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

Brasil Varandas Pinto, A.L.

Part II

The Brazilian evaluation system in perspective

Research evaluation in Brazil and the Netherlands

4

> *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

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 con-

clude 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)

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 struggling 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