



Universiteit
Leiden
The Netherlands

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

Brasil Varandas Pinto, A.L.

Citation

Brasil Varandas Pinto, A. L. (2023, October 24). *Advancing the evaluation of graduate education: towards a multidimensional model in Brazil*. Retrieved from <https://hdl.handle.net/1887/3645840>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3645840>

Note: To cite this publication please use the final published version (if applicable).

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 (Balachevsky 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 (CNPGE) 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 CNPGE (1974), it is estimated that approximately 7000 students enrolled 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 (CNPGE, 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 (CNPGE) and the other from the agency itself. The CNPGE 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 CNPGE proposed comprehensive support measures to improve standards, including financial resources, advisory services, information dissemination, and quality recognition mechanisms (CNPGE, 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 scholarly 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, undoubtedly 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 (Ulhoa, 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 find-

ings 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 (CON-FAP), 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 acquired 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							Total
	1	2	3	4	5	6	7	
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").

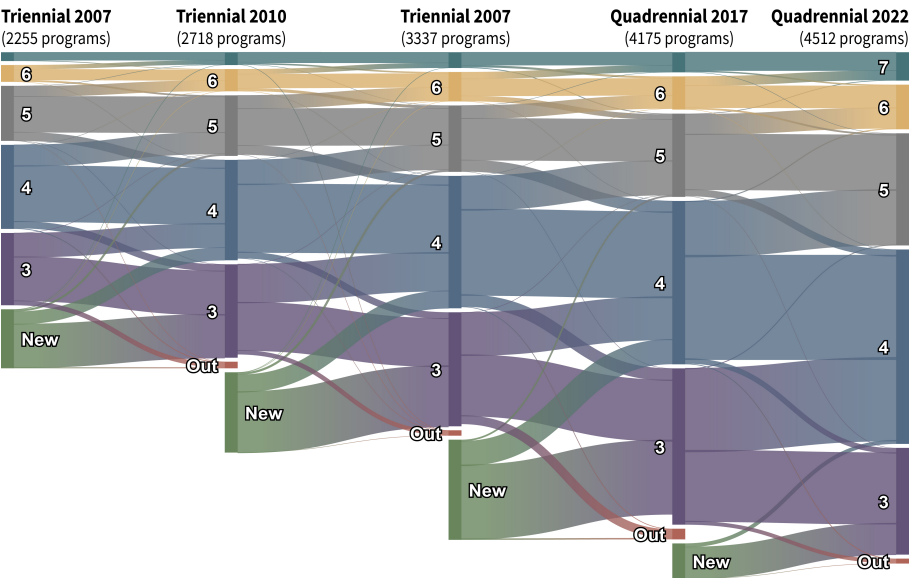


Figure 3.1.: Flow of graduate program grades over the 2007–2022 national evaluations

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.

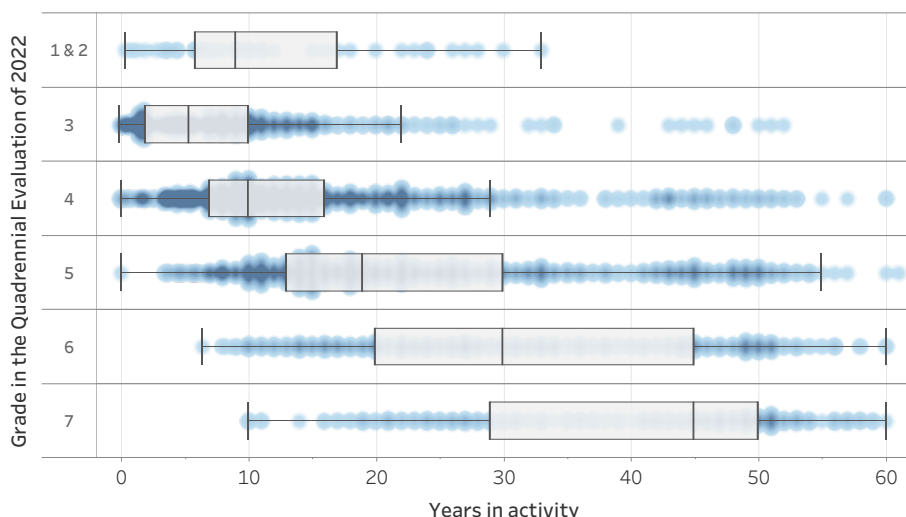


Figure 3.2.: Relation between national evaluation grade (2022) and age of PPG

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.

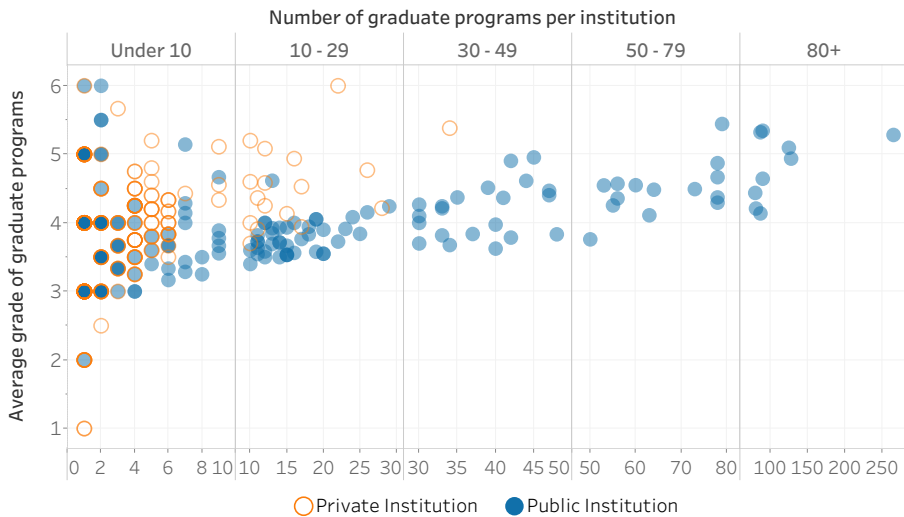


Figure 3.3.: Average grade of PPG vs total number of programs per institution (National Evaluation of 2022)

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.

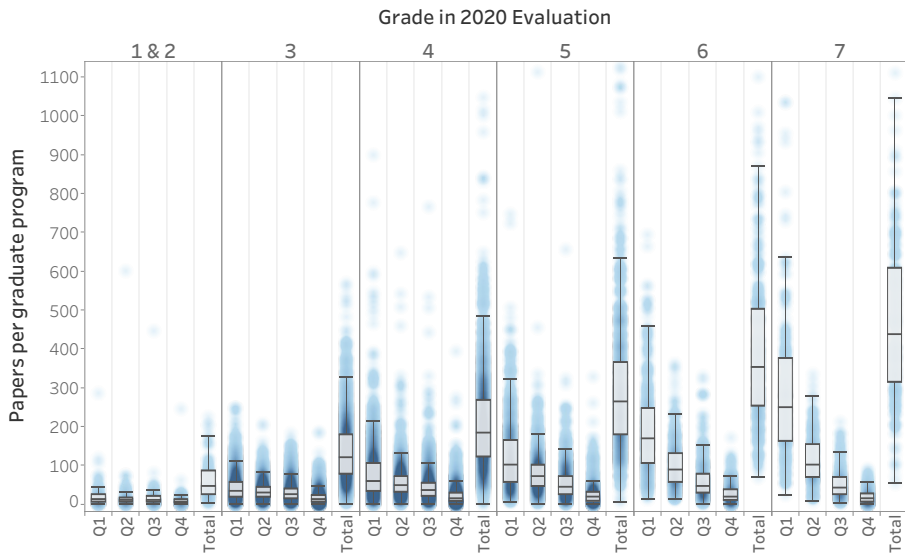


Figure 3.4.: Number of publications per PPG (per Qualis quartile and in total), according to grade obtained at the 2022 evaluation

[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.

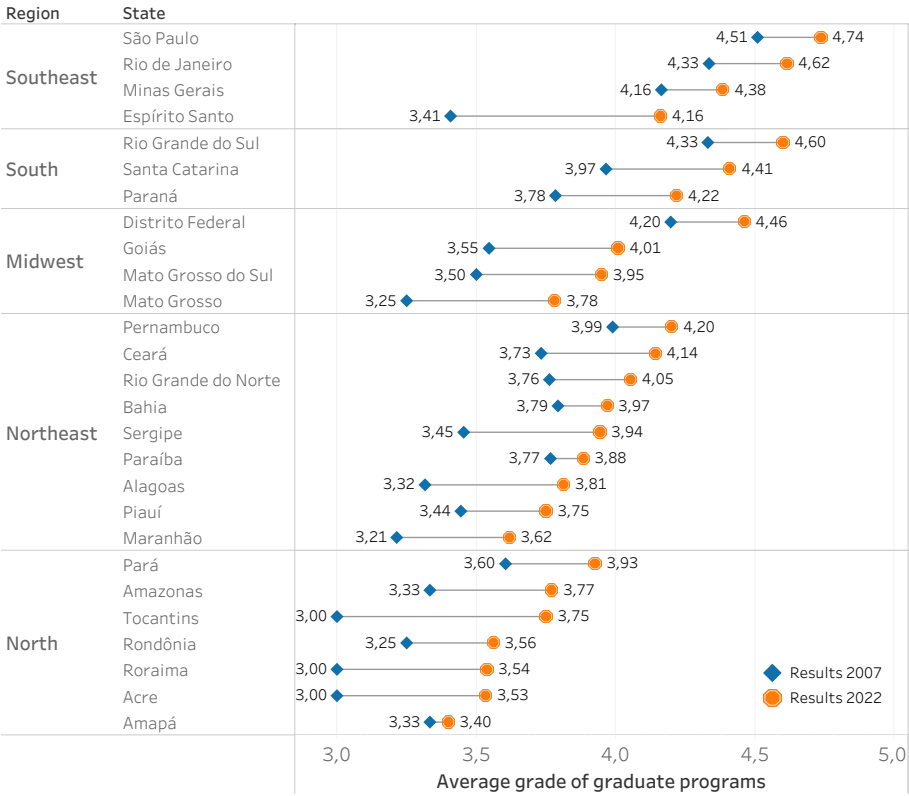


Figure 3.5.: Variation in grade average, per state, from the evaluations of 2007 to 2022

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.

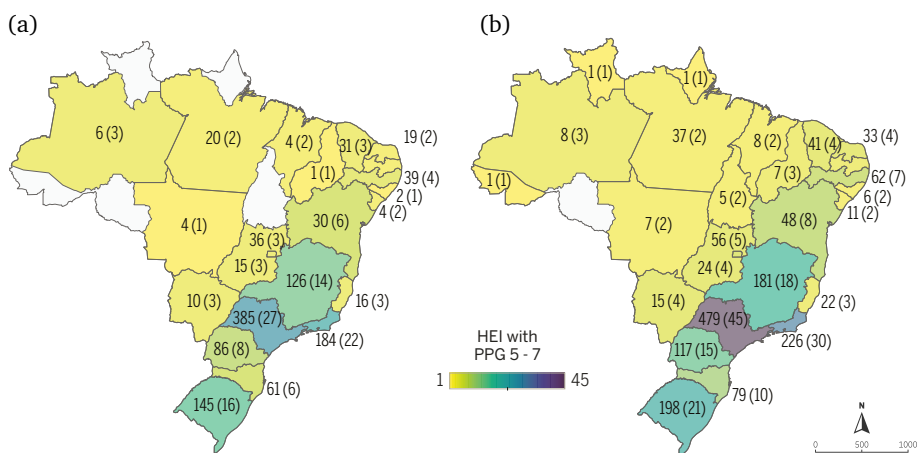


Figure 3.6.: Geographical distribution of graduate programs (grades 5-7) and the number of institutions offering them in (a) 2007 and (b) 2022

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 asymmetries 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 asymmetries 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, homogenization, 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, I. 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., ... Bão, 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/42SWTOM>
- 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.
- Ulhoa, 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>

