DF. Retrieved from 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*. Brasília, DF.

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.

Fanelli, A. G. D. (2016). U-Multirank and Latin American Universities. *International Higher Education*, (84), 7–9. https://doi.org/10.6017/ihe.2016.84.9108

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). https://doi.org/10.1007/978-94-007-3005-2_10

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). https://doi.org/10.1007/978-94-007-3005-2_6

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

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. https://doi.org/10.29024/sar.31

Hazelkorn, E., & Gibson, A. (2017). Global science, national research, and the question of university rankings. *Palgrave Communications*, *3*(1), 21. https://doi.org/10.1057/s41599-017-0011-6

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. https://doi.org/10.1590/s1414-4077/s1414-40772020000200006

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. https://doi.org/10.22347/2175-2753v12i37.2825

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. https://doi.org/10.20336/rbs.256

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, *110*(2), 967–990. https://doi.org/10.1007/s11192-016-2212-y

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*. Brasília, DF: CAPES. Retrieved from https://bit.ly/2PmAf0q

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

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. https://doi.org/10.1007/s11192-005-0008-6

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). https://doi.org/10.1007/978-94-007-3005-2_1

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

Verhine, R. E., Lima, A. C. S., Carvalho, D. B. B. d., Leite, D., Bertolin, J. C. G., Dantas, L. V., . . . Rosa, D. L. (2019). *Autoavaliação de Programas de Pós-Graduação*. Brasília, DF: CAPES. Retrieved from https://bit.ly/3jTSRS8

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

Waaijer, C. J. (2018). Why the public nature of university rankings torpedoes their value for benchmarking: an institutional researcher's opinion. In *Proceedings of the 23rd international conference on science and technology indicators* (pp. 1031–1037). Leiden, Netherlands.

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 considered 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

Reference List

AAAS. (2022). The Scope of the “Humanities” for purposes of the Humanities Indicators. Retrieved April 20, 2022, from https://bit.ly/36uQCAH

ABE. (1929). *O problema universitário brasileiro: Inquérito promovido pela Secção de Ensino Technico e Superior da Associação Brasileira de Educação*. Rio de Janeiro: Encadernadora.

ABRASCO. (2019). Apreciação da proposta de Qualis Periódico Referência. Retrieved March 2023, from https://bit.ly/3mSLm2a

Agostini, R. (2019). MEC cortará verba de universidade por ’balbúrdia’ e já enquadra UnB, UFF e UFBA. *Estadão*. Retrieved from https://bit.ly/3pYA0KW

Alonso-Gamboa, J. O., & Russell, J. M. (2012). Latin American scholarly journal databases: a look back to the way forward. *Aslib Proceedings*, *64*(1), 32–45. https://doi.org/10.1108/00012531211196693

Alperin, J. P. (2014). Citation databases omit local journals. *Nature*, *511*(7508), 155. https://doi.org/10.1038/511155c

Amado, A. M., Tomanari, G. A. Y., Fernandes, L. M. R., Dalmonte, E. F., Magalhães, J. S. d., Siqueira, V. B. C., . . . Oliveira, T. M. (2020). *Relatório do GT do Qualis Periódicos Humanidades*. CAPES. Brasília, DF. Retrieved April 12, 2023, from https://bit.ly/3KSUzAF

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

Andrews, L. (2017). How can we demonstrate the public value of evidence-based policy making when government ministers declare that the people ‘have had enough of experts’? *Palgrave Communications*, *3*(1), 11. https://doi.org/10.1057/s41599-017-0013-4

Audy, J. L. N. (2020). O novo modelo de avaliação da CAPES [webinar]. YouTube: Unesp Oficial. Retrieved August 13, 2021, from https://youtu.be/TyNrLIilnHU

Balbachevsky, E. (2004). Graduate education: emerging challenges to a successful policy. In C. Brock & S. Schwartzman (Eds.), *The challenges of education in Brazil* (pp. 209–228). Oxford, UK: Triangle Journals.

Balbachevsky, 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). Rio de Janeiro: Nova Fronteira.

Balbachevsky, E., & Schwartzman, S. (2010). The graduate foundations of research in Brazil. *Higher Education Forum*, *7*(1), 85–101.

Bali, A. S., Capano, G., & Ramesh, M. (2018). Anticipating and designing for policy effectiveness. *Policy and Society*, *38*(1), 1–13. https://doi.org/10.1080/14494035.2019.1579502

Barata, R. B. (2016). Dez coisas que você deveria saber sobre o Qualis. *Revista Brasileira de Pós-Graduação*, *13*(30), 13–40. https://doi.org/10.21713/2358-2332.0.947

Barata, R. B. (2019). Necessary changes in the evaluation of graduate programs in Brazil. *Interface - Comunicação, Saúde, Educação*, *23*. https://doi.org/10.1590/interface.180635

Barata, R. B. (2020). Programas de pós-graduação profissionais: Por que precisamos deles? *International Journal of Business & Marketing*, *5*(2), 30–34.

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

Barroso, J. (2005). O Estado, a educação e a regulação das políticas públicas. *Educação & Sociedade*, *26*(92), 725–751. https://doi.org/10.1590/s0101-73302005000300002

Bawa, A. (2008). South Africa. In M. Nerad & M. Heggelund (Eds.), *Toward a global phd? forces and forms in doctoral education worldwide* (pp. 117–130). Seattle, WA: University of Washington Press.

Becher, T. (1993). Graduate Education in Britain: The View from the Ground. In B. R. Clark (Ed.), *The research foundations of graduate education: Germany, britain, france, united states, japan* (pp. 115–153). Berkeley, CA: University of California Press.

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

Ben-David, J., & Zloczower, A. (2009). Universities and academic systems in modern societies. *European Journal of Sociology*, *3*(1), 45–84. https://doi.org/10.1017/s0003975600000527

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation*, *23*(2), 166–173. https://doi.org/10.1093/reseval/rvu002

Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P.-C., & Proudman, V. (2021). *OA Diamond Journals Study. Part 1: Findings*. Zenodo. Retrieved from https://bit.ly/3vibnrL

Brasil, A. (2018). Exploring differences in scientific production originated from professional and academic graduate courses in Brazil. In *Proceedings of the 23rd international conference on science and technology indicators* (pp. 27–40). https://doi.org/10.6084/m9.figshare.9443693.v1

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2021a). A national evaluation push towards increased societal impact: The Brazilian experience in valuing broader research outputs. In *Proceedings of the eu-spri conference 2021*. Extended abstract, Oslo, Norway: Eu-SPRI. Retrieved from https://www.euspri2021.no/abstracts/

Brasil, A. (2021b). Beyond the Web of Science: an overview of Brazilian papers indexed by regionally relevant databases. In *Proceedings of the 18th international conference on scientometrics & informetrics* (pp. 193–204). Leuven, Belgium: KU Leuven. Retrieved from https://bit.ly/3jOHQBR

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.

Brasil, A. (2021e). When indicators are not enough: The Brazilian Qualis experience in combining metrics with peer-review. In *Proceedings of the nordic workshop of bibliometrics and research policy*, Odense, DK: University of Southern Denmark.

Brasil, A. (2022). Multidimensionality through self-evaluation: from theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation*, (53). https://doi.org/10.22163/fteval.2022.546

Brasil, A. (2023a). Between bibliometrics and peer review: The evolution and challenges of Brazil's Qualis system. preprint. Retrieved from https://bit.ly/412rlnA

Brasil, A. (2023b). Rethinking a national classification of research and graduate education. *Profesional de la Información*, *32*(2), e320224. https://doi.org/10.3145/epi.2023.mar.24

Brasil, A. (2023c). The dynamics of a national evaluation system: The Brazilian model and its role in developing research and graduate education in the country. *SocArXiv*. https://doi.org/10.10100/xxx.202xxxx

Brasil, A., & Trevisol, J. V. (2023). *ENRESSH country report: Brazil*. OSF. https://doi.org/10.17605/OSF.IO/ZH465

Brasil, A., Trevisol, J. V., & Drooge, L. v. (2022). Research evaluation in Brazil and The Netherlands: A comparative study. In *Eu-spri 2022*, Utrecht, NL. Retrieved December 31, 2022, from https://bit.ly/3gRMNuo

Brennan, J., & Shah, T. (2000). Quality assessment and institutional change: Experiences from 14 countries. *Higher Education*, *40*(3), 331–349. https://doi.org/10.1023/a:1004159425182

Brigatti, F. (2023). Economistas se rebelam contra mudança em ranking de revistas científicas. *Folha de São Paulo*. Retrieved from https://bit.ly/3AesptT

British Council. (2014). *Learning English in Brazil: Understanding the aims and expectations of the Brazilian emerging middle classes*. São Paulo, Brazil. Retrieved from https://bit.ly/3b3lbeZ

Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy and Systems*, *15*(1), 35. https://doi.org/10.1186/s12961-017-0192-x

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. https://doi.org/10.3152/095820208x280907

Cano, W. (2015). Crise e industrialização no Brasil entre 1929 e 1954: A reconstrução do Estado Nacional e a Política Nacional de Desenvolvimento. *Brazilian Journal of Political Economy*, *35*(3), 444–460. https://doi.org/10.1590/0101-31572015v35n03a04

Capano, G. (2010). A Sisyphean task: Evaluation and institutional accountability in Italian higher education. *Higher Education Policy*, *23*(1), 39–62. https://doi.org/10.1057/hep.2009.19

Capano, G., Regini, M., & Turri, M. (2016). *Changing Governance in Universities*. London: Palgrave Macmillan.

CAPES. (1981). *Plano Nacional de Pós-Graduação - PNPG 1982-1985*. Brasília, DF: CAPES.

CAPES. (1985). *Plano Nacional de Pós-Graduação - PNPG 1986-1989*. Brasília, DF: CAPES.

CAPES. (1995b). *Programa de flexibilização do modelo de pós-graduação senso estrito em nível de mestrado*. Brasília, DF. Retrieved from https://bit.ly/3sp5Aic

CAPES. (2003). *QUALIS: Concepção e diretrizes básicas*. Brasília. Retrieved July 24, 2020, from https://bit.ly/3u77Iuy

CAPES. (2004). *Plano Nacional de Pós-Graduação - PNPG 2005-2010*. Brasília, DF: CAPES.

CAPES. (2009). Capes lança aplicativo gráfico para consulta de dados. Retrieved December 7, 2021, from https://bit.ly/3IsaOBw

CAPES. (2010). *Plano Nacional de Pós-Graduação - PNPG 2011-2020*. Brasília, DF: CAPES.

CAPES. (2017a). Avaliação Quadrienal 2017: Hotsite. Retrieved July 4, 2020, from http://avaliacaoquadrienal.capes.gov.br/

CAPES. (2018a). Avaliação comparada da pós-graduação. Seminários: Repensando a Avaliação. Retrieved February 2022, from https://bit.ly/3HyExbc

CAPES. (2018b). Avaliação da produção intelectual de programas de pós-graduação. Seminários: Repensando a Avaliação. Retrieved February 2022, from https://bit.ly/3AYGFqb

CAPES. (2018c). *Avaliação Quadrienal em números*. Brasília, DF. Retrieved from https://bit.ly/2Y2rPvO

CAPES. (2019a). *Aprimoramento do processo de avaliação da pós-graduação*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3wDRgDU

CAPES. (2019b). Meio-Termo: refletir para avançar. Retrieved April 21, 2023, from https://bit.ly/41rui2k

CAPES. (2020a). *Metodologia do Qualis Referência: Documento base aprovado no CTC 198, de setembro de 2020*. CAPES. Brasília, DF. Retrieved April 20, 2023, from https://bit.ly/41w96rW

CAPES. (2020d). Sobre as áreas de avaliação. Retrieved April 9, 2021, from https://bit.ly/3s0XB9f

CAPES. (2020e). *Tabela de Áreas de Conhecimento/Avaliação*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3I7infy

CAPES. (2021b). Avaliação Quadrienal. Retrieved June 18, 2021, from https://bit.ly/35wv16S

CAPES. (2022a). CAPES divulgará os resultados parciais. Retrieved May 13, 2023, from https://bit.ly/3W0LObG

CAPES. (2022b). Resultado da Avaliação Quadrienal 2017-2020. Retrieved January 5, 2023, from https://bit.ly/3Gj9BfE

CAPES. (2023a). Aumento no valor das bolsas marca ações da CAPES. Retrieved May 14, 2023, from https://bit.ly/3nWhkep

CAPES. (2023b). *Documento técnico do Qualis Periódicos*. CAPES. Brasília, DF. Retrieved April 12, 2023, from https://bit.ly/3KTmUGY

CAPES. (2023c). Metodologia do Qualis Referência: Quadriênio 2017-2020. Retrieved April 12, 2023, from https://bit.ly/3my156n

Cardoso, M. S. (2015). Sistema Geocapes de dados estatísticos. Brasília, DF. Retrieved December 2021, from https://bit.ly/3DyZmR1

Castro, C. d. M. (2006). A CAPES na visão de um ex-diretor. *Análise*, *17*(2), 360–376. Retrieved from https://bit.ly/2VNnRGY

Castro, C. d. M., & Soares, G. A. D. (1983). Avaliando as avaliações da Capes. *Revista de Administração de Empresas*, *23*(3), 63–73. https://doi.org/10.1590/s0034-75901983000300007

Castro, M. H. d. M. (2015). Higher education policies in Brazil: A case of failure in market regulation. In S. Schwartzman, R. Pinheiro, & P. Pillay (Eds.), *Higher education in the brics countries* (pp. 271–289). Dordrecht: Springer.

Centraal Bureau voor de Statistiek. (2021). *Nederland in cijfers: Aan de hand van 36 vragen verbeeld*. Den Haag: CBS. Retrieved February 2022, from https://bit.ly/3gv46h6

Centraal Bureau voor de Statistiek. (2022). 40 percent international first-year students at Dutch universities. CBS. Retrieved May 2, 2023, from https://bit.ly/40Y9YEG

CGEE. (2016). *Mestres e doutores 2015: Estudos da demografia da base técnico-científica brasileira*. Brasília, DF: CGEE.

CGEE. (2019). *Brasil: Mestres e Doutores 2019*. CGEE. Brasília. Retrieved November 26, 2022, from https://mestresdoutores2019.cgee.org.br

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., ... Borges, B. (2021). *Shiny: Web application framework for r*. R package version 1.7.1. Retrieved from https://CRAN.R-project.org/package=shiny

Chavarro, D., Ràfols, I., & Tang, P. (2018). To what extent is inclusion in the Web of Science an indicator of journal 'quality'? *Research Evaluation*, *27*(2), 106–118. https://doi.org/10.1093/reseval/rvy001

Clarivate Analytics. (2019). *Research in Brazil: Funding excellence*. Retrieved from https://bit.ly/3gdf5vY

Clark, B. R. (Ed.). (1993). *The research foundations of graduate education: Germany, Britain, France, United States, Japan*. Berkeley, CA: University of California Press.

CNPG. (1974). *Plano Nacional de Pós-Graduação - PNPG 1975-1979*. Brasília, DF: MEC.

Cohen, A., & Steege, M. v. d. (1982). An Historical Overview of the State and Higher Education in the Netherlands. *European Journal of Education*, *17*(3), 271. https://doi.org/10.2307/1502985

Colglazier, B. (2016). Encourage governments to heed scientific advice. *Nature*, *537*(7622), 587–587. https://doi.org/10.1038/537587a

Comissão Especial de Acompanhamento do PNPG 2011-2020. (2018). *Proposta de aprimoramento do modelo de avaliação da PG*. CAPES. Brasília, DF. Retrieved from 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*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3wyR097

Copiello, S. (2020). Business as Usual with Article Processing Charges in the Transition towards OA Publishing: A Case Study Based on Elsevier. *Publications*, *8*(1), 3. https://doi.org/10.3390/publications8010003

Córdova, R. d. A. (2001). *CAPES: Origem, Realizações, Significações (1951-2001)*. Brasília, DF.

Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and Mechanisms for Evaluating Government-Funded Research. *American Journal of Evaluation*, *28*(4), 437–457. https://doi.org/10.1177/1098214007308290

Cunha, L. A. (2007). *A universidade temporã: o ensino superior da colônia à era Vargas* (3rd ed.). São Paulo: Editora UNESP.

Cuppen, H. (2021). Opeens nieuwe naam voor vereniging universiteiten. Retrieved August 8, 2023, from https://bit.ly/468FByR

Cury, C. R. J. (2005). Quadragésimo ano do parecer CFE nº 977/65. *Revista Brasileira de Educação*, (30), 07–20. https://doi.org/10.1590/s1413-24782005000300002

CWTS. (2022). Leiden Ranking. Retrieved May 24, 2023, from https://www.leidenranking.com/

da Fonseca, M. L. M. (2015). A institucionalização da pesquisa científica brasileira: Os primeiros anos de atuação do Conselho Nacional de Pesquisas (CNPq). *Parcerias Estratégicas*, *18*(36), 253–268.

David, P. A. (1985). Clio and the Economics of QWERTY. *The American Economic Review*, *74*(2), 332–337.

Davidovich, L., & Ribeiro, R. J. (2021). Manifestação da ABC e SBPC contra suspensão da Avaliação Quadrienal 2021. Retrieved March 10, 2022, from https://bit.ly/3q12v8j

de Almeida, E. C. E., & Guimarães, J. A. (2013). Brazil's growing production of scientific articles—how are we doing with review articles and other qualitative indicators? *Scientometrics*, *97*(2), 287–315. https://doi.org/10.1007/s11192-013-0967-y

de Almeida, E. C. E., Guimarães, J. A., & Alves, I. T. G. (2010). Dez anos do Portal de Periódicos da Capes: Histórico, evolução e utilização. *Revista Brasileira de Pós-Graduação*, *7*(13), 218–246. https://doi.org/10.21713/2358-2332.2010.v7.194

de Andrade, D., Rizzi, P., & de Almeida, S. F. M. (2005). Experiência em parceria estratégica: mestrado profissional em Engenharia Aeronáutica, ITA-Embraer. *Revista Brasileira de Pós-Graduação*, *2*(4), 48–60. https://doi.org/10.21713/2358-2332.2005.v2.77

de Biasi, K. (2019). *Solving the Change Paradox by means of trust*. https://doi.org/10.1007/978-3-658-23912-1

Deem, R., & Dowle, S. (2020). The UK Doctorate: History, Features and Challenges. In M. Yudkevich, P. G. Altbach, & H. d. Wit (Eds.), *Trends and issues in doctoral education: A global perspective* (1st ed., pp. 152–178). https://doi.org/10.4135/9789353885991.n7

Dellagostin, O. (2020). Semelhanças e diferenças entre os estados no financiamento à pesquisa e na pós-graduação [webinar]. presentation. Retrieved November 13, 2020, from https://youtu.be/JFBKShfJKDM

de Vasconcelos, S. M. R. (2008). *Ciência no Brasil: Uma abordagem cienciométrica e lingüística* (Doctoral dissertation, Universidade Federal do Rio de Janeiro).

Dias, A. M. I., Therrien, J., & Farias, I. M. S. d. (2017). As áreas da educação e de ensino na Capes: identidade, tensões e diálogos. *Revista Educação e Emancipação*, *10*(1), 34–57. https://doi.org/10.18764/2358-4319.v10n1p34-57

Dias, M. (2019). Bolsonaro diz que manifestantes contra cortes na educação são idiotas úteis e massa de manobra. *Folha de São Paulo*. Retrieved from https://bit.ly/4386EZ0

Donnelly, C. A., Boyd, I., Campbell, P., Craig, C., Vallance, P., Walport, M., ... Wormald, C. (2018). Four principles to make evidence synthesis more useful for policy. *Nature*, *558*(7710), 361–364. https://doi.org/10.1038/d41586-018-05414-4

dos Santos, P. J. P., de Oliveira, T. M., Paixão, F. d. A. J., Pascutti, P., Amado, A. M., Oliveira, D. d. H., ... Mamiya, E. N. (2019). *Proposta do GT Qualis Periódicos*. Brasília, DF.

Dourado, C. (2023). Fuga de cérebros, a diáspora de cientistas brasileiros. Retrieved May 14, 2023, from https://bit.ly/3MabZs6

Education First. (2020). *EF English proficiency index: A ranking of 100 countries and regions by English skills*. Retrieved from https://www.ef.nl/epi/

Else, H. (2018). How Unpaywall is transforming open science. *Nature*, *560*(7718), 290–291. https://doi.org/10.1038/d41586-018-05968-3

Elton, L. (2004). Goodhart's Law and Performance Indicators in Higher Education. *Evaluation & Research in Education*, *18*(1-2), 120–128. https://doi.org/10.1080/09500790408668312

Elven, M. v. (2023). Hoe het Nederlandse hoger onderwijs populair werd bij internationals. Retrieved May 3, 2023, from https://bit.ly/44lwv10

European Ministers of Education. (1999). *Bologna Declaration*. Bologna. Retrieved from https://bit.ly/3OxS8lR

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.

Fanelli, A. G. D. (2016). U-Multirank and Latin American Universities. *International Higher Education*, (84), 7–9. https://doi.org/10.6017/ihe.2016.84.9108

Farmer, R. (2010). How to Influence Government Policy with Your Research: Tips from Practicing Political Scientists in Government. *PS: Political Science & Politics*, *43*(4), 717–719. https://doi.org/10.1017/s1049096510001368

Fávero, M. d. L. d. A. (2006). A universidade no Brasil: Das origens à reforma universitária de 1968. *Educar em Revista*, *22*(28), 17–36. https://doi.org/10.1590/s0104-40602006000200003

Fealing, K. H., Lane, J., Marburger III, J. H., & Shipp, S. (Eds.). (2011). *The Science of Science Policy: A Handbook*. Stanford: Stanford University Press.

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In S. Bartling & S. Friesike (Eds.), *Opening science* (pp. 17–47). Springer, Cham.

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.), *Multi-dimensional ranking: The design and development of U-Multirank* (pp. 167–177). https://doi.org/10.1007/978-94-007-3005-2_10

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). https://doi.org/10.1007/978-94-007-3005-2_6

Ferrari, L. (2023). Pesquisadores criticam novo 'ranking' de revistas científicas da Capes, o Qualis. *Estadão*. Retrieved from https://bit.ly/41EGuMY

Ferreira, M. d. M., & Moreira, R. d. L. (Eds.). (2002). *CAPES, 50 anos: Depoimentos ao CPDOC/FGV*. Rio de Janeiro: Fundação Getúlio Vargas.

Fink, D. (2006). The professional doctorate: Its relativity to the Ph.D. and relevance for the knowledge economy. *International Journal of Doctoral Studies*, *1*(1), 35–44. https://doi.org/10.28945/59

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

Furlaneto, A. (2019). Acadêmicos de universidades em todo o mundo assinam manifesto contra cortes de verbas para Ciências Humanas no Brasil. *O Globo*. Retrieved from http://glo.bo/42Y5QGE

Gadd, E., Holmes, R., & Shearer, J. (2021). Developing a Method for Evaluating Global University Rankings. *Scholarly Assessment Reports*, *3*(1), 1–19. https://doi.org/10.29024/sar.31

Gaieck, W., Lawrence, J., Montchal, M., Pandori, W., & Valdez-Ward, E. (2020). Opinion: Science policy for scientists: A simple task for great effect. *Proceedings of the National Academy of Sciences*, *117*(35), 20977–20981. https://doi.org/10.1073/pnas.2012824117

Galleron, I., Ochsner, M., Spaapen, J., & Williams, G. (2017). Valorizing SSH research: Towards a new approach to evaluate SSH research' value for society. *fteval Journal for Research and Technology Policy Evaluation*, *44*, 35–41. https://doi.org/10.22163/fteval.2017.274

Garfield, E. (1995). Quantitative analysis of the scientific literature and its implications for science policymaking in Latin America and the Caribbean. *Bulletin of the Pan American Health Organization (PAHO)*, *29*(1), 87–95.

Gatti, B., André, M., Fávero, O., & Candau, V. M. F. (2003). O modelo de avaliação da CAPES. *Revista Brasileira de Educação*, (22), 137–144. https://doi.org/10.1590/s1413-24782003000100012

Gaughan, M., & Robin, S. (2004). National science training policy and early scientific careers in France and the United States. *Research Policy*, *33*(4), 569–581. https://doi.org/10.1016/j.respol.2004.01.005

Gellert, C. (1993). The German Model of Research and Advanced Education. In B. R. Clark (Ed.), *The research foundations of graduate education: Germany, britain, france, united states, japan* (pp. 5–44). Berkeley, CA: University of California Press.

Gershon, R. A. (2013). Innovation failure: A case study analysis of Eastman Kodak and Blockbuster Inc. In A. B. Albarran (Ed.), *Media management and economics research in a transmedia environment* (pp. 46–68). New York: Routledge.

Geuna, A., & Martin, B. R. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva*, *41*(4), 277–304. https://doi.org/10.1023/b:mine.0000005155.70870.bd

Gibbs, W. W. (1995). Lost science in the third world. *Scientific American*, *273*(2), 92–99. https://doi.org/10.1038/scientificamerican0895-92

Giolo, J., Leher, R., & Sguissardi, V. (2020). *Future-se: Ataque à autonomia das instituições federais de educação superior e sua sujeição ao mercado*. São Carlos, SP: Diagrama Editorial.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193. https://doi.org/10.1023/a:1014848323806

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*(3), 357–367. https://doi.org/10.1023/a:1022378804087

Gluckman, P. (2014). Policy: The art of science advice to government. *Nature*, *507*(7491), 163–165. https://doi.org/10.1038/507163a

Goedegebuure, L., & Westerheijden, D. F. (1991). Changing balances in Dutch higher education. *Higher Education*, *21*(4), 495–520. https://doi.org/10.1007/bf00134986

Gouvêa, F. C. F. (2012). A institucionalização da pós-graduação no Brasil: O primeiro decênio da CAPES (1951-1961). *Revista Brasileira de Pós-Graduação*, *9*(17), 373–397. https://doi.org/10.21713/2358-2332.2012.v9.312

Hammarfelt, B., & de Rijcke, S. (2014). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, *24*(1), 63–77. https://doi.org/10.1093/reseval/rvu029

Hanzen, E. (2021). Maior crise da história da Capes é agravada com renúncia coletiva de coordenadores de área e consultores. *Jornal da Universidade*. Retrieved from https://bit.ly/3nYkApy

Hathaway, O. A. (2003). Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.384595

Hazelkorn, E., & Gibson, A. (2017). Global science, national research, and the question of university rankings. *Palgrave Communications*, *3*(1), 21. https://doi.org/10.1057/s41599-017-0011-6

HEFCE. (2015). *Annual report and accounts 2014-15*. Higher Education Funding Council for England. Bristol. Retrieved from https://bit.ly/3c9sfbC

Herculano-Houzel, S. (2013). A profissão cientista não existe no Brasil. Retrieved November 26, 2022, from https://youtu.be/CF8UeugZW7c

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*(2), 251–261. https://doi.org/10.1016/j.respol.2011.09.007

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

Holanda, L. (2021). Capes encerra repasses de verbas a institutos e surpreende cientistas. *Agência Brasil*. Retrieved from https://bit.ly/3szBkTb

Horta, J. S. B., & Moraes, M. C. M. d. (2005). O sistema CAPES de avaliação da pós-graduação: Da área de educação à grande área de ciências humanas. *Revista Brasileira de Educação*, *10*(30), 95–116. https://doi.org/10.1590/s1413-24782005000300008

Hortale, V. A. (2003). Modelo de avaliação CAPES: desejável e necessário, porém, incompleto. *Cadernos de Saúde Pública*, *19*(6), 1837–1840. https://doi.org/10.1590/s0102-311x2003000600027

Hostins, R. C. L. (2006). Os Planos Nacionais de Pós-graduação (PNPG) e suas repercussões na pós-graduação brasileira. *Perspectiva*, *24*(1), 1–28.

INEP. (2022). *Censo da Educação Superior 2021: Notas estatísticas*. INEP/MEC. Brasília, DF. Retrieved May 2, 2023, from https://bit.ly/42h1Yje

Jubb, M., Plume, A., Oeben, S., Brammer, L., Johnson, R., Bütün, C., & Pinfield, S. (2017). *Monitoring the transition to Open Access*. Universities UK. Retrieved from https://bit.ly/3vqUHPY

Kobzar, E., & Roshchin, S. (2020). Russian Doctoral Education: Between Teaching and Research. In M. Yudkevich, P. G. Altbach, & H. d. Wit (Eds.), *Trends and issues in doctoral education: A global perspective* (1st ed., pp. 127–151). https://doi.org/10.4135/9789353885991.n6

Kyvik, S., & Tvede, O. (2010). The Doctorate in the Nordic Countries. *Comparative Education*, *34*(1), 9–25. https://doi.org/10.1080/03050069828315

Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation Systems: What are they and why study them? *Evaluation*, *14*(2), 157–169. https://doi.org/10.1177/1356389007087537

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. https://doi.org/10.1590/s1414-4077/s1414-40772020000200006

Leite, M. F. B., Viana, A. B. N., & Pereira, G. G. (2010). Métodos quantitativos na avaliação da CAPES: Uma pesquisa bibliográfica. *FACEF Pesquisa - Desenvolvimento e Gestão*, *9*(2), 166–174. Retrieved from https://bit.ly/2ZEYg3X

Leydesdorff, L., Etzkowitz, H., Ivanova, I., & Meyer, M. S. (2017). The Measurement of Synergy in Innovation Systems: Redundancy Generation in a Triple Helix of University-Industry-Government Relations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2937647

Lim, H., Kim, S. J., & Shin, J. C. (2020). Rapid Development and Current Rethinking in Doctoral Education in South Korea. In M. Yudkevich, P. G. Altbach, & H. d. Wit (Eds.), *Trends and issues in doctoral education: A global perspective* (1st ed., pp. 363–386). https://doi.org/10.4135/9789353885991.n15

London, S. (1996). Understanding change: The dynamics of social transformation. Retrieved July 8, 2021, from https://bit.ly/3k0MFq7

Lord Robbins, L. C. (1963). *The Robbins Report*. Retrieved from https://bit.ly/2WrVdeS

Machado, R. (2022). Mais 3 pesquisadores ligados à Capes renunciam e aumentam debandada. *Folha de São Paulo*. Retrieved from https://bit.ly/3MMCEdW

Maciel, C. E., Trierweiller, A. C., & Ferenhof, H. A. (2019). Perspectiva do Usuário na Utilização da Plataforma Sucupira como Sistema de Avaliação: Uma Analise Exploratória. (pp. 1965–1975). https://doi.org/10.5151/9cidi-congic-4.0350

Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, *29*(4), 507–548. https://doi.org/10.1023/a:1007113830879

Maia, M. A. Q. (2020). *Sistema de informação da pós-graduação brasileira: Avaliação da Plataforma Sucupira* (Doctoral dissertation, Universidade Federal de Minas Gerais).

Marburger III, J. H. (2011). Why policy implementation needs a Science of Science policy. In K. H. Fealing, J. Lane, J. H. Marburger III, & S. Shipp (Eds.), *The science of science policy: A handbook* (pp. 31–55). Stanford: Stanford University Press.

Marcovitch, J., Goldemberg, J., Axel-Berg, J., Righetti, S., dos Santos, S. M., de Oliveira, L. N., . . . Guimarães, J. A. C. (2018). *Repensar a universidade: desempenho acadêmico e comparações internacionais* (J. Marcovitch, Ed.). https://doi.org/10.11606/9788571661868

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. https://doi.org/10.22347/2175-2753v12i37.2825

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

Marques, J. (2021). Crise na Capes ameaça qualidade da pesquisa e divisão de verbas para pós-graduação. *Estadão*. Retrieved from https://bit.ly/3o12nb2

Martinelli, A. E., Lima, A. C. S., Costa, F. N., Fernandes, L. M. R., Brito, M. V. H., Possamai, O., . . . Avellar, S. O. d. C. (2019). *Impacto e relevância econômica e social*. Retrieved from https://bit.ly/3vTjxFO

Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). *Revista Brasileira de Sociologia*, *6*(13), 9–26. https://doi.org/10.20336/rbs.256

Martins, P. (2019). Cortes nas universidade federais são 'colapso anunciado com calma e precisão'. Retrieved May 19, 2023, from https://bit.ly/3MEbeZF

Maxwell, T. W., & Kupczyk-Romanczuk, G. (2009). Producing the professional doctorate: the portfolio as a legitimate alternative to the dissertation. *Innovations in Education and Teaching International*, *46*(2), 135–145. https://doi.org/10.1080/14703290902843760

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

Ministry of Education: People's Republic of China. (2022). Educational Statistics. Retrieved from https://bit.ly/3yiASfp

Moacyr, P. (1937). *A instrução e o Império: Subsídios para a história da educação no Brasil*. São Paulo: Companhia Editora Nacional.

Moed, H. F. (2006). *Citation analysis in research evaluation*. Dordrecht: Springer.

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, *110*(2), 967–990. https://doi.org/10.1007/s11192-016-2212-y

Moed, H. F., & Visser, M. S. (2007). *Developing bibliometric indicators of research performance in computer science: An exploratory study*. Leiden.

Molas-Gallart, J. (2012). Research Governance and the Role of Evaluation: a comparative study. *American Journal of Evaluation*, *33*(4), 583–598. https://doi.org/10.1177/1098214012450938

Molas-Gallart, J. (2014). Research evaluation and the assessment of public value. *Arts and Humanities in Higher Education*, *14*(1), 111–126. https://doi.org/10.1177/1474022214534381

Montana, J., & Wilsdon, J. (2021). Analysts, advocates and applicators: three discourse coalitions of UK evidence and policy. *Evidence & Policy*, *20*(20), 1–14. https://doi.org/10.1332/174426421x16112601473449

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*. Brasília, DF: CAPES. Retrieved from https://bit.ly/2PmAf0q

Monteiro, A. L., Meneghetti, S. M. P., Suarez, P. A. Z., Rubira, A. F., Zarbin, l. J. G., Rezende, C. M. d., . . . Coltro, W. K. T. (2021). Carta aberta dos Coordenadores e Consultores da Área de Química na CAPES. Retrieved May 14, 2022, from https://bit.ly/3sBhVkO

Monteiro, V. (2015). Regulamentação da profissão de cientista em debate. *Jornal da Ciência*. Retrieved from https://bit.ly/3GNrm8s

Morato, N. (2015). Estratégias e perspectivas para a pós-graduação são debatidas em encontro. Retrieved May 13, 2023, from https://bit.ly/44TT2ST

Mounier, P. (2018). 'Publication favela' or bibliodiversity? Open access publishing viewed from a European perspective. *Learned Publishing*, *31*(S1), 299–305. https://doi.org/10.1002/leap.1194

Moura, M. (2019). Universidades públicas realizam mais de 95% da ciência no Brasil. Retrieved November 26, 2022, from https://bit.ly/3gEVbxw

Muenchen, R. A. (2019). The popularity of data science software. Retrieved December 28, 2021, from http://r4stats.com/articles/popularity/

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton & Oxford: Princeton University Press.

Nader, H., & Davidovich, L. (2021). Crise na Capes e a batalha pela Ciência brasileira. *Piauí*. Retrieved May 13, 2023, from https://bit.ly/44ZdMsw

Nader, H., Palis, J., Sampaio, F., Garcia, F. et al.. (2015). Carta ao ministro da Educação contra os possíveis cortes no Portal de Periódicos da CAPES. letter. Retrieved August 5, 2020, from http://www.sbpcnet.org.br/site/arquivos/arquivo%5C_467.pdf

Nerad, M. (2008). United States of America. In M. Nerad & M. Heggelund (Eds.), *Toward a global phd? forces and forms in doctoral education worldwide* (pp. 278–297). Seattle, WA: University of Washington Press.

Nobre, L. N., & de Freitas, R. R. (2017). A evolução da pós-graduação no Brasil: Histórico, políticas e avaliação. *Brazilian Journal of Production Engineering*, *3*(2), 18–30. Retrieved from https://bit.ly/38sAuMO

NUFFIC. (2023). Higher education in The Netherlands. Retrieved May 2, 2023, from https://bit.ly/3oPGfAf

Ochsner, M., Kulczycki, E., & Gedutis, A. (2018). The Diversity of European Research Evaluation Systems. In *Proceedings...* (pp. 1235–1241). Leiden, NL.

Ochsner, M., Kulczycki, E., Gedutis, A., & Girkontaitė, A. (2019). The Diversity of Research Evaluation Systems. Vilnius, LT. Retrieved April 18, 2023, from https://bit.ly/3KL3HpQ

Ochsner, M., Kulczycki, E., Gedutis, A., & Peruginelli, G. (2020). National Research Evaluation Systems. In R. Ball (Ed.), *Handbook bibliometrics* (pp. 99–106). https://doi.org/10.1515/9783110646610-011

OECD. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. Retrieved from http://oe.cd/frascati

OECD. (2021). Education at a Glance 2021. *Education at a Glance*. The data set is available at https://bit.ly/3nePKoB. https://doi.org/10.1787/b35a14e5-en

Oliva, G., Tourinho, E. Z., Neto, E. d. M., Audy, J. L. N., Bevilacqua, L., Laplane, M. F., . . . Báo, S. N. (2017). *Comissão Especial para Análise do Sistema e Processo de Avaliação da Qualidade da Pós-graduação Brasileira*. Brasília.

Packer, A. L. (2010). The SciELO Open Access: A Gold Way from the South. *Canadian Journal of Higher Education*, *39*(3), 111–126. https://doi.org/10.47678/cjhe.v39i3.479

Paixão, R. B., Bruni, A. L., Becker, J. L., & Tenório, R. M. (2014). Avaliação de mestrados profissionais: construção e análise de indicadores à luz da multidimensionalidade. *Ensaio: Avaliação e Políticas Públicas em Educação*, *22*(83), 505–532. https://doi.org/10.1590/s0104-40362014000200010

Pascutti, P. (2016). Entendendo o Qualis Periódicos CAPES. Campinas, SP. Retrieved April 2022, from https://bit.ly/40leSuZ

Passarelli, H. (2019). 'Ideia de universidade para todos não existe', diz ministro da Educação. *Valor Econômico*. Retrieved from http://glo.bo/3IBD6uY

Pavan, C., & Barbosa, M. C. (2018). Article processing charge (APC) for publishing open access articles: the Brazilian scenario. *Scientometrics*, *117*(2), 805–823. https://doi.org/10.1007/s11192-018-2896-2

Peternelli Jr, R. S. (2019). *PEC 206/2019. Dispõe sobre a cobrança de mensalidade pelas universidades públicas*. Câmara dos Deputados. Brasília, DF. Retrieved May 19, 2023, from https://bit.ly/3BEQ6Mr

Plotly Technologies. (2015). Collaborative data science. Montreal, QC: Plotly Technologies Inc. Retrieved from https://plot.ly

Polonsky, M. J., & Whitelaw, P. (2005). What Are We Measuring When We Evaluate Journals? *Journal of Marketing Education*, *27*(2), 189–201. https://doi.org/10.1177/0273475305276842

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org

Ràfols, I., Molas-Gallart, J., Chavarro, D., & Robinson-Garcia, N. (2016). On the dominance of quantitative evaluation in peripherall countries: Auditing research with technologies of distance. *SSRN Electronic Journal*, 1–22. https://doi.org/10.2139/ssrn.2818335

Raju, R., Claassen, J., Pietersen, J., & Abrahamse, D. (2020). An authentic flip subscription model for Africa: library as publisher service. *Library Management*, *41*(6/7), 369–381. https://doi.org/10.1108/lm-03-2020-0054

Reategui, E., Pires, A. d. S., Carniato, M., & Franco, S. R. K. (2020). Evaluation of Brazilian research output in education: confronting international and national contexts. *Scientometrics*, *125*(1), 427–444. https://doi.org/10.1007/s11192-020-03617-z

Rezende, C. (2020). Weintraub: 'Não quero sociólogo, antropólogo e filósofo com meu dinheiro'. Retrieved May 19, 2023, from https://bit.ly/3WhXmHw

Ribeiro, D. (1961). Universidade de Brasília. *Revista Brasileira de Estudos Pedagógicos*, *36*(83), 161–230.

Ribeiro, M. (2021). Ministro da Educação, Milton Ribeiro, é o convidado do Sem Censura. Retrieved May 25, 2023, from https://bit.ly/4391lZU

Ribeiro, R. J. (2008). Quality in Research: CAPES' point of view. In J. L. N. Audy & M. C. Morosini (Eds.), *Innovation and quality in the university* (pp. 127–169). Porto Alegre: EDIPUCRS.

Ribeiro, R. J. (2022a). Carta da SBPC à CAPES sobre audiência pública para debater a Avaliação Quadrienal. Retrieved from https://bit.ly/3Ki9QYx

Ribeiro, R. J. (2022b). Depoimento do presidente da SBPC sobre a Avaliação Quadrienal da CAPES. Rio de Janeiro. Retrieved from https://bit.ly/3tp0FyM

Rijksoverheid. (2023). Tertiary (higher) education. Retrieved May 2, 2023, from https://bit.ly/3nmJChP

Rossini, M. C. (2021). Entenda o corte de R$ 600 milhões no orçamento da ciência brasileira. *Superinteressante*. Retrieved March 12, 2022, from https://bit.ly/3t1lsJH

Rothen, J. C. (2018). A universidade brasileira na Reforma Francisco Campos de 1931. *Revista Brasileira de História da Educação*, *8*(17), 141–160.

Rothen, J. C. (2019). O ensino superior e a Nova Gestão Pública: Aproximações do caso brasileiro com o francês. *Revista Eletrônica de Educação*, *13*(3), 970–995. https://doi.org/10.14244/198271993549

Saldaña, P. (2020). Em meio a pandemia, governo Bolsonaro investe contra pesquisa em ciências humanas. Retrieved August 10, 2020, from https://bit.ly/2PHEUH8

Saldaña, P. (2021a). Capes tem déficit de R$ 124 mi para pagar bolsas de formação docente até fim do ano. *Folha de São Paulo*. Retrieved from https://bit.ly/3a0WGm1

Saldaña, P. (2021b). Diretor de Avaliação da Capes pede demissão após renúncia de pesquisadores. *Folha de São Paulo*. Retrieved from https://bit.ly/42SWT0M

Santos, P. J. P. d., Oliveira, T. M. d., Paixão, F. d. A. J., Pascutti, P., Amado, A. M., Oliveira, D. d. H., ... Guterres, S. S. (2020). *Relatório do GT Qualis Periódicos*. CAPES. Brasília, DF.

Sapolsky, H. M., & Taylor, M. Z. (2011). Politics and the Science of Science Policy. In K. H. Fealing, J. Lane, J. H. Marburger III, & S. Shipp (Eds.), *The science of science policy: A handbook* (pp. 9–22). Stanford: Stanford University Press.

Sarmento, R., & Costa, V. (2017). *Comparative approaches to using R and Python for statistical data analysis*. Hershey, PA: IGI Global.

Saviani, D. (2020). Meio século de pós-graduação no Brasil: Do período heróico ao produtivismo pela mediação de um modelo superior às suas matrizes. *Movimento - Revista de Educação*, *7*(14). https://doi.org/10.22409/mov.v7i14.46475

SBPC. (2021a). Carta aberta da SBPC sobre a CAPES. Retrieved March 10, 2022, from https://bit.ly/3CnkQ4o

SBPC. (2021b). Entidades se manifestam em defesa da Capes. Retrieved May 14, 2022, from https://bit.ly/38xTKN8

SBPC, & ABC. (2020a). A história da CAPES na voz dos seus ex-presidentes [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

SBPC, & ABC. (2020b). Qual o papel da CAPES na construção da ciência e da pesquisa no Brasil? [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

Scholten, W., van Drooge, L., & Diederen, P. (2018). *Excellence is extra-ordinary: Thirty years of focus on excellence in Dutch science policy*. The Hague.

Schwartzman, S. (1982). Avaliando a pós-graduação: a prática da teoria. *Instituto Universitário de Pesquisas do Rio de Janeiro*, (10).

Schwartzman, S. (1989). A ciência no período de pós-guerra. presentation. Brasília, DF.

Schwartzman, S. (1991). *A space for science: The development of the scientific community in Brazil*. University Park, Pennsylvania: Penn State Press.

Schwartzman, S., Pinheiro, R., & Pillay, P. (2015). *Higher education in the BRICS countries* (S. Schwartzman, R. Pinheiro, & P. Pillay, Eds.). Dordrecht: Springer.

Serafim, M. P., & Dias, R. d. B. (2020). A importância da ciência e das universidades públicas na resolução de problemas sociais. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, *25*(1), 1–4. https://doi.org/10.1590/s1414-40772020000100001

Sguissardi, V. (2006). A avaliação defensiva no 'modelo CAPES de avaliação': É possível conciliar avaliação educativa com processos de regulação e controle do Estado? *Perspectiva*, *24*(1), 49–88. Retrieved from https://bit.ly/2VKO88E

Siler, K., Haustein, S., Smith, E., Larivière, V., & Alperin, J. P. (2018). Authorial and institutional stratification in open access publishing: the case of global health research. *PeerJ*, *6*, e4269. https://doi.org/10.7717/peerj.4269

Siqueira, M. B. (2019). Sucupira - A Platform for the Evaluation of Graduate Education in Brazil. *Procedia Computer Science*, *146*(2019), 247–255. https://doi.org/10.1016/j.procs.2019.01.081

Siriwardhana, C. (2015). Promotion and Reporting of Research from Resource-Limited Settings. *Infectious diseases*, *8*, 25–9. https://doi.org/10.4137/idrt.s16195

Sobrinho, J. D. (2003). Avaliação da educação superior: Regulação e emancipação. *Avaliação: Revista da Avaliação da Educacão Superior*, *8*(2), 31–47. Retrieved from https://bit.ly/3I5MR2m

Soma, N. Y., Alves, A. D., & Yanasse, H. H. (2016). O Qualis Periódicos e sua utilização nas avaliações. *Revista Brasileira de Pós-Graduação*, *13*(30), 45–61. https://doi.org/10.21713/2358-2332.2016.v13.1128

Spaapen, J., & Drooge, L. v. (2011). Introducing 'productive interactions' in social impact assessment. *Research Evaluation*, *20*(3), 211–218. https://doi.org/10.3152/095820211x12941371876742

Spagnolo, F., & Souza, V. C. (2004). O que mudar na avaliação da Capes? *Revista Brasileira de Pós-Graduação*, *1*(2), 8–34. https://doi.org/10.21713/2358-2332.2004.v1.38

Stern, F. L. (2019). A criação da área de avaliação ciências da religião e teologia na Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). *ESPAÇOS - Revista de Teologia e Cultura*, *1*(26), 73–91. Retrieved from https://espacos.itespteologia.com.br/espacos/article/view/62

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Sucupira, N. (1980). Antecedentes e primórdios da pós-graduação. *Forum Educacional*, *4*(4), 3–18.

Sugimoto, C. R., & Larivière, V. (2018). *Measuring Research: What everyone needs to know*. https://doi.org/10.1093/wentk/9780190640118.001.0001

Taylor, J. (2008). Quality and standards: The challenge of the professional doctorate. *Higher Education in Europe*, *33*(1), 65–87. https://doi.org/10.1080/03797720802228191

Tenório, E. (2015). Regulamentação da profissão "cientista"no Brasil. Retrieved November 26, 2022, from https://bit.ly/3AOij3k

Tijssen, R. J. W., Mouton, J., van Leeuwen, T. N., & Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation*, *15*(3), 163–174. https://doi.org/10.3152/147154406781775904

Torfing, J. (2009). Rethinking path dependence in public policy research. *Critical Policy Studies*, *3*(1), 70–83. https://doi.org/10.1080/19460170903158149

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, *9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z. eprint: 1810.08473

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

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

Ulhôa, M. T. d. (2017). As consequências do QUALIS artístico. *ARJ – Art Research Journal / Revista de Pesquisa em Artes*, *3*(3), XLII–LI. https://doi.org/10.36025/arj.v3i3.12079

UNESCO. (2015). *ISCED fields of education and training 2013: Detailed field descriptions*. UNESCO Institute for Statistics. Montreal. Retrieved from http://dx.doi.org/10.15220/978-92-9189-179-5-en

UNL. (2022). Universiteiten van Nederland. Retrieved January 21, 2022, from https://www.universiteitenvannederland.nl/

Van Noorden, R. (2017). The science that's never been cited. *Nature*, *552*(7684), 162–164. https://doi.org/10.1038/d41586-017-08404-0

van Drooge, L. (2021a). Research evaluation in context 1: Introducing research evaluation in the Netherlands. Retrieved January 2022, from https://bit.ly/3IsQZJE

van Drooge, L. (2021b). Research evaluation in context 2: One joint protocol, three criteria and four aspects. Retrieved January 2022, from https://bit.ly/3Atb67G

van Drooge, L., de Jong, S., Faber, M., & Westerheijden, D. (2013). *Twenty years of research evaluation* (tech. rep. No. 8). Rathenau Instituut. The Hague. Retrieved from http://bit.ly/3cXp4mK

van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

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. https://doi.org/10.1007/s11192-005-0008-6

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). https://doi.org/10.1007/978-94-007-3005-2_1

Vasconcellos, J. (2021). Presidente da Capes nega a dissolução de conselho. *Correio Braziliense*. Retrieved from https://bit.ly/3KGMk7F

Veiga, J. M. (2021). Decisão judicial autoriza Capes a retomar Avaliação Quadrienal; retrieved May 13, 2023, from https://bit.ly/3BoBaSE

Vereniging Hogescholen. (2023). Instroom, inschrijvingen en diploma's. Retrieved May 2, 2023, from https://bit.ly/3Hwirbd

Verhine, R. E. (2008). Avaliação da CAPES: Subsídios para a reformulação do modelo. In D. Mancebo, J. d. R. Silva Jr, & J. F. de Oliveira (Eds.), *Reformas e políticas: Educação superior e pós-graduação no brasil* (pp. 165–188). Campinas: Alínea.

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

Verhine, R. E., & Dantas, L. (2009). Reflexões sobre o sistema de avaliação da CAPES a partir do V Plano Nacional de Pós-graduação. *Revista de Educação Pública*, *18*(37), 295–310.

Verhine, R. E., & de Freitas, A. A. d. S. M. (2012). A avaliação da educação superior: modalidades e tendências no cenário internacional. *Revista Ensino Superior Unicamp*, (7), 16–39. Retrieved from https://bit.ly/2C9herj

Verhine, R. E., & Souza, Â. R. d. (2021). Compreendendo a Crise Recente na Pós-Graduação Brasileira. Retrieved May 13, 2023, from https://bit.ly/42F0dNq

Verhine, R. E., Lima, A. C. S., Carvalho, D. B. B. d., Leite, D., Bertolin, J. C. G., Dantas, L. V., ... Rosa, D. L. (2019). *Autoavaliação de Programas de Pós-Graduação*. Brasília, DF: CAPES. Retrieved from https://bit.ly/3jTSRS8

Verhine, R. E., Faljoni-Alario, A., Federizzi, L. C., Ferreira, F., & Oliveira, T. M. d. (2021). *Equivalência entre conceitos e notas*. CAPES. Brasília, DF. Retrieved June 21, 2023, from https://bit.ly/3NF4tXP

Viana, W. F. (2018). *Sistema CAPES de avaliação da pós-graduação stricto sensu: Um estudo de caso da área de Administração Pública no Brasil* (Doctoral dissertation, Universidade de Brasília).

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *arXiv preprint arXiv:2005.10732v2*.

VSNU, KNAW, & NWO. (2003). *Standard Evaluation Protocol (2003-2009)*. The Hague. Retrieved from https://bit.ly/3grtaGP

VSNU, KNAW, & NWO. (2009). *Standard Evaluation Protocol (2009-2015)*. The Hague. Retrieved from https://bit.ly/3wujG32

VSNU, KNAW, & NWO. (2016). *Standard Evaluation Protocol (2015-2021)*. The Hague. Retrieved from https://bit.ly/3iJqtC3

VSNU, NFU, KNAW, NWO, & ZonMw. (2019). *Room for everyone's talent*. The Hague. Retrieved January 30, 2023, from https://bit.ly/3Jt9ilw

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

Waaijer, C. J. (2018). Why the public nature of university rankings torpedoes their value for benchmarking: an institutional researcher's opinion. In *Proceedings of the 23rd international conference on science and technology indicators* (pp. 1031–1037). Leiden, Netherlands.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

Waltman, L., Boekhout, H., & van der Weijden, I. (2019). A large scale bibliometric analysis of differences between male and female researchers: Time trends and career paths. In *Proceedings...* Amsterdam, Netherlands.

Weert, E. d., & Boezerooy, P. (2007). *Higher education in the Netherlands*. Center for Higher Education Policy Studies (CHEPS). Enschede, NL.

Westerheijden, D. F., Cremonini, L., Kolster, R., Kottmann, A., Redder, L., Soo, M., ... Weert, E. d. (2008). *New Degrees in the Netherlands: Evaluation of the Bachelor-Master Structure and Accreditation in Dutch Higher Education*. Ministerie van Onderwijs, Cultuur en Wetenschappen. Enschede, NL. Retrieved from https://bit.ly/3tNdAM4

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org

Winter, E., Brasil, A., Moraes, C. L. d., Granjeiro, J. M., Filho, J. R. d. S. V., Nichiata, L. Y. I., ... Pascutti, P. G. (2019). *Produção técnica*. Brasília. Retrieved from https://bit.ly/3t06SiJ

Yamashita, M. (2023). Lista Qualis e a duvidosa qualidade dos critérios de qualidade. *Questão de Ciência*. Retrieved from https://bit.ly/3UShMq1

Datasets

Brasil, A. (2021c). Brazilian evaluation areas, broad areas and groups, including original names in Portuguese and English translations. Dataset. https://doi.org/10.5281/zenodo.5795662

BRASIL. Ministério da Educação. (2021). Cadastro e-MEC de Instituições e Cursos de Educação Superior. Brasília, DF: MEC. Retrieved from https://emec.mec.gov.br

CAPES. (2020b). Orçamento da CAPES: Evolução em Reais. Brasília, DF: CAPES. Retrieved from https://bit.ly/37NYz0P

CAPES. (2020c). Portal de Periódicos. Retrieved August 17, 2021, from https://periodicos.capes.gov.br/

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/

CAPES. (2023d). Qualis Periódicos. Brasília, DF: CAPES. Retrieved from http://qualis.capes.gov.br

Clarivate Analytics. (2022). Web of Science Core Collection. CWTS in-house version. Retrieved from https://bit.ly/3rHmCJT

CNPq. (2020). Portal de Dados Abertos do CNPq. Retrieved from http://dadosabertos.cnpq.br/

DOAJ. (2021). Public data dump. Retrieved from https://doaj.org/docs/public-data-dump/

Eurostat. (2022). Graduates by education level, programme orientation, completion, sex and age. Retrieved from https://bit.ly/3yg5kXf

IBGE. (2021). Sistema IBGE de Recuperação Automática. SIDRA. Brasília, DF. Retrieved from https://sidra.ibge.gov.br/

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

Latindex. (2021). Latindex Directory. Retrieved from https://www.latindex.org/

OECD. (2022). Purchasing power parities (PPP) (indicator). Retrieved from http://doi.org/10.1787/1290ee5a-en

RedALyC. (2021). Sistema de Información Científica RedALyC: Colección de Revistas. Retrieved from https://www.redalyc.org

SciELO. (2021). SciELO Analytics. Retrieved from https://analytics.scielo.org

US Census Bureau. (2020). Educational attainment in the United States: 2019. Retrieved from https://bit.ly/3dUYh9a

Legislation

BRASIL. (1891). *Constituição da República Federativa do Brasil*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1893). *Lei nº 173, de 10 de setembro de 1893. Regulamenta a organização de associações que se fundarem para fins religiosos, Moraes, scientíficos, artísticos, políticos ou de simples recreio, nos termos do art. 72 da Constituição*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1915). *Decreto nº 11.530, de 18 de Março de 1915. Reorganiza o ensino secundario e o superior na Republica*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1920). *Decreto nº 14.343, de 07 de setembro de 1920. Institue a Universidade do Rio de Janeiro*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1931a). *Decreto nº 19.850, de 11 de Abril de 1931. Exposição de motivos*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1931b). *Decreto nº 19.851, de 11 de Abril de 1931. Dispõe que o ensino superior no Brasil obedecerá o sistema universitário*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1931c). *Decreto nº 19.852, de 11 de Abril de 1931. Dispõe sobre a organização da Universidade do Rio de Janeiro*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1951). *Decreto nº 29.741, de 11 de julho de 1951. Institui uma comissão para promover a Campanha Nacional de Aperfeiçoamento de Pessoal de Nível Superior*. Diário Oficial da União. Rio de Janeiro, RJ.

BRASIL. (1961). *Lei nº 4024, de 20 de dezembro de 1961. Fixa as diretrizes e bases da educação nacional*. Diário Oficial da União. Brasília, DF.

BRASIL. (1974a). *Decreto nº 73.411, de 4 de janeiro de 1974. Institui o Conselho Nacional de Pós-Graduação e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1974b). *Decreto nº 74.299, de 18 de julho de 1974. Dispõe sobre a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1981). *Decreto nº 86.791, de 28 de dezembro de 1981. Extingue o Conselho Nacional de Pós-Graduação e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1988). *Constituição da República Federativa do Brasil*. Senado Federal do Brasil. Brasília, DF.

BRASIL. (1990a). *Lei nº 8028, de 12 de abril de 1990. Dispõe sobre a organização da Presidência da República e dos Ministérios, e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1990b). *Medida Provisória nº 150, de 15 de março de 1990. Dispõe sobre a organização da Presidência da República e dos Ministérios e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (1992). *Lei nº 8405, de 9 de janeiro de 1992. Autoriza o Poder Executivo a instituir como fundação pública a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e dá outras providências*. Diário Oficial da União. Brasília, DF.

BRASIL. (2007). *Decreto nº 6316, de 20 de dezembro de 2007. Aprova o Estatuto da CAPES* (Decree No. 6316/2007). Diário Oficial da União. Brasília, DF. Retrieved May 13, 2023, from https://bit.ly/42oR4Z3

BRASIL. (2014). *Lei nº 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação*. Diário Oficial da União. Brasília, DF.

BRASIL. Ministério da Educação. (2009). *Portaria Normativa nº 17, de 28 de dezembro de 2009. Dispõe sobre o mestrado profissional no âmbito da CAPES (Revogada pela Portaria nº 389/2017)* (Ordinance No. 17/2009). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3H4FEyg

BRASIL. Ministério da Educação. (2017). *Portaria nº 389, de 23 de março de 2017. Dispõe sobre o mestrado e doutorado profissional no âmbito da pós-graduação stricto sensu*. Diário Oficial da União. Brasília, DF.

CAPES. (1995a). *Portaria nº 47, de 17 de outubro de 1995. Determina a implantação, na CAPES, de procedimentos apropriados à recomendação, acompanhamento e avaliação de cursos de mestrado dirigidos à formação profissional* (Ordinance No. 47/1995). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/33CLf0V

CAPES. (1998). *Portaria nº 80, de 16 de dezembro de 1998. Dispõe sobre o reconhecimento dos mestrados profissionais e dá outras providências*. Diário Oficial da União. Brasília, DF.

CAPES. (2011). *Portaria nº 83, de 6 de Junho de 2011. Cria 4 áreas de avaliação: Biodiversidade, Ciências Ambientais, Ensino e Nutrição.* Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/383xcmX

CAPES. (2014). *Resolução nº 5, de 11 de Dezembro de 2014. Altera a periodicidade da avaliação do SNPG* (Resolution No. 5/2014). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3BK5Ywu

CAPES. (2016a). *Portaria nº 141, de 14 de setembro de 2016. Define e disciplina as formas de colaboração e os procedimentos de escolha dos consultores científicos* (Ordinance No. 141/2016). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3v5NG7B

CAPES. (2016b). *Portaria nº 234, de 15 de Dezembro de 2016. Alteração de denominação de áreas de avaliação.* Diário Oficial da União. Brasília, DF.

CAPES. (2017b). *Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal*. Diário Oficial da União. Brasília, DF.

CAPES. (2017c). *Portaria nº 60, de 20 de março de 2019. Dispõe sobre o mestrado e doutorado profissionais* (Ordinance No. 60/2019). Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/3BGiXzo

CAPES. (2018d). *Portaria nº 150, de 4 de julho de 2018. Institui o Grupo de Trabalho (GT) do Qualis Periódicos*. Diário Oficial da União. Brasília, DF.

CAPES. (2018e). *Portaria nº 271, de 14 de dezembro de 2018. Institui o GT do Qualis Periódicos das Humanidades* (Ordinance No. 271/2018). Diário Oficial da União. Brasília, DF. Retrieved April 21, 2023, from https://bit.ly/3mW7bhh

CAPES. (2019c). *Ordinance nº 8, de 31 de janeiro de 2019. Institui o GT do Qualis Periódicos das Ciências da Vida e Ciências Exatas, Tecnológicas e Multidisciplinar* (Ordinance No. 8/2019). Diário Oficial da União. Brasília, DF. Retrieved April 21, 2023, from https://bit.ly/3mUUvXY

CAPES. (2021d). *Portaria nº 122, de 5 de Agosto de 2021. Consolida os parâmetros e os procedimentos gerais da Avaliação Quadrienal de Permanência da pós-graduação stricto sensu no Brasil*. Diário Oficial da União. Brasília, DF. Retrieved from https://bit.ly/384xDgE

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*. Brasília, DF.

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

Ministerie van Onderwijs, Cultuur en Wetenschap. (1992). *Wet op het hoger onderwijs en wetenschappelijk onderzoek*. OCW. Den Haag. Retrieved from https://bit.ly/33YmT1l

Author Index

A

Expanding GeoCapes usefulness for research

The majority of the science and technology research conducted in Brazil takes place within a national system of research and graduate education (SNPG) (SBPC and ABC, 2020b). This system has been designed as the object of public policy from the 1950s, through a series of actions such as the launch of a national campaign to prepare highly qualified personnel (BRASIL, 1951) and then through the Sucupira Report, a document that shaped the SNPG into a model close to the one that exists today (CFE, 1965). Since its conception, an evaluation system has been an integral part of the SNPG, with a periodic evaluation of graduate programs in continuous operation since the 1970s (Brasil, 2020).

The original campaign in charge of training higher education personnel became the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES), today a public foundation that has been in charge of evaluating the SNPG for nearly five decades (Brasil, 2020). To perform the central assessment of graduate education, CAPES has consistently implemented strategies to collect data from graduate programs (PPG) and higher education institutions (HEI). The first time the agency conducted its collection was in 1975, and data from the first few years served as evidence for the first evaluation of graduate programs in the country, which took place in 1978 (Siqueira, 2019).

The initially paper-based data collection was replaced in 1987, when CAPES launched its first information system for that purpose: DataCapes. According

This chapter is under review as: Brasil, A. (2023). Expanding GeoCapes usefulness for research: Introducing the GeoCapes R package.

to R. J. Ribeiro (2008), from the collection of annual information on the performance of all graduate programs in Brazil, DataCapes held one of the most detailed databanks of higher education in the world, with available information that caters to the evaluation needs of all fields of knowledge.

Over the next few years, CAPES' information systems followed the technological evolution, leading to three iterations of a new system designed to gather data from graduate programs. The new "Coleta CAPES" was operational from 1996 to 2013, moving from an initial offline system to an online one, eventually connecting to other databases such as the national Lattes curriculum. From 2013, the system was replaced by the more robust Sucupira Platform, which has been the subject of several studies over the years, including Maciel et al. (2019), Siqueira (2019), and Maia (2020).

Information systems designed to collect data from the SNPG have evolved considerably over time, providing evaluation committees with unprecedented information to assess the quality of research in Brazil. Despite that, for many years, "Coleta CAPES" was considered a black box, as its data was made available mainly to the evaluation committees. Neither society nor graduate programs could access the information supporting the evaluations. That started to change in July 2009, with the launch of CAPES' Georeferenced Information System (GeoCapes). This system was designed to bring the public a series of datasets around graduate education, presented with various levels of aggregation, and included a dashboard to explore the available information from different perspectives (CAPES, 2009; Cardoso, 2015; Maia, 2020).

From its launch, GeoCapes has been valuable for the accountability of invested resources and as a tool for interested users to understand and analyse Brazilian graduate education. Today, the system remains relevant, including around two decades of data on graduate programs, student body, faculty members, international cooperation, research funding, and even institutional access to scientific publications (from the Brazilian Portal of Journals). Nevertheless, GeoCapes presents a series of limitations when users need to perform more than fundamental analyses (Cardoso, 2015; Maia, 2020).

This paper introduces a GeoCapes R package developed to address some of the limitations seen in the original system. These include the need for further data standardisation and cleaning, the lack of tools for proper longitudinal analyses, and the need to enrich datasets with information to broaden potential research.

Additionally, the presented R package also addresses the language barrier of GeoCapes data, offering a curated translation of necessary content into English, making the datasets ready for international research as well.

A.1 GeoCapes benefits and limitations

CAPES introduces its GeoCapes system as a georeferenced tool with information on graduate education. Accessible at http://geocapes.capes.gov.br, the interface offers geographical and analytic visualisations of data for select indicators, including choropleth maps for a distribution perspective. Figure A.1 shows GeoCapes main page, displaying a map with the distribution of graduate programs in 2020. On the right side, visualisations provide information by broad area, course levels offered, and type of institution. Map selections filter the information displayed on the rest of the dashboard (CAPES, 2021e).

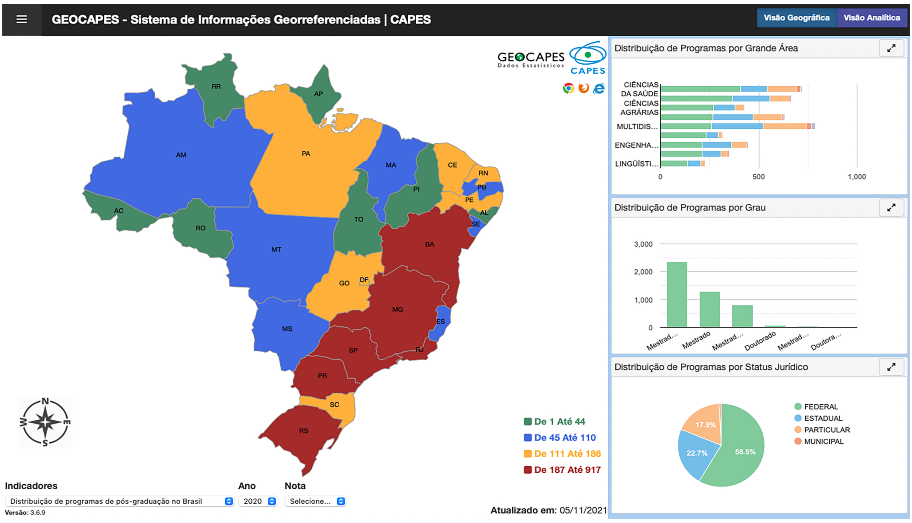
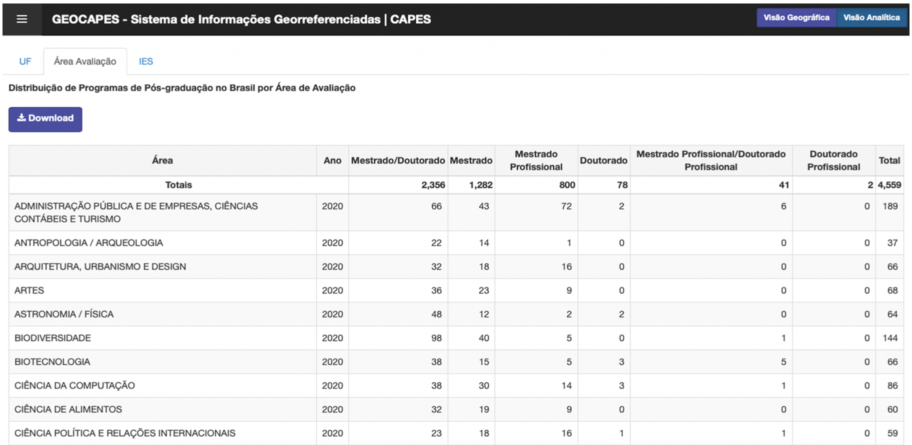


Figure A.1.: Geographical view of the GeoCapes system, displaying the distribution of graduate programs in Brazil in 2020

Figure A.2 shows GeoCapes analytical view, where the dashboard's underlying data is shown in a table format. For the selected dataset, three levels of aggregation are offered: state, evaluation area, and higher education institution (HEI).

To explore the panel, users can choose different years and even download data according to the desired level of aggregation.



GEOCAPES - Sistema de Informações Georreferenciadas | CAPES

Visão Geográfica Visão Analítica

UF Área Avaliação IES

Distribuição de Programas de Pós-graduação no Brasil por Área de Avaliação

Download

Área	Ano	Mestrado/Doutorado	Mestrado	Mestrado Profissional	Doutorado	Mestrado Profissional/Doutorado Profissional	Doutorado Profissional	Total
Totais		2,356	1,282	800	78	41	2	4,559
ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO	2020	66	43	72	2	6	0	189
ANTROPOLOGIA / ARQUEOLOGIA	2020	22	14	1	0	0	0	37
ARQUITETURA, URBANISMO E DESIGN	2020	32	18	16	0	0	0	66
ARTES	2020	36	23	9	0	0	0	68
ASTRONOMIA / FÍSICA	2020	48	12	2	2	0	0	64
BIODIVERSIDADE	2020	98	40	5	0	1	0	144
BIOTECNOLOGIA	2020	38	15	5	3	5	0	66
CIÊNCIA DA COMPUTAÇÃO	2020	38	30	14	3	1	0	86
CIÊNCIA DE ALIMENTOS	2020	32	19	9	0	0	0	60
CIÊNCIA POLÍTICA E RELAÇÕES INTERNACIONAIS	2020	23	18	16	1	1	0	59

Figure A.2.: Analytical view of the GeoCapes system, displaying the distribution of graduate programs in Brazil aggregated by evaluation area

In addition to the distribution of graduate programs, GeoCapes includes six other datasets, all listed in Table A.1 alongside the respective data coverage and aggregation levels present in the analytical view. The availability of such diverse data sets contributes to the transparency of the Brazilian graduate system and serves as an essential source of information for higher education institutions, students, researchers, and policymakers.

Table A.1.: Datasets available in GeoCapes, including coverage period and available aggregation levels

Datasets	Coverage	State	Countries	Broad Area	Area	HEI
Distribution of graduate programs	1998–2020	✓			✓	✓
Distribution of faculty members	1998–2020	✓			✓	✓
Distribution of graduate students	1998–2020	✓			✓	✓
CAPES scholarship holders in Brazil	1995–2020	✓		✓	✓	✓
CAPES scholarship holders abroad	1998–2020		✓		✓	
CAPES investment in scholarships and funding	2002–2020	✓				
Access data to the Portal of Journals	2001–2020	✓				✓

The levels of aggregation provided by GeoCapes reveal limitations that could impact research interests. For instance, the Brazilian graduate system classifies its programs into 49 evaluation areas, grouped into nine broad areas and three upper groups. As an example, "Economics" is one of the evaluation areas in the broad area of "Applied Social Sciences", which is part of the "Humanities" upper group. GeoCapes lists evaluation areas for five datasets, but broad areas are only available for domestic scholarship holders. Therefore, if a researcher hopes to analyse graduate program growth in relation to scholarships granted in the upper group of "Life Sciences", GeoCapes' interface is not enough. For that study, it would be necessary to download the underlying datasets, enriching them with information on broad and upper groups of areas.

Since most datasets include the microlevel classification of evaluation areas, enriching the data would be straightforward if GeoCapes provided unique identifiers for the 49 areas. However, since that is not the case, users need to rely on the area names, which are not standardised. For instance, the graduate programs dataset includes 106 distinct evaluation areas, and most are duplicates from either trailing and inconsistent spacing or capitalisation issues. While cleaning the data could solve these, another problem comes from renaming evaluation areas over the years. The current area of "Journalism and Information", for example, was known as "Applied Social Sciences I" until it was rebranded, alongside two other areas, by Ordinance 234 (CAPES, 2016b).

Similar inconsistencies are seen across higher education institutions, cities, countries, and other fields. Those problems can be partially explained by the fact that GeoCapes does not seem to be based on a relational database, compiling visualisations from a structured source. Instead, the underlying information appears to be built from appending new aggregated data every year, imposing limitations that need to be addressed to conduct rigorous research.

A.2 Introducing the GeoCapes R package

R is an open-source language and environment for statistical computing and graphics, popular among researchers and data analysts across multiple fields of knowledge. The language's algorithms are powerful to work with large volumes of data, automating the tedious and error-prone manual work associated with spreadsheet use while also being shareable in public repositories, thus fostering openness and reproducibility (R Core Team, 2021; Sarmento and Costa, 2017).

The benefits of the R language make it valuable to deal with GeoCapes data, in part because the system's datasets are updated annually, but not simultaneously. Algorithms can seamlessly download and process the most recent data, saving users from handling new tables for every study. Additionally, the approach is more suitable to address new research questions that may arise from the data, as the code can be adapted and expanded as needed. Of course, the code complexity grows with the design of multiple usage scenarios, limiting its use to those users proficient in the chosen language. For that reason, this paper introduces an R package compiling a series of algorithms designed to work with GeoCapes data. In a package, thousands of lines of code are transformed into simple and easy to use functions, significantly improving accessibility.

To demonstrate its simplicity, Algorithm A.1 shows the commands to install and load GeoCapes for R; download and process the most recent version of the graduate program dataset; export the cleaned and enriched results to a CSV file; and even produce visualisations from the data. To use the code, R should be installed on the user's computer, and the software download with necessary instructions is available at https://www.r-project.org/. It is also recommended to install RStudio Desktop, an integrated development environment (IDE) that enhances the use of many R features (https://www.rstudio.com/).

Algorithm A.1: GeoCapes R package: installing, loading and operation example

```
# Install and load the latest version of the GeoCapes R Package
install("devtools")
devtools::install_github("AndreBrasil/geocapes")
library(geocapes)

# Download and clean the GeoCapes dataset of graduate programs
get_grad_programs()

# Exporting the dataset as a local file
save_grad_programs()

# Plots from the processed dataset
prog_year()
prog_hei()
```

The resulting data set from running the `get_grad_programs()` function is a table with 65.599 observations[1] of 16 variables, including the 13 originally

[1] The information considers the dataset made available in November, 2021.

provided by GeoCapes and three additional ones: evaluation area id, HEI id, and upper area groups. Cleaning algorithms are applied to fields such as city, HEI and institution type. A series of arguments are included in the download function to filter the dataset: *start year*, *end year*, *state*, *region*, and *HEI*. From that, `get_grad_programs(2010, 2015, region = "North")` retrieves records from 2010 to 2015, from the North region of Brazil, while `get_grad_programs(state_code = "SP")` limits the resulting table to PPG in the state of São Paulo. As the package is thoroughly documented, help files with descriptions, arguments, and examples are available for every function.

After downloading the chosen data set, the `save_grad_programs()` function can be used to export the resulting table as a CSV file. Those familiar with R can continue using the environment for data analyses, and the package also includes ready-made functions for that. Figure A.3, for instance, shows the plot created with the `prog_year()` function, which consists of a line graph of the number of graduate programs active per year in the 1998–2020 period.

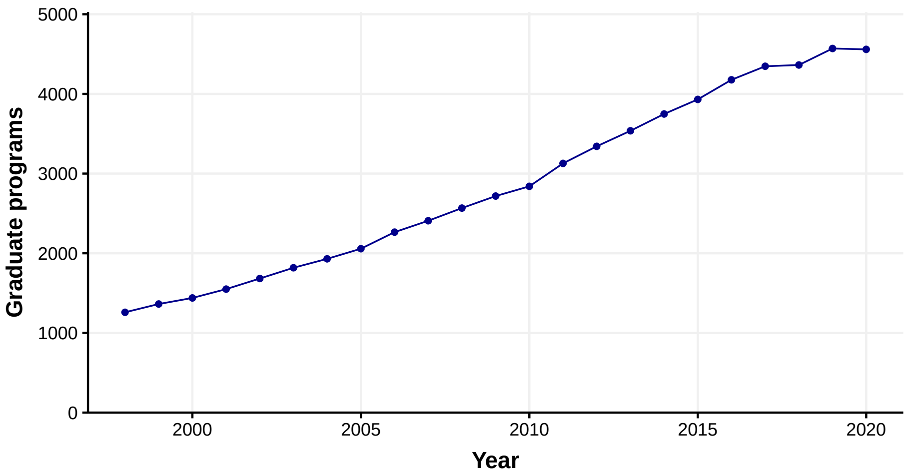


Figure A.3.: Line chart created by `prog_year()`, showing graduate programs per year

As happens with the download function, `prog_year()` includes arguments to filter the starting and end years displayed, but it can also detail the plotted line according to different variables. Figure A.4, for instance, shows the output of `prog_year(per_region = TRUE)`, a line chart of PPG revealing the asymmetrical distribution of graduate education across the five Brazilian regions.

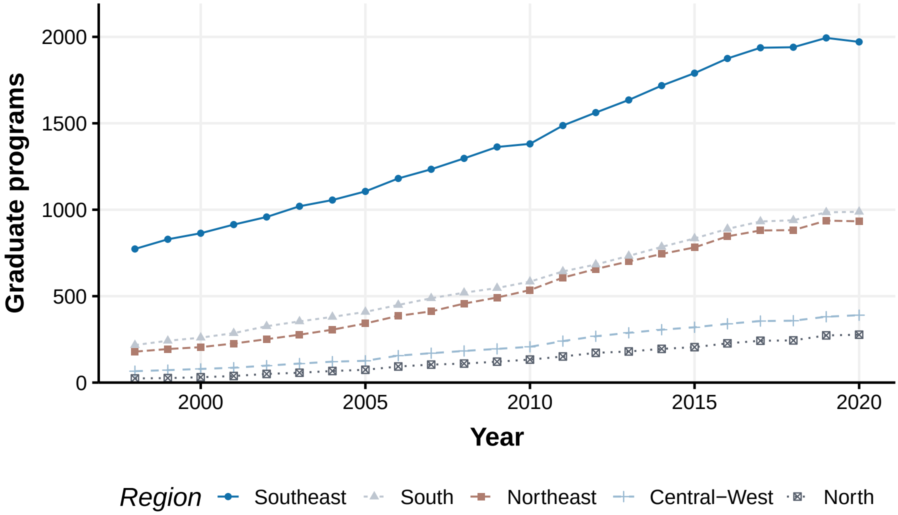


Figure A.4.: Line chart created by `prog_year(per_region = TRUE)`, showing graduate programs per year and region

Another example is seen in Figure A.5, which displays the result of running `prog_year(per_region = TRUE, per_level = TRUE)`. This command outputs a line chart for each region, detailing the number of programs according to the level of courses offered. The legend shows master's degrees (MA), doctoral degrees (DO), and the respective professional categories (MP and DP).

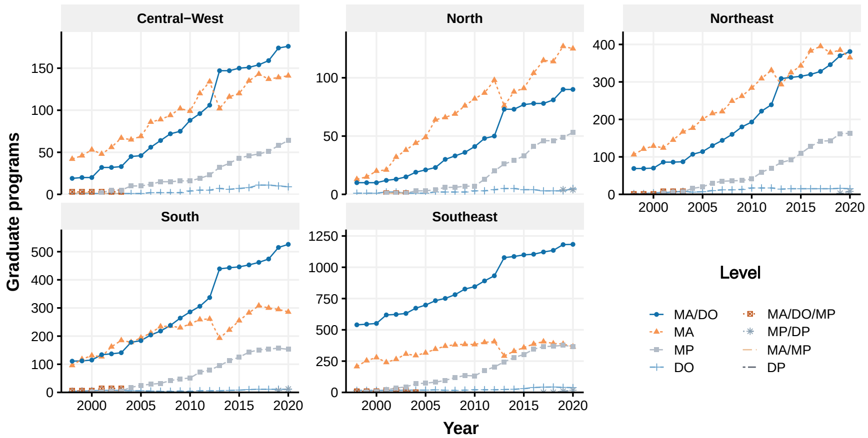


Figure A.5.: Line chart created by `prog_year(per_region = TRUE, per_level = TRUE)`, showing graduate programs per year and level, grouped by region

Figure A.5 reveals how graduate programs offering academic courses at the master's and doctoral levels have been the majority for decades in the Southeast of Brazil, but a more recent development in the South and Central West regions. In the case of the Northeast and North regions, the proportion of programs offering only the academic master's is still significant. The data support the arguments presented in Brasil (2020) for the valorisation of the master's course in its current format in the country – resembling a short doctorate – as it is still the primary alternative for scientific training in a substantial part of Brazil.

Still working with the PPG dataset, the GeoCapes package for R offers the `prog_hei()` function, which analyses the data from the perspective of HEI. The bar graph produced can be seen in Figure A.6, where the top 15 HEI in the number of graduate programs are shown (using the most recent year available in the dataset). Each bar is detailed by grade, which starts at *2* for this set of institutions and goes up to the maximum of *7*. Grade *A* represents programs that were recently accredited, but have not yet gone through a periodic evaluation. The available arguments allow the selection of reference years and the number of institutions to be shown, as well as the filtering of HEI from selected regions. Users are also allowed to visualise the results as a percentage stacked bar chart.

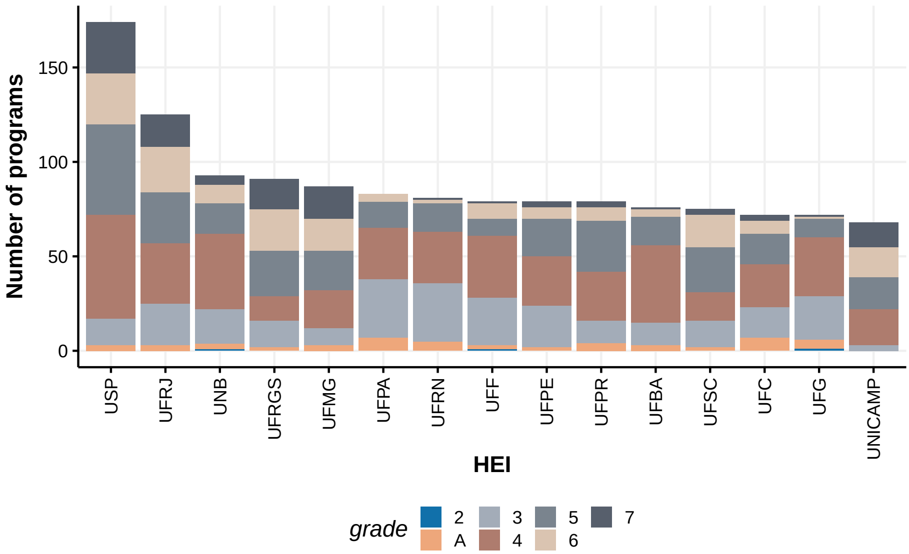


Figure A.6.: Top 15 HEI in number of graduate programs in Brazil, detailed by grade

A different example of the GeoCapes R package application comes from the function `get_funding()`, which downloads data on annual CAPES investment per state. While relevant, the original dataset has no information on region, and expenditures are only provided in Brazilian real (BRL), without any monetary correction over the years. As Brazil is not a country free of inflation, deflating investments would improve comparisson capabilities and converting values to US dollars (USD) would make the dataset useful for an international audience. With that in mind, internal functions were embedded into `get_funding()`, adding the following columns to the resulting dataset:

i) Brazilian geographical regions;

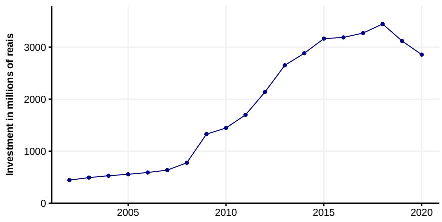
ii) The corresponding US dollar exchange rate for each record, obtained through the Application Programming Interface (API) of the Brazilian Central Bank;

iii) The informed investment converted to the corresponding dollar rate;

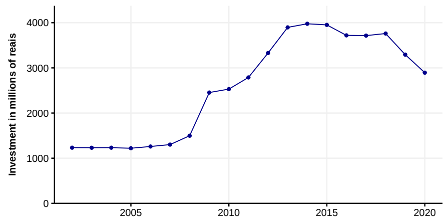
iv) Deflated values in BRL, corrected using the Extended National Consumer Price Index (IPCA), which is collected through the API of the Brazilian Institute of Applied Economic Research (IPEA).

As with any of the download functions provided by the package, the resulting data set can be saved as an Excel compatible CSV file, in this case by running `save_funding()`. Furthermore, a visualisation function is available – `funding_year()` – which can be used to plot a line graph of CAPES investment over time, as can be seen on Figure A.7. Arguments available allow users to select the start and end years plotted, choose the desired currency (either BRL or USD), and whether the monetary correction should be applied for investments in Brazilian real. It is also possible to plot the data according to region, as shown on Figure A.7d.

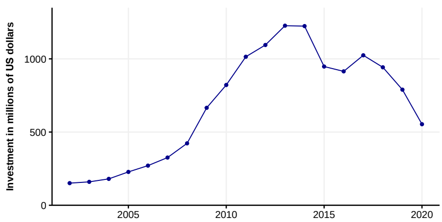
The GeoCapes funding dataset is quite valuable for observing the rise and fall in investments made by CAPES over nearly two decades. It helps researchers and general users identify the asymmetry in funding distribution, see the impact of devaluation of the Brazilian real on the dollar, and more. Combining funding information with data on scholarship holders in Brazil or abroad, growth of the graduate system, and details on students makes GeoCapes an even more powerful resource for answering relevant research questions about the Brazilian National System of Graduate Education.



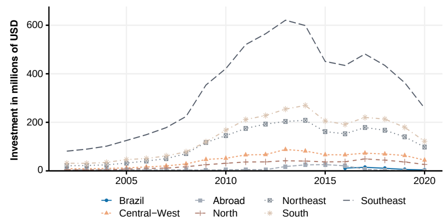
(a) `funding_year(currency = "real")`



(b) `funding_year(deflate = TRUE)`



(c) `funding_year(currency = "dollar")`



(d) `funding_year(currency = "dollar", per_region = TRUE)`

Figure A.7.: Line charts generated by the `funding_year()` function

A.3 Available functions in the R package

The initial release of the GeoCapes for R package includes more than 60 functions, of which 50 are available to the user. The remaining are internal functions created for tasks such as monetary correction, defining visualisation templates, and dataset translation. They are activated when necessary, without the need for user intervention, and the package documentation includes details about their operation.

The user-accessible functions are presented in two sets: one in Portuguese and the other in English. Both download, clean, and enrich datasets; save results to CSV files; and provide visual output for simple analyses. The only difference is that the first set does all that in the original language of the datasets, while the second applies a series of algorithms to translate column names and content such as countries, regions, evaluation areas, and groups, etc.

The GeoCapes R package does not use machine-generated translations. Instead, it relies on lookup tables curated from the author's previous research and international publications. For instance, the translation applied to evaluation areas, broad areas, and groups resulted from analysing the classification adopted by CAPES against those of the Organisation for Economic Co-operation and Development (OECD) and the International Standard Classification of Education (ISCED). The corresponding lookup table has been made available on a public repository (Brasil, 2021c).

The following subsections will present the user-accessible functions included in version 1.0 of the GeoCapes R package.

A.3.1 Downloading

The GeoCapes R package can download all the data sets listed in Table A.1. The available functions are shown in Table A.2, focussing on the English versions, but showing the Portuguese alternatives below them. Running any of those functions will make the corresponding dataset available in the R environment, cleaned and enriched. Arguments are available to allow users to pre-filter the original datasets to fit particular needs. For example, all functions accept the `start_year` and `end_year` parameters, while others may be used to get information for a single higher education institution or for a region or state.

Table A.2.: GeoCapes R package functions to download, clean, and enrich datasets

Function	Description
`get_grad_programs()` `baixar_ppg()`	Retrieves the complete dataset of active graduate programs (PPG) in Brazil, with information per year and according to HEI, city, evaluation area, etc. The full enriched table includes 65.599 records of 16 variables, which can be filtered by *start year*, *end year*, *state*, *region*, and *HEI*.
`get_faculty()` `baixar_docentes()`	Downloads the dataset with information on the distribution of faculty members aggregated by PPG, and detailed by affiliation type (permanent, contributor or visiting professor). The full dataset includes 77.655 observations of 19 variables, and the same filters as above apply.
`get_students()` `baixar_discentes()`	Similar to the previous function, the final table includes the number of graduate students per year and level across each individual PPG. The full dataset includes 137.820 observations of 20 variables. The same five filters remain available.

Continue...

Table A.2 Continued

Function	Description
`get_grants()` `baixar_bolsas()`	Downloads the dataset with grants awarded to each PPG, detailling level (masters, doctorate, postdoctorate, etc.) and funding program. The complete data set includes 154.916 observations of 28 variables. While the filters mentioned above also apply, users have the additional option to filter by *funding line*.
`get_international()` `baixar_internacional()`	Retrieves the dataset of scholarships granted for studies abroad, aggregated according to foreign institution and funding program. The resulting data set has 117.914 observations of 21 variables. Available filters are: *start year*, *end year*, *country*, *funding program*, and *HEI (foreign)*.
`get_funding()` `baixar_fomento()`	This function gets the table with information on CAPES' funding per year and state, enriching the data with columns on monetary correction (IPCA), currency convertion (from BRL to USD) and region. The resulting table contains only 538 observations of 8 variables, but can still be filtered by *start year*, *end year*, *state*, and *region*.
`get_pub_portal()` `baixar_portal()`	Retrieves information regarding usage of the Brazilian Portal of Journals, including access to reference databases and the number of full-text downloads, aggregated by HEI. The data set of 5603 observations of 11 variables can be filtered by *start year*, *end year*, *state*, *region*, and *HEI*.
`get_geocapes()` `baixar_geocapes()`	This function downloads all GeoCapes datasets, running the respective algorithms to clean and enrich them. All the filters mentioned above can be used here.

A.3.2 Saving

After running any of the download functions, the corresponding datasets will be made available in R's global environment. These can be explored by users – who can further clean, enrich, and analyse the tables – or just be exported as CSV files. Table A.3 lists the available functions to save data sets.

Table A.3.: GeoCapes R package functions to export downloaded datasets

Function	Description
`save_grad_programs()` `salvar_ppg()`	Exports the complete dataset of active graduate programs in Brazil as `geocapes_grad_prog.csv`.
`save_faculty()` `salvar_docentes()`	Exports the dataset on the distribution of faculty members per PPG as `geocapes_faculty.csv`.

Continue. . .

Table A.3 Continued

Function	Description
`save_students()` `salvar_discentes()`	Exports the dataset on the distribution of graduate students per PPG as `geocapes_students.csv`.
`save_grants()` `salvar_bolsas()`	Exports the dataset on CAPES scholarship holders in Brazil as `geocapes_grants.csv`.
`save_international()` `salvar_internacional()`	Exports the dataset on CAPES scholarship holders abroad as `geocapes_international.csv`.
`save_funding()` `salvar_fomento()`	Exports the dataset on CAPES investment in scholarships and funding as `geocapes_funding.csv`.
`save_pub_portal()` `salvar_portal()`	Exports the dataset on the use of the Brazilian Portal of Journals as `geocapes_pub_portal.csv`.
`save_geocapes()` `salvar_geocapes()`	Checks which datasets were downloaded by the user and runs the respective functions to export everything at once.

Each of the functions presented in Table A.3 will perform all saving operations on the datasets locally available. Thus, any filters applied during the download process will be reflected in the exported files. To represent the content of each CSV file, users can include a different filename to the functions. For instance, `save_international("geocapes_international_USA.csv")` can better reflect the output of scholarship holders if the download was filtered to show only grants awarded for studying or conducting research in the United States.

A.3.3 Visualising

The GeoCapes R package includes a series of functions to help users understand and visualise the downloaded data. Table A.4 lists those available in version 1.0, including the arguments users can apply to adjust the visualisations to reflect their needs.

Table A.4.: GeoCapes R package functions to visualise downloaded datasets

Function	Description
`prog_year()` `ppg_por_ano()`	Creates a line chart on the number of graduate courses active in Brazil, as seen on figures A.3 to A.5. Users can define *start year* and *end year* to show and choose to plot the lines by region and level.

Continue...

Table A.4 Continued

Function	Description
`prog_hei()` `ppg_por_ies()`	Creates a stacked bar chart on the distribution of graduate programs per HEI, detailed according to grade (Figure A.6). By default, the function plots the most recent data for the top 15 institutions in the number of PPG, but users can choose to visualise previous years and change the number of HEI to display. It is also possible to filter the data to show one or more regions and to output the results as a percent bar chart.
`faculty_year()` `docentes_ano()`	Creates a stacked bar chart of the number of faculty members over time, detailed by affiliation type. Users can change the range of the data displayed, and filter by region, state, or HEI.
`enrolled_year()` `matriculados_ano()`	This function creates a stacked bar chart similar to the previous one, but plots the number of enrolled students, detailed by level. All filters from `faculty_year()` apply.
`graduates_year()` `graduados_ano()`	Also similar to the previous function, including the available filters, but producing a stacked bar chart of the number of students graduated each year, again detailed by level.
`save_grants()` `salvar_bolsas()`	Creates a line chart on the number of scholarship holders in Brazil. The visualisation can be filtered by region, state and HEI, and users can select the desired range to display.
`international_year()` `internacional_ano()`	Creates a line chart on scholarship holders abroad, detailed for the top 10 countries. Users can select the period to be shown (defining *start year* and *end year*), change the number of countries to display, or even choose the ones to be shown. Two boolean filters are also available to exclude either undergraduate scholarships or the Science Without Borders program from the visualisation.
`funding_year()` `invest_ano()`	Creates a line chart on the evolution of CAPES funding over time, as seen in Figure A.7. Users can visualise investments in USD or BRL (nominal or deflated) and can detail the results by region. It is also possible to select the years to display.
`pub_portal_year()` `portal_ano()`	Creates a stacked bar chart on the annual use of the Brazilian Portal of Journals, detailed by type of access (database and full paper download). Once again, users can select the desired range to display, as well as filter by region, state and HEI.

For experienced R users, it is worth mentioning that the functions listed in Table A.4 rely on the ggplot2 data visualisation package designed by Wickham (2016), so the resulting objects can be further manipulated. For instance, running `prog_year() + geom_smooth(method = "lm")` will add a linear model to the original plot. Likewise, wrapping any of the visualisation functions in `ggplotly()` – from Plotly Technologies (2015) – will convert the figures into interactive ones, ready for embedding into web applications.

A.4 Conclusion and next steps

The launch of CAPES' Georeferenced Information System (GeoCapes) was a significant step towards a transparent system of research and graduate education in Brazil. Opening data on graduate programs, faculty members, student body, and different funding initiatives has made the system very valuable over the years, both for accountability and research purposes. However, significant limitations exist both in the system's public interface and in the underlying data, including language restrictions imposed by its Portuguese only approach. Therefore, this paper introduced a GeoCapes R package designed to share a series of algorithms developed to increase the value of CAPES' system for Brazilian researchers and an international audience.

In its initial release, the GeoCapes R package includes 25 user-accessible functions to deal with the original data in Portuguese and equivalent counterparts to perform the same tasks with additional translation into English. These functions download, clean, enrich, export, and plot data from GeoCapes, expanding the system's functionality and usefulness for researchers and other interested users. By providing simple access to complex data analyses, the functions are easy to use and are documented in the corresponding language.

Among the benefits of the GeoCapes for R package is the possibility of performing time series analyses as seen in A.3, A.4, A.5, and A.7. However, several other analyses are already possible with the package, and the arguments included in each function allow users to filter and adjust the various output files and visualisations. In that sense, an important next step in the package development is the expansion of potential usage scenarios, incorporating new analyses both from the author's perspective and from eventual user requests or collaboration.

Although expanding the package for R users may be relevant, it is crucial to acknowledge the steep learning curve associated with any programming language. Furthermore, researchers with a more qualitative profile usually adopt more intuitive statistical tools, such as SPSS. On the other hand, disciplines such as engineering favour the use of Python, in part because it is a general-purpose programming language (Muenchen, 2019). Thus, it is relevant to consider how to expand the scope of the package presented here in order to allow its use by groups of users who may benefit from the results offered, but who would have no interest or need to add R to their research toolset. In this sense, one of the next steps planned for the development of GeoCapes for R is the design of a graphical web interface, based on the Shinny platform created by Chang et al. (2021). In this way, users will be able to operate the package without typing a single line of code.

For now, the open source version of the GeoCapes R package – including its documentation and a space for users to report bugs or send feature requests – is available on the GitHub repository at https://github.com/AndreBrasil/geocapes.

References

BRASIL. (1951). *Decreto nº 29.741, de 11 de julho de 1951. Institui uma comissão para promover a Campanha Nacional de Aperfeiçoamento de Pessoal de Nível Superior*. Diário Oficial da União. Rio de Janeiro, RJ.

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. https://doi.org/10.33871/nupem.2020.12.27.222-253

Brasil, A. (2021c). Brazilian evaluation areas, broad areas and groups, including original names in Portuguese and English translations. Dataset. https://doi.org/10.5281/zenodo.5795662

CAPES. (2009). Capes lança aplicativo gráfico para consulta de dados. Retrieved December 7, 2021, from https://bit.ly/3IsaOBw

CAPES. (2016b). *Portaria nº 234, de 15 de Dezembro de 2016. Alteração de denominação de áreas de avaliação*. Diário Oficial da União. Brasília, DF.

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

Cardoso, M. S. (2015). Sistema Geocapes de dados estatísticos. Brasília, DF. Retrieved December 2021, from https://bit.ly/3DyZmR1

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., . . . Borges, B. (2021). *Shiny: Web application framework for r*. R package version 1.7.1. Retrieved from https://CRAN.R-project.org/package=shiny

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*. Brasília, DF.

Maciel, C. E., Trierweiller, A. C., & Ferenhof, H. A. (2019). Perspectiva do Usuário na Utilização da Plataforma Sucupira como Sistema de Avaliação: Uma Analise Exploratória. (pp. 1965–1975). https://doi.org/10.5151/9cidi-congic-4.0350

Maia, M. A. Q. (2020). *Sistema de informação da pós-graduação brasileira: Avaliação da Plataforma Sucupira* (Doctoral dissertation, Universidade Federal de Minas Gerais).

Muenchen, R. A. (2019). The popularity of data science software. Retrieved December 28, 2021, from http://r4stats.com/articles/popularity/

Plotly Technologies. (2015). Collaborative data science. Montreal, QC: Plotly Technologies Inc. Retrieved from https://plot.ly

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org

Ribeiro, R. J. (2008). Quality in Research: CAPES' point of view. In J. L. N. Audy & M. C. Morosini (Eds.), *Innovation and quality in the university* (pp. 127–169). Porto Alegre: EDIPUCRS.

Sarmento, R., & Costa, V. (2017). *Comparative approaches to using R and Python for statistical data analysis*. Hershey, PA: IGI Global.

SBPC, & ABC. (2020b). Qual o papel da CAPES na construção da ciência e da pesquisa no Brasil? [webinar]. Brasília, DF: SBPCnet. Retrieved August 17, 2021, from https://youtu.be/o1ztaCBihAk

Siqueira, M. B. (2019). Sucupira - A Platform for the Evaluation of Graduate Education in Brazil. *Procedia Computer Science*, *146*(2019), 247–255. https://doi.org/10.1016/j.procs.2019.01.081

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org

Additional Resources

B

This dissertation is grounded in an applied research approach, with the aim of making it relevant and impactful on evaluation processes. It provides a variety of practical suggestions to improve Brazil's graduate evaluation system. Given its complexity, extensive nature, and English language, the work might not be easily accessible to everyone. Thus, a list of deliverables tied to this dissertation's core essence are provided here. This inventory is geared towards a more practical understanding and application of the research.

B.1 Policy brief series

In the realm of academic research, it is of paramount importance to ensure that the results and findings of studies are not limited to the walls of universities or the pages of scholarly publications. It is equally crucial to see these results translated into actionable insights, specifically tailored for policymakers who wield the power to promote change. To this end, a suite of policy briefs has been prepared to communicate some of the key recommendations derived from the various chapters of this dissertation.

Figure B.1 showcases the cover of the first policy brief in the series, developed from insights included in Chapter 9. This chapter advocates for the use of purchasing power parity (PPP) as a preferred index to negotiate transformative agreements for open-access publishing in Brazil. Other policy briefs within the collection available at https://doi.org/10.17605/osf.io/8jsvu explore a myriad of topics. These include, but are not limited to, the role of evaluation in the recognition of open-access publishing, and recommendations regarding the reconfiguration of the disciplinary classification currently upheld by CAPES.

POLICY BRIEF #01

CAPES

CWTS
Meaningful metrics

PRÁTICAS DE NEGOCIAÇÃO DE APCS:

A PARIDADE DO PODER DE COMPRA COMO PARÂMETRO MAIS EQUITATIVO

Sumário Executivo

Este documento sugere que a CAPES utilize o índice de Paridade do Poder de Compra (PPP) como parâmetro normalizador da estrutura econômica brasileira ao negociar acordos de transformação com editoras científicas. As metodologias adotadas por muitas editoras para concessão de waivers de publicação empregam o modelo Research4Life ou a classificação do Banco Mundial com base no PIB. Ambas negligenciam as consideráveis disparidades econômicas nacionais, exigindo uma alternativa mais equitativa. Este resumo apresenta evidências de que o PPP pode ser usado como possível alternativa. No entanto, sua apresentação não descarta o valor de consultas com especialistas econômicos para explorar alternativas mais adequadas ao cenário de negociação.

Figure B.1.: Cover page of one of the policy briefs being prepared for CAPES

B.2 Website: BrScience

In conjunction with the outcomes delineated within this dissertation, a website has been designed to offer an extensive scientometric analysis of the strengths, weaknesses, opportunities, and threats of the Brazilian science system. The BrScience website (https://brscience.cwts.nl) includes information on scientific output, citation impact, international and industry collaboration, and more. The core data for the initial development came from the Web of Science, covering the period 2013-2022. However, additional databases have been subsequently added and the presented analyses now include information from Scopus, OpenAlex, and other relevant sources.

The primary objective of the effort is to offer an overview of the Brazilian scientific landscape from a scientometric perspective, while also providing policymakers with evidence for the effective allocation of investments and the recognition of past efforts. The development of this website is taking place in consultation with CAPES, and the initial single author approach for the development has been opened to include collaborations from many other researchers interested in sharing their analyses on the Brazilian scientific landscape.

Figure B.2.: BrScience website, available at https://brscience.cwts.nl

B.3 Dashboard: Literature explorer

During this doctoral research, an exhaustive study was undertaken, encompassing the examination of close to 1500 research items. This large collection included articles, books, chapters, reports, and legislation, among other types of document. To facilitate the exploration of the cited works, each individual chapter in this dissertation includes a dedicated reference section. Furthermore, the end of the book also features a consolidated reference list of all cited works, followed by a complete author index.

However, for those who wish to gain a more detailed perspective of the entire literary landscape explored, an interactive dashboard has been created and made available at https://andrebrasil.github.io/viz/references.html. This resource enables users to dive into all the reviewed works, cited or not, and provides direct links to any content with a DOI or a permanent URL. The interface also offers multiple filter capabilities, including by author, publication type, language, and by the chapters in which the works were cited.

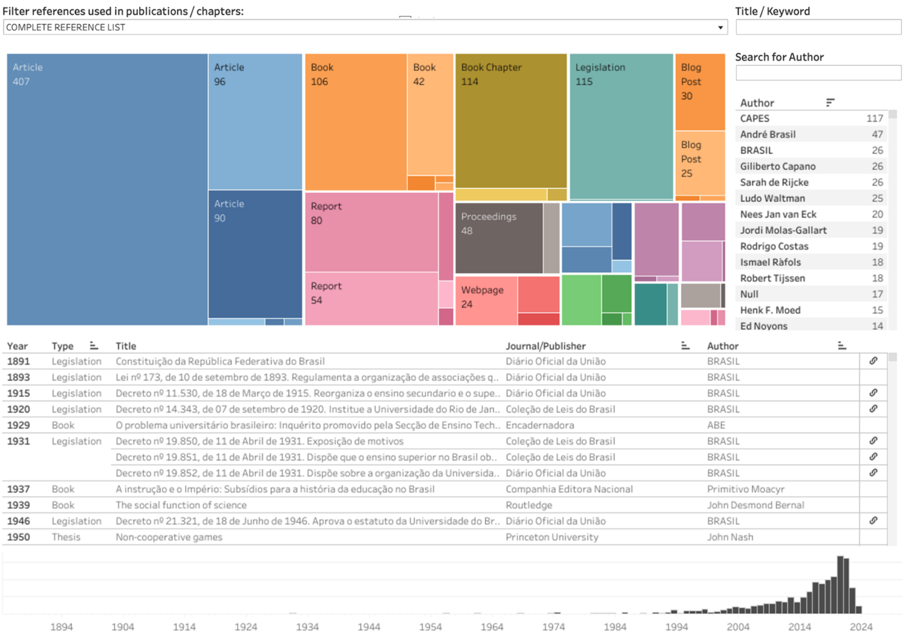


Figure B.3.: Main view of the bibliography explorer

B.4 Dashboard: Brazilian legislation on graduate education

Throughout this dissertation, more than 100 legislative items were examined, including laws, ordinances, and decrees. To assist those seeking to understand the construction process of the Brazilian National System of Graduate Education (SNPG) and its evaluation, a timeline has been prepared with these documents. The tool is available at https://andrebrasil.github.io/viz/timeline.html, and not only displays the researched legislation, but also provides direct access to the original texts.

It is crucial to note that the Brazilian National System of Graduate Education (SNPG) is predominantly regulated by the Ministry of Education and by CAPES. As such, the legislation mapped in this timeline at times serves to mould the system, while in other instances it plays a role in defining evaluation processes and standards. Therefore, its importance in understanding the Brazilian evaluation system is considerable.

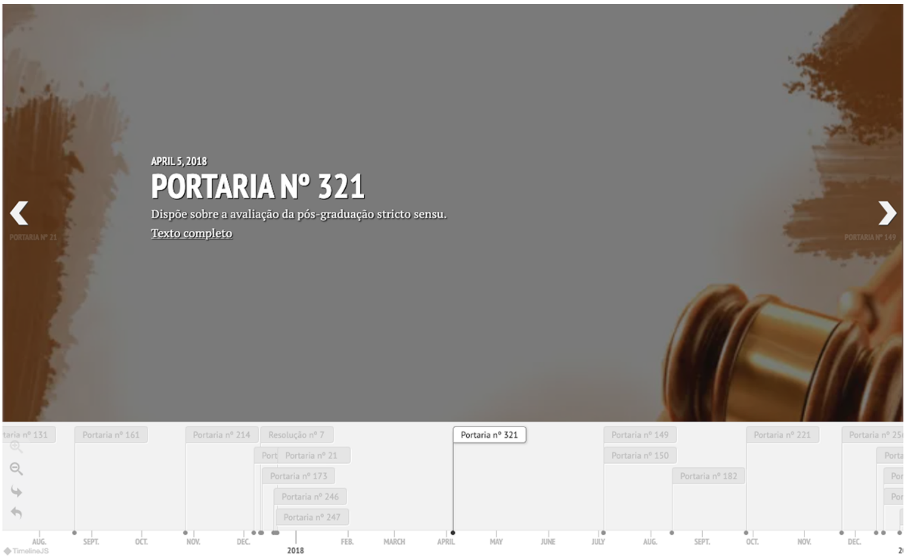


Figure B.4.: Legislation timeline available at https://andrebrasil.github.io/viz/timeline.html

Summary

The importance of evaluation in shaping the trajectory of the Brazilian scientific landscape cannot be overstated. It provides institutions, policymakers, and other stakeholders with tools to assess the efficacy and performance of its Brazilian National System of Graduate Education (SNPG). This system is the cornerstone of research in Brazil, and its evaluation influences funding allocation, institutional reputation, and the accreditation of graduate programs.

In the 1970s, the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) was tasked with overseeing the national evaluation system. Committed to championing research and graduate education excellence, the agency has for decades recognised and harnessed the transformative power of evaluation, consistently improving its methods through collaboration between its highly qualified staff, the academic community, and evaluation stakeholders.

However, advancing an ongoing evaluation system is a significant challenge, especially because necessary changes cannot be implemented in a disruptive way; they must be incremental. Furthermore, the rigours of public policy management pose a challenge to rethink the evaluation process, as the task is often overshadowed by everyday urgencies that can hinder the possibility of diving into the depth and breadth of available research that could support it.

This dissertation is rooted in such a context. After spending a decade as a CAPES policy officer, I had the opportunity to step away from my daily activities to explore graduate education and pathways to advance its evaluation. The chosen location for this research was the Centre for Science and Technology Studies (CWTS), the birthplace of the Leiden Manifesto for research metrics. This manifesto has significantly influenced recent reforms in Brazilian evaluation, underscoring the importance of transparency, diversity, and context in assessment. Within this nurturing environment, a rigorous research period culminated in this body of work, including articles published in various journals

and conferences. Part of the journey and my research approach are described in the Introduction, with the narrative unfolding into four subsequent parts.

Part I sets the stage, focussing on understanding the Brazilian National System of Graduate Education and its evaluation system. Inspired by the theory of path dependence and its implications for science policy reform, two chapters delve into the journey of these systems, with the aim of comprehending their development over recent decades. Chapter 2 embarks on a thorough exploration of the Brazilian science system, mapping its origins and growth influenced by pivotal policy decisions. Chapter 3 shifts the focus to the historical landscape of the evaluation of research and graduate education in Brazil and how the system has matured into an integral component of the national scientific environment.

Part II seeks to contextualise the Brazilian system on an international stage. For this, Chapter 4 provides a comparative analysis of research evaluation systems in Brazil and the Netherlands, illuminating how different trajectories and decisions have shaped their respective designs. This comparison highlights unique attributes, challenges, and outcomes; all arising from different historical and policy contexts. Brazil's performance-driven system propels its scientific endeavours, but risks fostering research uniformity, potentially stifling innovation. In contrast, the Dutch system prioritises research quality and societal relevance, encouraging diverse research paths. Lessons from both countries are valuable, but their distinct histories serve as evidence against a one-size-fits-all approach.

Another analysis carried out in Part II of the dissertation relates to the role that graduate education plays in Brazilian science. For this, Chapter 5 used a wealth of resources and data to dissect the design ethos of a science system focused on graduate education, unveiling its notably high influence on academic publishing in Brazil and providing fresh insights into the differences between the SNPG and other science systems. These differences are empirically recognised and can easily be inferred from the development trajectory of the SNPG. However, this dissertation seeks to find bibliometric evidence to support them.

The following Part III delves into some of the many defining characteristics and instruments of the Brazilian evaluation system, with the aim of analysing its development and current status, and contextualising its strengths and weaknesses. In this sense, Chapter 6 examines the Brazilian research classification adopted for the evaluation of graduate education. The chapter traces how the existing organisation around 49 areas was shaped from the influence of disciplinary and administrative elements, demonstrating how their restructuring is in order to address local and international incoherences.

Chapters 7 and 8 dive into various aspects of journal publishing in Brazil, including evaluation, database coverage, and multilingualism. Central to the discussion is the Qualis classification system and its role in assessing the quality of scholarly publications in the country. Although imperfect and in need of continuous evolution, Qualis should be recognised as an important achievement of the evaluation conducted in Brazil, as it combines quantitative and qualitative methods to produce the journal classification central in the country's evaluation.

Chapter 9 expands the analysis of journal publications in Brazil, investigating some of its well-established open access practises. The study highlights the relevance of local journals and databases such as SciELO as important alternatives to allow researchers to publish both in the local language and in English, often open access without paying any Article Processing Charges (APCs). The chapter also explores the potential adoption of the Purchasing Power Parity (PPP) index to help equalise the publishing costs for Brazilian researchers in the international arena.

Finally, Part III concludes with a broader perspective on open science. Considering that openness often goes beyond access to scholarly publications, Chapter 10 reviews efforts to diversify Brazilian research output to include a series of technical and technological products destined for a wider audience than the scientific community. The chapter analyses how CAPES has evolved its evaluation system to value and induce growth in that type of production, also looking at the first observable effects of the innitiative.

The final Part IV addresses challenges facing the evaluation system due to the exponential growth of the SNPG. Chapter 11 explores some of these challenges, especially the dangers of simplistic solutions to complex evaluation problems. An example of such a danger was the investigation conducted by CAPES on the possibility of implementing a multidimensional evaluation model close to that adopted by U-Multirank. The model discussed would emphasise overreliance on quantitative indicators, when a more holistic approach would be required to embrace the diverse academic landscape of Brazil. A proposed solution is in increasing the role of self-assessment in Brazilian evaluation, emphasising the need for further involvement of the country's higher education institutions (HEI) in the much needed exercise of their institutional autonomy.

The dissertation concludes with Chapter 12, which synthesises the research findings and examines the Brazilian evaluation system through the lens of the ten principles of the Leiden Manifesto. From this perspective, the final chapter pinpoints vital concerns that demand attention and offers tailored recommendations to enhance the nation's assessment landscape. Central to these recommendations is the imperative to acknowledge the shortcomings of the current evaluation system. The graduate program serves as the primary assessment unit within the Brazilian system, and the SNPG is on track to encompass 5000 of these programs, spread across over 400 educational institutions. Continuing with a centralised, comparative and simultaneous national evaluation is not sustainable if the system aims to promote innovation and diversity.

The future of Brazilian evaluation requires some changes from the current top-down approach. First, it should foster partnerships with institutions for more inclusive multidimensional self-assessment strategies. Evaluation should also shift toward fostering educational and guidance-orientated outcomes instead of simply punitive or rewarding ones. Given its influential role, the Brazilian evaluation system is uniquely positioned to catalyse meaningful change, and its evolution can push the SNPG to become a more multifaceted system with room for varied research profiles, thus better addressing societal needs.

Nederlandse Samenvatting

Het belang van evaluatie in de geschiedenis van de Braziliaanse wetenschap staat buiten kijf. Het biedt instellingen, beleidsmakers en andere belanghebbenden de middelen om de effectiviteit en prestaties van hun Nationaal Postgraduate Systeem (SNPG) te beoordelen. Dit systeem is de steunpilaar van het onderzoek in Brazilië en de beoordeling ervan beïnvloedt de verdeling van middelen, de institutionele reputatie en de accreditatie van Postgraduateprogramma's.

In de jaren 1970 kreeg de Coördinatie voor de Verbetering van Hoger Onderwijs Personeel (CAPES) de missie om toezicht te houden op het nationale evaluatiesysteem. Het agentschap, dat zich ten doel heeft gesteld uitmuntend onderzoek en postgraduate onderwijs te bevorderen, erkent en benut al decennia de kracht van evaluatie, waardoor veranderingen tot stand komen. In samenwerking met het hooggekwalificeerde team, de academische gemeenschap en de stakeholders van evaluatie worden de gehanteerde methoden voortdurend verfijnd.

Het verder ontwikkelen van een evaluatiesysteem is echter een belangrijke uitdaging, vooral omdat noodzakelijke veranderingen niet ontwrichtend moeten worden doorgevoerd, maar stapsgewijs. Bovendien vormen de eisen van openbaar beleidsmanagement een uitdaging om het evaluatieproces te heroverwegen, omdat de missie vaak overschaduwd wordt door ad hoc bezigheden die een grondig gebruik van beschikbaar onderzoek dat dergelijke taken zou kunnen ondersteunen, kunnen bemoeilijken.

Dit proefschrift is geworteld in deze context. Na tien jaar bij het Directoraat Evaluatie van CAPES, heb ik mijn normale werkzaamheden tijdelijk stilgelegd om het postgraduatesysteem en evaluatievoortgang te onderzoeken. De gekozen locatie voor dit onderzoek was het Centrum voor Wetenschaps- en Technologiestudies (CWTS), waar het Leidse Manifest ontstond. Dit manifest heeft in belangrijke mate bijgedragen aan recente hervormingen in de Braziliaanse evaluatie en benadrukt het belang van transparantie, diversiteit en context in evaluatie. Binnen deze omgeving leidde een periode van intensief onderzoek tot dit werk, inclusief artikelen die zijn gepubliceerd in verschillende tijdschriften en conferenties. Een deel van dit traject en mijn onderzoeksopzet worden beschreven in de Inleiding, waarna het verhaal zich ontvouwt in vier delen.

In Deel I wordt uitgelegd hoe het postgraduatesysteem en het evaluatiesysteem functioneren. Geïnspireerd door de theorie van pad afhankelijkheid en de implicaties voor hervorming van wetenschapsbeleid, wordt in twee hoofdstukken ingegaan op de geschiedenis van deze systemen om hun ontwikkeling in de afgelopen decennia te begrijpen. Hoofdstuk 2 bevat een verkenning van het Braziliaanse wetenschapssysteem, de oorsprong, groei en de invloed van de politiek op het systeem. Hoofdstuk 3 verlegt de focus naar een historisch overzicht van de evaluatie van onderzoek en postgraduateonderwijs in Brazilië en toont de centrale rol waarvan in het wetenschappelijke landschap.

Deel II plaatst het Braziliaanse systeem in een internationale context. Tegen deze achtergrond biedt Hoofdstuk 4 een vergelijkende analyse van onderzoeksevaluatiesystemen in Brazilië en Nederland, waarbij wordt benadrukt hoe verschillende trajecten en beslissingen van invloed zijn geweest op de vormgeving van die systemen. De vergelijking belicht unieke kenmerken, uitdagingen en resultaten die voortvloeien uit verschillende historische en politieke contexten. Het Braziliaanse systeem is prestatiegericht, wat wetenschappelijke ontwikkeling aanmoedigt, maar innovatie kan belemmeren door het streven naar onderzoeksuniformiteit. Het Nederlandse systeem geeft daarentegen prioriteit aan de onderzoekskwaliteit en de relevantie ervan voor de maatschappij, en moedigt diverse onderzoekstrajecten aan. Lessen uit beide landen zijn waardevol, maar hun verschillende achtergronden pleiten tegen een uniforme aanpak.

Een andere analyse in Deel II betreft de rol van postgraduateonderwijs in de Braziliaanse wetenschap. In dit verband is in Hoofdstuk 5 gebruikt gemaakt van een groot aantal bronnen en gegevens om het ontwerpethos van een postgraduate georiënteerd wetenschapssysteem te ontleden, de opmerkelijke invloed ervan op academische publicaties in Brazilië aan te tonen en nieuwe inzichten te bieden in de verschillen tussen de SNPG en andere wetenschapssystemen. Deze verschillen zijn empirisch vast te stellen en kunnen gemakkelijk worden afgeleid uit het ontwikkelingstraject van de SNPG. Deze dissertatie probeert echter door middel van bibliometrische analyse bewijs te vinden om de verschillen daadwerkelijk vast te stellen.

In Deel III wordt ingegaan op enkele van de belangrijkste kenmerken en instrumenten van het Braziliaanse evaluatiesysteem, met als doel de ontwikkeling en huidige status ervan te analyseren en de sterke en zwakke punten in context te plaatsen. Hoofdstuk 6 onderzoekt het Braziliaanse classificatiesysteem voor postgraduate evaluatie. Het hoofdstuk beschrijft hoe de huidige organisatie rond 49 evaluatiegebieden tot stand is gekomen onder invloed van disciplinaire en administratieve factoren, en laat zien hoe herstructurering nodig is om lokale en internationale inconsistenties aan te pakken.

De Hoofdstukken 7 en 8 behandelen verschillende criteria voor publicatie in tijdschriften in Brazilië, waaronder de beoordeling, databasedekking en meertaligheid. Het classificatiesysteem Qualis en zijn rol in de beoordeling van academische publicaties is een centraal element in deze discussie. Hoewel dit systeem niet perfect is en voortdurend geëvalueerd dient te worden, moet Qualis worden erkend als een belangrijke beoordelingsinstrument in Brazilië, omdat het kwantitatieve en kwalitatieve methoden combineert die essentieel zijn voor het nationale beoordelingssysteem.

Hoofdstuk 9 breidt de analyse van tijdschriftpublicaties in Brazilië uit en onderzoekt enkele van hun gevestigde open access praktijken. Het benadrukt het belang van lokale tijdschriften en databases, zoals SciELO, als belangrijke alternatieven om onderzoekers in staat te stellen zowel in de lokale taal als in het Engels te publiceren, vaak met open toegang zonder artikelverwerkingskosten (APC) te hoeven betalen. Het hoofdstuk onderzoekt ook de mogelijke toepassing van de koopkrachtpariteit-index (PPP) om de publicatiekosten voor Braziliaanse onderzoekers op het internationale toneel gelijk te trekken.

Ten slotte biedt Deel III een breder perspectief op open wetenschap. Aangezien de gewenste openheid vaak verder gaat dan academische publicaties, behandelt Hoofdstuk 10 inspanningen om de productie uit Braziliaans onderzoek te diversifiëren naar technische en technologische producten voor een breder publiek dan alleen de wetenschappelijke gemeenschap. Het hoofdstuk onderzoekt hoe CAPES zijn beoordelingssysteem heeft ontwikkeld om deze productie te waarderen en aan te moedigen, en kijkt naar de eerste effecten van dit initiatief.

Deel IV gaat in op de uitdagingen waarmee het beoordelingssysteem wordt geconfronteerd door de exponentiële groei van SNPG. Hoofdstuk 11 verkent enkele van deze uitdagingen, in het bijzonder de gevaren van simplistische oplossingen voor complexe beoordelingsproblemen. Een voorbeeld van dit gevaar was het onderzoek van CAPES naar de mogelijke implementatie van een multidimensionaal beoordelingsmodel vergelijkbaar met dat van U-Multirank. Het besproken model zou de nadruk leggen op een te grote afhankelijkheid van kwantitatieve indicatoren, terwijl een meer holistische benadering nodig zou zijn om de academische diversiteit van Brazilië te omarmen. Een voorgestelde oplossing is om de rol van zelfevaluatie in de Braziliaanse evaluatie te vergroten, waarbij de nadruk wordt gelegd op de behoefte aan een grotere betrokkenheid van de instellingen voor hoger onderwijs bij de noodzakelijke uitoefening van hun institutionele autonomie.

Het proefschrift sluit af met Hoofdstuk 12, dat de bevindingen van het onderzoek samenvat en het Braziliaanse beoordelingssysteem bekijkt door de lens van de tien principes van het Leidse Manifest. Vanuit dit perspectief belicht het laatste hoofdstuk cruciale aandachtspunten en biedt het specifieke aanbevelingen om het nationale beoordelingslandschap te verbeteren. Tot de belangrijkste aanbevelingen behoort de noodzaak om de tekortkomingen van het huidige beoordelingssysteem te erkennen. Het postgraduateprogramma is de belangrijkste eenheid van beoordeling binnen het Braziliaanse systeem en de SNPG bevat al 5000 van dergelijke programma's in meer dan 400 instellingen voor hoger onderwijs. Doorgaan met gecentraliseerde, vergelijkende en gelijktijdige nationale beoordeling is niet houdbaar als het systeem innovatie en diversiteit wil bevorderen.

De toekomst van de Braziliaanse beoordeling vereist een aantal veranderingen ten opzichte van de huidige top-down benadering. Ten eerste moeten allianties met instellingen worden aangemoedigd om meer inclusieve multidimensionale zelfevaluatiestrategieën te ontwikkelen. De beoordeling moet ook worden aangepast om meer aandacht te besteden aan de educatieve dimensie van SNPG, gerelateerd aan de opleiding van personeel. Het is ook essentieel dat de Braziliaanse beoordeling meer formatief wordt, in plaats van het huidige model dat gebaseerd is op een systeem van straffen en belonen. Gezien de onmiskenbare invloed die het heeft, is de Braziliaanse beoordeling goed gepositioneerd om significante veranderingen teweeg te brengen in de SNPG zodat het zich ontwikkelt tot een veelzijdiger systeem dat ruimte biedt aan een verscheidenheid aan onderzoek profielen die tegemoet kunnen komen aan de uiteenlopende behoeften van de samenleving.

Sumário em Português

A importância da avaliação na trajetória da ciência brasileira é inquestionável. Ela fornece às instituições, gestores de políticas públicas e demais interessados ferramentas para avaliar a eficácia e o desempenho do seu Sistema Nacional de Pós-Graduação (SNPG). Este sistema é a pedra angular da pesquisa no Brasil, e sua avaliação influencia a alocação de recursos, a reputação institucional e a acreditação de programas de pós-graduação.

Na década de 1970, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) foi incumbida de supervisionar o sistema nacional de avaliação. Comprometida em defender a excelência em pesquisa e pós-graduação, a fundação reconhece e faz uso do poder transformador da avaliação há décadas, aprimorando continuamente seus métodos com a colaboração entre sua equipe altamente qualificada, a comunidade acadêmica e os interessados na avaliação.

No entanto, avançar um sistema de avaliação de forma contínua é um desafio significativo, principalmente porque as mudanças necessárias não podem ser implementadas de forma disruptiva; elas devem ser incrementais. Além disso, os rigores da gestão de políticas públicas representam um desafio para repensar o processo de avaliação, já que a missão é muitas vezes ofuscada por urgências cotidianas que podem dificultar a devida exploração da pesquisa disponível que poderia dar suporte a tal tarefa.

Esta dissertação está enraizada nesse contexto. Após passar uma década como servidor da Diretoria de Avaliação da CAPES, tive a oportunidade de me afastar das minhas atividades diárias para investigar a pós-graduação e os caminhos para avançar em sua avaliação. O local escolhido para esta pesquisa foi o Centre for Science and Technology Studies (CWTS), local de origem do Manifesto de Leiden para métricas de pesquisa. Tal manifesto influenciou significativamente reformas recentes na avaliação brasileira, destacando a importância da transparência, diversidade e contexto na avaliação. Dentro deste ambiente propício, um rigoroso período de pesquisa culminou neste corpo de trabalho, incluindo artigos publicados em diversos periódicos e conferências. Parte da jornada e da minha abordagem de pesquisa são descritas na Introdução, com a narrativa desdobrando-se em quatro partes subsequentes.

A Parte I prepara o terreno, focando em entender o Sistema Nacional de Pós-Graduação e seu sistema de avaliação. Inspirados pela teoria da dependência de trajetória e suas implicações para a reforma de políticas científicas, dois capítulos mergulham na história desses sistemas, com o objetivo de compreender seu desenvolvimento nas últimas décadas. O Capítulo 2 embarca numa exploração do sistema científico brasileiro, mapeando suas origens, crescimento e influências políticas. O Capítulo 3 muda o foco para o panorama histórico da avaliação de pesquisa e da pós-graduação no Brasil, mostrando o amadurecimento do sistema para se tornar um componente central no cenário científico nacional.

A Parte II contextualiza o sistema brasileiro no cenário internacional. Para isso, o Capítulo 4 oferece uma análise comparativa dos sistemas de avaliação de pesquisa no Brasil e nos Países Baixos, iluminando como diferentes trajetórias e decisões moldaram seus respectivos designs. A comparação destaca atributos, desafios e resultados únicos; frutos de diferentes contextos históricos e políticos. O sistema Brasileiro é focado em desempenho, o que impulsiona seu desenvolvimento científico, mas tende a promover a uniformidade da pesquisa, potencialmente sufocando a inovação. Em contraste, o sistema holandês prioriza a qualidade da pesquisa e a relevância para a sociedade, incentivando trajetórias de pesquisa diversas. Lições de ambos os países são valiosas, mas suas histórias distintas servem como evidência contra a adoção de uma abordagem única para todos.

Outra análise realizada na Parte II da dissertação relaciona-se com o papel que a pós-graduação desempenha na ciência brasileira. Para isso, o Capítulo 5 utilizou uma riqueza de recursos e dados para dissecar o ethos de design de um sistema de ciência focado na pós-graduação, revelando sua influência notavelmente alta na publicação acadêmica no Brasil e fornecendo novos insights sobre as diferenças entre o SNPG e outros sistemas de ciência. Essas diferenças são empiricamente reconhecidas e podem ser facilmente inferidas a partir da trajetória de desenvolvimento do SNPG. No entanto, esta dissertação busca encontrar evidências bibliométricas para apoiá-las.

A Parte III aprofunda-se em algumas das características e instrumentos centrais do sistema de avaliação brasileiro, com o objetivo de analisar seu desenvolvimento e status atual, contextualizando seus pontos fortes e fracos. Nesse sentido, o Capítulo 6 examina o sistema de classificação de pesquisa brasileira adotado para a avaliação da pós-graduação. O capítulo traça como a organização existente em torno de 49 áreas de avaliação, nove grandes áreas e três colégios foi moldada a partir da influência de elementos disciplinares e administrativos, demonstrando como sua reestruturação é necessária para abordar incoerências locais e internacionais.

Os Capítulos 7 e 8 exploram diversos aspectos da publicação em periódicos no Brasil, incluindo sua avaliação, cobertura de bases de dados e multilinguismo. O sistema de classificação Qualis e seu papel na avaliação da qualidade das publicações acadêmicas no país é um elemento central nessa discussão. Embora imperfeito e necessitando de evolução contínua, o Qualis deve ser reconhecido como uma importante conquista da avaliação realizada no Brasil, pois combina métodos quantitativos e qualitativos para produzir a classificação de periódicos que é de grande importância no sistema de avaliação do país.

O Capítulo 9 expande a análise sobre as publicações em periódicos no Brasil, investigando algumas de suas práticas consolidadas de acesso aberto. O estudo destaca a relevância de periódicos e bases de dados locais, tais como o SciELO, como alternativas importantes para permitir que os pesquisadores publiquem tanto no idioma local quanto em inglês, muitas vezes em acesso aberto sem pagar nenhuma taxa de processamento de artigos (APC). O capítulo também explora a possível adoção do índice de Paridade do Poder de Compra (PPP) para ajudar a equalizar os custos de publicação para os pesquisadores brasileiros no cenário internacional.

Por fim, a Parte III conclui com uma perspectiva mais ampla sobre ciência aberta. Considerando que a abertura desejada frequentemente vai além do acesso a publicações acadêmicas, o Capítulo 10 revisa esforços para diversificar a produção oriunda da pesquisa brasileira para incluir uma série de produtos técnicos e tecnológicos destinados a um público mais amplo do que a comunidade científica. O capítulo analisa como a CAPES evoluiu seu sistema de avaliação para valorizar e induzir o crescimento desse tipo de produção, observando também os primeiros efeitos da iniciativa.

A Parte IV aborda desafios que o sistema de avaliação enfrenta devido ao crescimento exponencial do SNPG. O Capítulo 11 explora alguns desses desafios, especialmente os perigos de soluções simplistas para problemas complexos de avaliação. Um exemplo desse perigo foi a investigação realizada pela CAPES sobre possível implantação de um modelo de avaliação multidimensional próximo ao adotado pelo U-Multirank. O modelo discutido enfatizaria a excessiva confiança em indicadores quantitativos, quando uma abordagem mais holística seria necessária para abraçar a diversidade acadêmica do Brasil. Uma solução proposta está em aumentar o papel da autoavaliação na avaliação brasileira, enfatizando a necessidade de maior envolvimento das instituições de ensino superior (IES) no tão necessário exercício de sua autonomia institucional.

A dissertação conclui com o Capítulo 12, que sintetiza as descobertas da pesquisa e examina o sistema de avaliação brasileiro através da lente dos dez princípios do Manifesto de Leiden. A partir desta perspectiva, o capítulo final destaca preocupações vitais que exigem atenção e oferece recomendações específicas para aprimorar o cenário de avaliação nacional. Entre as principais recomendações está a necessidade de reconhecer as falhas do atual sistema de avaliação. O programa de pós-graduação serve como a principal unidade de avaliação dentro do sistema brasileiro, e o SNPG está a caminho de abranger 5000 destes programas, distribuídos em mais de 400 instituições de ensino superior. Continuar com uma avaliação nacional centralizada, comparativa e simultânea não é sustentável se o sistema pretende promover inovação e diversidade.

O futuro da avaliação brasileira requer algumas mudanças em relação à abordagem top-down atual. Primeiramente, deve-se incentivar parcerias com instituições para adotar estratégias de autoavaliação multidimensional mais inclusivas. A avaliação também precisa de ajustes para dar mais evidência à dimensão educacional do SNPG, relacionada à formação de pessoal. Também é essencial que a avaliação brasileira tenha um caráter mais formativo, ao invés do modelo atual baseado em um sistema de punição e recompensas. Dada a sua inegável influência, a avaliação brasileira está bem posicionada para induzir mudanças significativas no SNPG, de forma que este evolua para um sistema mais multifacetado, onde existe espaço para uma variedade de perfis de pesquisa capazes de atender às distintas necessidades da sociedade.

About the Author

André Brasil

www.andrebrasil.net

a.l.brasil@cwts.leidenuniv.nl

@andrebrasilbsb

andre-brasil

André Brasil is a researcher at the Centre for Science and Technology Studies (CWTS | Leiden University), focusing on national evaluation systems, scientometrics, scholarly publishing, open science, responsible research and innovation (RRI) and diversity in science, especially concerning multilingualism and geographic inclusion. As part of his activities, he is a member of the UNESCO Chair for Diversity and Inclusion in Global Science and a research fellow at the Research on Research Institute (RoRI). André is also affiliated with the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES), working at the Division of Studies and Research in Evaluation.

André completed his secondary education at Colégio Objetivo in 1993. With a drive for a varied academic foundation, he earned a bachelor's degree in Languages & Literature alongside an associate degree in Marketing Management from the Methodist University of São Paulo (UMESP). Enhancing his profile, he earned specialist degrees in Business Intelligence and Teaching in Higher Education. In 2018, he earned his Master's degree in Public Policy and Development from the Brazilian Institute for Applied Economic Research (IPEA). His doctoral research was conducted at CWTS with the focus on improving the Brazilian evaluation system for research and graduate education, a key strategy to ensure the quality and development of Science & Technology in the country.

List of publications

Articles in this dissertation

Brasil, A. (2020a). Building a national system of research and graduate education: How did the university become the house of science in Brazil? *Revista NUPEM*, *25*(1), 178–198. https://doi.org/10.1590/s1414-407720200001000010

Brasil, A. (2021b). A national evaluation push towards increased societal impact: The Brazilian experience in valuing broader research outputs. *SocArXiv*. https://doi.org/10.31235/osf.io/s98x3

Brasil, A. (2022c). Multidimensionality through self-evaluation: from theory to practice in the Brazilian graduate system. *fteval Journal for Research and Technology Policy Evaluation*, (53). https://doi.org/10.22163/fteval.2022.546

Brasil, A. (2023c). Expanding GeoCapes usefulness for research: Introducing the GeoCapes R package. Manuscript submitted for publication.

Brasil, A. (2023f). Rethinking a national classification of research and graduate education. *Profesional de la Información*, *32*(2), e320224. https://doi.org/10.3145/epi.2023.mar.24

Brasil, A. (2023h). The dynamics of a national evaluation system: The Brazilian model and its role in developing research and graduate education in the country. *SocArXiv*. https://doi.org/10.31235/osf.io/wf5kn

Brasil, A., & Trevisol, J. V. (2023b). Research evaluation in Brazil and the Netherlands: A comparative study. *SocArXiv*. https://doi.org/10.31235/osf.io/er9fa

Other articles

Bernstein, M. J., Nielsen, M. W., Alnor, E., **Brasil, A.**, Birkving, A. L., Chan, T. T., . . . Mejlgaard, N. (2022). The Societal Readiness Thinking Tool: A Practical Resource for Maturing the Societal Readiness of Research Projects. *Science and Engineering Ethics*, *28*(1), 6. https://doi.org/10.1007/s11948-021-00360-3

Trevisol, J. V., & **Brasil, A.** (in press). As políticas de avaliação da pesquisa e da pós-graduação nos países baixos: Lições sobre autoavaliação e autonomia institucional. *Revista Brasileira de Educação*, *27*.

Trevisol, J. V., Bastiani, S., & **Brasil, A.** (2020). As dinâmicas da Pós-graduação em Santa Catarina: evolução e perspectivas (1969 - 2018). *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, *25*(1), 178–198. https://doi.org/10.1590/s1414-407720200001000010

Book chapters

Brasil, A., & Trevisol, J. V. (2021). O Sistema Nacional de Pós-Graduação no Paraná: Características e Dinâmicas de Desenvolvimento. In V. d. S. Oliveira & F. G. Costa (Eds.), *Pós-graduação e pesquisa: Formação e desafios no contexto da UENP*. Cornélio Procópio: Editora UENP.

Conference full papers

Brasil, A. (2019a). 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. (2021c). A national evaluation push towards increased societal impact: The Brazilian experience in valuing broader research outputs. In *Eu-SPRI Conference 2021*, Oslo, Norway.

Brasil, A. (2021d). Beyond the Web of Science: An overview of Brazilian papers indexed by regionally relevant databases. In *18th Conference of the International Society for Scientometrics and Informetrics (ISSI)*, Leuven, Belgium.

Brasil, A. (2022f). Rethinking a national classification system of research and graduate education. In *26th International Conference on Science, Technology and Innovation Indicators (STI)*, Granada, Spain.

Brasil, A. (2023a). Between Bibliometrics and Peer Review: The Evolution and Challenges of Brazil's Qualis System. In *27th International Conference on Science, Technology and Innovation Indicators (STI)*. https://doi.org/10.55835/6442f8d5389baf2145b264d9

Brasil, A. (2023i). The impact of graduate education on academic publishing: A comparative analysis from a Brazilian perspective. In *19th Conference of the International Society for Scientometrics and Informetrics (ISSI)*, Bloomington, IN.

Brasil, A., Chen, C. M.-L., & van Leeuwen, T. (2022). Analysing Open Access publishing beyond Plan S: Motivations for publishing in Hybrid and Bronze OA formats. In *26th International Conference on Science, Technology and Innovation Indicators (STI)*, Granada, Spain.

Brasil, A., Trevisol, J. V., & van Drooge, L. (2022). Research evaluation in Brazil and the Netherlands: A comparative study. In *Eu-SPRI Conference 2022*, Utrecht, Netherlands.

Brasil, A., & van Leeuwen, T. (2022). The unseen costs of article processing charges: The different realities of Brazil and the Netherlands. In *26th International Conference on Science, Technology and Innovation Indicators (STI)*, Granada, Spain.

Conference / Workshop abstracts and posters

Brasil, A. (2019c). The Brazilian Portal of Journals: Government funded access to scientific publications [Poster], In *WTMC Autumn Workshop: Open*, Deursen-Dennenburg, Netherlands.

Brasil, A. (2020b). Evolving the evaluation of graduate education [Poster], In *CWTS PhD Workshop*, Leiden, Netherlands.

Brasil, A. (2021e). 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.

Brasil, A. (2021h). When indicators are not enough: The Brazilian Qualis experience in combining metrics with peer-review [Abstract], In *Book of abstracts of the Nordic Workshop on Bibliometrics and Research Policy*, Odense, DK: University of Southern Denmark.

Brasil, A. (2022b). Multidimensionality through self-evaluation: From theory to practice in the Brazilian graduate system [Poster], In *REvaluation'22*, Vienna, Austria.

Brasil, A. (2023g). Rethinking evaluation to help steer a national science system [Extended abstract], In *Eu-SPRI Conference 2023*, Brighton, United Kingdom.

Brasil, A., Meijer, I., & Llorente, C. (2022). Responsibility in research evaluation practices: Lessons from a global discussion [Poster], In *REvaluation'22*, Vienna, Austria.

Jong, S., Klippe, W., **Brasil, A.**, Chan, T., & Meijer, I. (2020). Online tools to bridge the gap between research policy and academic practice [Short paper], In *Eu-SPRI, https://euspri2020.nl (Conference cancelled - Work presented at virtual Eu-SPRI track)*, Utrecht, Netherlands.

Trevisol, J. V., & **Brasil, A.** (2020). A autoavaliação nas políticas da pós-graduação brasileira [Extended abstract], In *Proceedings of the XIII Reunião Científica da ANPEd-Sul*, Blumenau, SC.

Scientific reports

Brasil, A., & Trevisol, J. V. (2023a). *ENRESSH country report: Brazil*. ENRESSH.

Calero-Medina, C., & **Brasil, A.** (2020, March). *Bibliometric analysis of research collaboration between Brazil and the Netherlands 2008–2018: Report for Netherlands Innovation Network*. Leiden, Netherlands: CWTS BV.

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*. Brasília, Brazil: CAPES. Retrieved from https://bit.ly/2PmAf0q

van Drooge, L., & **Brasil, A.** (2022, February). *Bibliometric analysis of the faculty of Veterinary Medicine 2008–2018*. Leiden, Netherlands: CWTS BV.

Winter, E., **Brasil, A.**, Moraes, C. L. d., Granjeiro, J. M., Filho, J. R. d. S. V., Nichiata, L. Y. I., ... Pascutti, P. G. (2019). *Produção técnica*. Brasília, Brazil: CAPES. Retrieved from https://bit.ly/3t06SiJ

Scientific blog posts and magazine articles

Brasil, A. (2020d, March 12). Sorbonne declaration on research data rights [Leiden Madtrics]. Retrieved from https://leidenmadtrics.nl/articles/sorbonne-declaration-on-research-data-rights

Brasil, A. (2021f, September 7). Move away from metrics, but beware of subjectivity [SUPER MoRRI Blog]. Retrieved February 2022, from https://bit.ly/32V9nM2

Brasil, A. (2021g, June 1). Responsibility in research evaluation practices: Lessons from a global discussion [SUPER MoRRI Blog]. Retrieved from https://super-morri.eu/responsibility-in-research-evaluation/

Brasil, A. (2023d, March 4). Open Science Knowledge Platform: A journey to a dynamic resource [Leiden Madtrics]. Retrieved from https://leidenmadtrics.nl/articles/sorbonne-declaration-on-research-data-rights

Brasil, A., & Waltman, L. (2022). Make analytics inclusive: Publication databases still miss too much of the world's research [Research Europe]. Retrieved February 20, 2022, from http://bit.ly/123412

Brasil, A., Eun, S., van de Klippe, W., & Ràfols, I. (2020, April 30). Covid-19: What do funders consider relevant research? [Leiden Madtrics]. Retrieved from https://leidenmadtrics.nl/articles/covid-19-what-do-funders-consider-relevant-research

Invited talks

Brasil, A. (2019b). Gender policies in science in Brazil [Presentation]. In CWTS & Elsevier's International Center for the Study of Research (Org.), *Workshop Gender Inequalities in Science*: October 2019, Leiden, Netherlands.

Brasil, A. (2020c). Societal Readiness Thinking Tool [Presentation]. In Fraunhofer Institut (Org.), *4th Workshop of the NewHorrizon Social Lab Health*: January 2020, Berlin, Germany.

Brasil, A. (2020e). The Brazilian System of Research and Graduate Education: Overview and potencial for cooperation with LUMC [Presentation]. In Leiden University Medical Center (Org.), *Mini-seminar Brazilian-Leiden collaborations: Novel opportunities*: February 2020, Leiden, Netherlands.

Brasil, A. (2021a). (E)valuating research topics poorly covered by bibliometric databases [Presentation]. In European Network of Research Evaluation in the Social Sciences and Humanities (ENRESSH) (Ed.), *EvalHum General Assembly 2021*: December 2021.

Brasil, A. (2022a). Diversity and the opening of science. In Grupo Montevideo & Coimbra Group (Org.), *Joint high-level workshop: Strengthening academic collaboration between Latin America and Europe*: December 2022, Montevideo, Uruguay. Retrieved from https://www.coimbra-group.eu/policy-seminars/

Brasil, A. (2022d). Multilingualism at risk: The pressure to internationalise scholarly output in Latin America. In J. Pölönen, Z. Taskin, & E. Kulczycki (Org.), *STI 2022 special track on language bias in evaluative metrics*: September 2022, Granada, Spain. Retrieved from https://sti2022.org/program/

Brasil, A. (2022e). O impacto do sistema brasileiro de avaliação nos periódicos científicos. In Associação Brasileira de Editores Científicos (Org.), *ABEC Meeting 2022*: October 2022, Botucatu, Brazil. Retrieved from https://www.abecbrasil.org.br/eventos/meeting_2022/meeting22.php

Brasil, A. (2022g). The Brazilian model for the evaluation of scholarly output [Presentation], In *Federation of Finnish Learned Societies*: September 2022, Helsinki, Finland.

Brasil, A. (2022h). Valuing local language and regional relevance in research. In Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Org.), *International Open Science Week 2022*: October 2022, Ankara, Turkey. Retrieved from https://etkinlik.ulakbim.gov.tr/event/77/

Brasil, A. (2023b). Current challenges and prospects for the internationalization of science. In Federal University of Paraná (Org.), *Seminar on the Internationalization of Postgraduate Education and Research*: March 2023, Curitiba, Brazil. Retrieved from https://www.youtube.com/watch?v=7K8DqCDjx0I

Brasil, A. (2023e). Purchasing power parity and the unseen costs of APCs. In CAPES (Org.), *Collaborative workshop: Open access and transformative agreements*: May 2023, Brasília, Brazil. Retrieved from https://www.youtube.com/watch?v=m4LsuZofXt0

Kaiser, F., & **Brasil, A.** (2021). A Metodologia U-Multirank [Roundtable]. In ABAVE (Org.), *VII CONBRATRI – Avaliação educacional no contexto de desigualdades*: September 27–29, 2021, Brasília, Brazil.

Colophon

This document was designed by André Brasil using the *Clean Thesis* style as a starting point. The thesis was typeset using pdfLaTeX and the LaTeX 2$_\varepsilon$ environment, using the following packages:

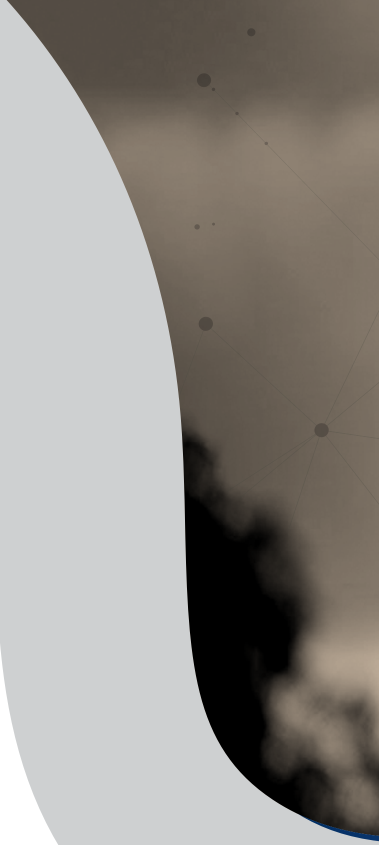
adjustbox	charter	glossaries	microtype	tabulary
amsfonts	cleanthesis	graphicx	multicol	textcomp
amsmath	csquotes	hyperref	multirow	tgheros
amssymb	enumitem	inputenc	rotating	threeparttable
array	float	listings	setspace	tocloft
babel	fontawesome	lmodern	siunitx	verbatim
biblatex	fontenc	longtable	subcaption	xcolor
booktabs	footmisc	lscape	tabto	xkeyval
caption	geometry	makecell	tabularx	

The original *Clean Thesis* style was developed by Ricardo Langner, based on user guide documents from Apple. The template is available at http://cleanthesis.der-ric.de/.

Most figures in this document were created using a combination of R, Tableau Desktop, and Flourish Studio. The code, data, static and interactive visualisations generated for this thesis can be accessed from the project Github repository:

https://github.com/AndreBrasil/advancing.evaluation

CWTS
Meaningful metrics



"Without evaluation, there is no quality at the graduate education level, which is where scientific research is conducted in Brazil."

Renato Janine Ribeiro