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

Brasil Varandas Pinto, A.L.

7

Between Bibliometrics and Peer Review

> *Qualis should not be considered an appropriate source for classifying the quality of scientific journals for purposes other than the evaluation of graduate programs.*
>
> **— Rita Barata**

Since the 1970s, Brazil has endeavoured to evaluate research and graduate education through a system that serves as a critical determinant of accreditation, permanence, and funding allocation (Brasil, 2021b; C. B. Martins, 2018). However, with the natural expansion of this system, evaluating scientific production qualitatively, a vital component of the process, became increasingly difficult (CAPES, 2003; Hortale, 2003). In response to these challenges, the Qualis ranking system was established in 1998 to assess the quality of academic journals as a proxy for the research contained therein (CAPES, 2003).

This chapter delves into the genesis of Qualis, examining its initial conception and its evolution over the years. Throughout its development, Qualis has shown adaptability to the dynamic nature of academic research and assessment, but cumulative improvements have led to or revealed fragilities in the system, some of which result from misuse of the ranking by external actors (M. F. B. Leite et al., 2010; Soma et al., 2016; Spagnolo and Souza, 2004). However, since its first

This chapter is an expanded version of the conference paper: Brasil, A. (2023). Between Bibliometrics and Peer Review: The Evolution and Challenges of Brazil's Qualis System. *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. Leiden, The Netherlands. https://doi.org/10.55835/6442f8d5389baf2145b264d9

use, Qualis has played an important role in the evaluation of research and graduate education in Brazil, adding to its foundation of continuous improvement (Hortale, 2003; C. B. Martins, 2018).

To better understand Qualis' current role, particular emphasis is placed on the two most recent national evaluations in which Qualis was adopted: The Quadrennial Evaluations of 2017 and 2022. The 2017 evaluation was crucial because the adopted assessment model was the culmination of two decades of minor adjustments, leading to an evident need for a more significant reform (Barata, 2019; CAPES, 2018b; PNPG Committee, 2018). This reform was planned and implemented in the following years, but put to the test in the evaluation of 2022 (Amado et al., 2020; CAPES, 2023b; PNPG Committee, 2020; Reategui et al., 2020; R. J. Ribeiro, 2022a; R. J. Ribeiro, 2022b). This latest iteration of Qualis reflects a more robust approach to the evaluation of journal publications, addressing some of the previous limitations while acknowledging its inherent strengths. However, the current Qualis system is still flawed, and while this study acknowledges its progress, it also pinpoints areas that warrant further improvement. Thus, building upon previous critiques (Barata, 2016; Barata, 2019; CAPES, 2018b; PNPG Committee, 2018; PNPG Committee, 2020), this study proposes alternative strategies to address the remaining weaknesses and improve the general reliability and efficacy of the Qualis system.

7.1 Qualis conception and early developments

Brasil (2023b) extensively examined the way in which several fundamental principles of the Brazilian evaluation system emerged as a result of resource constraints. For almost half a century, the primary unit of assessment within the national evaluation framework has been the graduate program (PPG). This was due to the expansion of the Brazilian National System of Graduate Education (SNPG), which experienced a growth in the number of PPGs, master's and doctoral students, and scholarships available. Despite this expansion, the personnel responsible for processing all concessions remained limited in number. Consequently, a decision was made to delegate the assessment of individuals to their respective graduate programs. CAPES would evaluate PPGs, award them with a specific number of scholarships, and then allow programs to carry out their own internal selection process to determine the allocation of available

grants. This model persists today and was adopted similarly when the Qualis system was first developed in 1998.

During that time, the Brazilian evaluation went through a significant restructuring. For example, a standardised evaluation form was adopted to evaluate graduate programs (PPG) in various disciplines, resulting in more consistent and comparable outcomes (Monteiro et al., 2019). Abílio Baeta Neves, CAPES' president during the reform of the evaluation system, discussed part of the motivation behind the changes in an interview with Ferreira and Moreira (2002). Neves explained that Brazil's PPGs had already achieved a reasonably high quality level, and many of the adopted indicators made it easy for a large number of these programs to achieve the highest possible evaluation score. For example, he cited the number of faculty members with doctoral degrees as an indicator used in the 1996 evaluation. However, a simple analysis revealed that faculty without doctorates were already scarce, rendering the indicator obsolete. Consequently, if no new criteria were introduced, efforts to differentiate programs could be compromised. Therefore, a decision was made to improve the evaluation of the quality and international integration of graduate education. Attaining this goal required replacing indicators such as the mere number of published articles, used at the time, with more significant metrics.

Qualis was the answer to the pressing problem. Taking into account the lack of resources to qualitatively assess all the paper production in the country, CAPES opted to classify scientific publishing outlets, assuming that articles accepted by indexed journals with a peer review system would guarantee a certain level of quality (Barata, 2016; Glänzel and Moed, 2002). The aim of Qualis was never to become a journal ranking. The idea, as highlighted by Neves in Ferreira and Moreira (2002), was to identify journals that should count as having scientific relevance in the national or international scenario, distinguishing what circulates knowledge and matters in each field from what serves as an auxiliary tool in graduate training and qualification, even though such dedicated publication channels can be vital for the development of scientists-in-training.

Qualis was then created with the primary purpose of classifying the journals listed by graduate programs in the annual data collection system used by CAPES to map the work conducted by faculty members and graduate students. The first classification system assessed journals in two main dimensions: quality or relevance in a specific scientific field (A for high, B for average, or C for low);

and their circulation (1 for international, 2 for national, and 3 for local). Three additional rankings were added to the nine possible combinations: *SR* – which means that the areas did not have enough information to classify the journal; *IP* – improper, meaning that it was not considered a scientific journal by the area committees; and *NC* – not classified by the area (CAPES, 2003).

Therefore, CAPES would collect the complete list of publications from graduate programs throughout the country and make them available to each evaluation area committee for classification. The areas had some flexibility with respect to the evaluation methods adopted, but taking into account the general guidelines provided by a council of area and agency representatives (CTC-ES). As expected, evaluations would rely more on database indicators in the areas of science, technology, engineering, and mathematics (STEM), and qualitative methods for those in the social sciences and humanities (SSH) (Barata, 2016; CAPES, 2003; Soma et al., 2016). Some additional characteristics of the original Qualis are:

i) Qualis is not a comprehensive list of journals. It contains only those with publications reported by graduate programs during each evaluation cycle;

ii) Classification is *ex post*, so journals are ranked according to the assessment performed after the publications are reported to CAPES by graduate programs. No *ex ante* component is present in the classification, so no expectation of future performance can be derived from a Qualis result;

iii) Qualis is a temporary list, not a cumulative one. That means that classifications from one cycle are not transported to the following one, meaning that some journals will leave the list, others will be included, and those that remain may receive a different classification;

iv) Journals can have multiple classifications across evaluation areas, as the same journal can be used to publish papers from graduate programs in different areas, and each committee conducts an independent analysis.

Taking into account the premisses and characteristics listed for the original Qualis, its primary purpose is reinforced and it becomes clear that "Qualis should not be considered an appropriate source for classifying the quality of scientific journals for purposes other than the evaluation of graduate programs" (Barata, 2016, p. 17). As a consequence, PPG should not use Qualis to hire staff, as the candidates' publications may have taken place in journals that are not in the current Qualis. Qualis should also be used with caution when selecting

journals to publish, as a journal classified in the top strata in one evaluation cycle might not be granted the same level in the next. However, some evaluation areas progressively incorporated other purposes for Qualis, such as making select journals more attractive for prospective authors by artificially inflating their ranking, or adding journals not reported in the data collection to the list, also aiming to stimulate publications in journals considered important in each field (CAPES, 2003).

With some of these distortions already being incorporated to the basic premisses defined in its beginnings, Qualis was used for a whole decade undergoing just minor evolutive adjustments after each new evaluation cycle. After the 2007 national evaluation, CAPES considered that it was time for a more significant change in the classification.

7.2 Reviewing Qualis for a new phase

The original Qualis was conceived around a major turning point for scientific publishing. The end of the 20th century was a dynamic period that reshaped publishing and citation practises. CAPES, for example, launched its Portal of Journals in 2000, an online platform that provided graduate programs throughout the country with access to 1.419 digital journals, a number that would multiply to tens of thousands in the following years (Brasil, 2020). With the digitalisation of scientific publishing, classifying journal publications on a scale that considered local, national, or international circulation made little sense. Thus, the Qualis classification was restructured after extensive discussions within CAPES and the evaluation area committees. The main change was the replacement of the double scale for circulation and quality with a new single scale of seven strata: *A1*, *A2*, *B1*, *B2*, *B3*, *B4*, and *B5*, plus an additional stratum *C*[1] for publications that did not meet the minimum criteria established in each area (Soma et al., 2016). The following rules applied to the new scale, which continued to be independently attributed by each evaluation area.

[1] Some area committees made use of an additional stratum, *NP*, to classifly journals considered not to be of scientific nature. However, since both *C* and *NP* journals would be excluded from the calculation of any indicator used in the evaluation process, many areas did not care to make a distinction between the two strata, which have been used irregularly over the years. For this study, both strata will be unified as *C*, as it is the case in the CAPES Qualis reports available at https://qualis.capes.gov.br/.

i) Fewer journals should be classified as *A1* than as *A2*;

ii) The number of journals in *A1* + *A2* can account for a maximum of 25% of the journals listed in the area;

iii) Similarly, *A1* + *A2* + *B1* cannot add up to more than 50% of the journals;

iv) All strata must be populated, with only *C* as a possible exception.

The rules behind the new Qualis scale reveal that the previous model may have led to overpopulation of the upper stratum in some areas. This becomes clear, as the review is said to be primarily motivated by the need to recover the gradual loss of discriminatory power experienced over the years (Barata, 2016; Soma et al., 2016). Although every journal could be ranked *A* on the previous scale, the new top *A1* rank was limited to be around the 12% percentile. *A2* would include the remaining top-quartile journals and *B1* those above the median. Some discriminatory power was regained, even though it was kept at a level of detail sufficient for the evaluation of graduate programs, but broad enough so that Qualis would not become a competitive ranking.

Despite additional advances in data collection, little else changed in the new Qualis. Old rules remained in place and old problems persisted. For instance, an exception to the original Qualis purpose allowed areas to inflate rankings of journals they wanted to promote, and committees also used the mechanism to devalue journals due to issues such as pertinence, relevance, and adherence to the area. Qualis became a classification of more than quality within the specific purpose for which it was intended. This has led to inconsistencies, such as in the case of journals with completely distinct classifications between areas.

An example of this issue is shown in Table 7.1, which shows a multitude of different Qualis classifications attributed to "Evaluation: Journal of Higher Education Evaluation", a multilingual Brazilian journal founded in 1996. Operating under the Diamond Open Access model (free for authors to publish without any article processing charges), the journal is indexed in databases such as the Scientific Electronic Library Online (SciELO) but not in the more international ones such as Scopus and Web of Science. Although the journal has been valued as *A1* by the evaluation area of Education since the 2010 edition of Qualis, the same is not true for other areas. Table 7.1 lists the various strata attributed to the journal in the 2017 Qualis, together with the number of papers published by each area in the 2013–2016 evaluation period (CAPES, 2023d).

Table 7.1.: Example of a journal with multiple strata in the Qualis 2017 classification

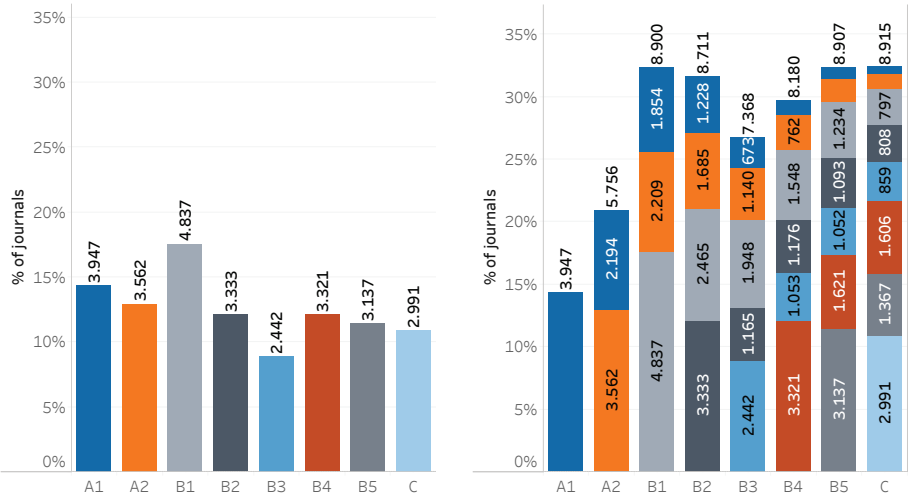
Journal	Stratum	Evaluation area	Papers 2013-2016
Evaluation: Journal of higher education evaluation ISSN: 1414-4077	A1	Education	72
	A2	Literature and linguistics	3
		Interdisciplinary	32
		Teaching and learning	3
	B1	Arts	1
		Journalism and information	2
		Social work	1
		Sociology	1
		Agricultural sciences	3
		Environmental sciences	2
	B2	Physical education, therapy and rehabilitation	2
		Public health	1
		Business and administration, accounting and tourism	43
		Economics	1
		Psychology	3
	B3	Engineering I	1
		Dental studies	1
		Political Science and international relations	1
	B4	Engineering III	5
		Medicine II	1
		Nutritional science	1
	B5	Pharmacy	2
	C	Medicine I	1

Considering Qualis as a system that links strata to percentiles with its *A1*, *A2*, and *B1* limitations, it makes sense that areas such as “Business and administration...” would rather not give a ranking above *B2* to an Education journal, as the higher classification could be used for a journal with closer connections to the area. However, it is difficult to consider adequate that areas are allowed to challenge the quality assessment of a discipline about one of its own journals, to the extreme of ranking it as *C*, like Medicine I has done for "Evaluation...". With that, the area is saying that the journal is not considered good or relevant enough to even be considered in the evaluation process, as *C* journals do not add any value to the assessment of a PPG’s scientific production.

Furthermore, while the example in Table 7.1 refers to a locally relevant journal, the problem affects even internationally established journals. For instance, the official Qualis 2017 results reveal that while “Science” has been classified as *A1* in most areas, it was also classified as *A2* in Economics and in “Business...”, and *B1* in Law.

Although areas may be able to demonstrate a variety of methods used in their journal classification and they may also justify any posterior adjustments they

have performed, the extreme variations across all areas weaken Qualis as a trustworthy instrument in the national evaluation performed by CAPES. The extent of these variations can be seen in Figure 7.1, with (a) showing the distribution of the best classification granted to each journal in the 2017 Qualis, and (b) showing all classifications obtained per journal.



(a) Best classification per journal

(b) All classifications per journal

Figure 7.1.: Strata distribution of the Qualis 2017 classification of journals

Figure 7.1a displays the highest rank achieved by academic journals in the 2017 Qualis. The figure presents a distribution of 27.570 unique ISSN numbers that combine the results of the 49 evaluation areas. However, since the number of journals varies by area, the distribution of the first three strata deviates slightly from the *A1*, *A2*, *B1* rules. For example, almost 15% of the journals were rated *A1* by at least one evaluation area. *A2* is close to what was expected, as is the large percentage of journals in *B1*, which could even be higher since it represents the entire second quartile of journals. Figure 7.1b shows the same data as Figure 7.1a, but with all other journal classifications stacked on top of the highest-ranking bars. For example, of the 3947 journals classified as *A1* in at least one area, 2194 were ranked as *A2* by one or more areas, while 1854 were also classified as *B1*, etc. Reaching the lowest stratum, *C*, are 158 journals. The same pattern can be observed for all other rankings.

Although CAPES (2023b) recognises that the multiplicity of strata for the same journal was the biggest challenge offered by Qualis, the agency defended the process, arguing that diversity in classifications was not an inconsistency, but a reflexion of how each journal was valued with respect to its relevance to each evaluation area. Technically, the argument is correct, and the criteria and indicators used by each of the 49 evaluation committees to build their Qualis lists are publicly available in detailed reports, published before the classification takes place and after discussion and approval by the council of area and agency representatives (CTC-ES) (CAPES, 2020d; Soma et al., 2016). However, even the evaluation director at the time of the 2017 Qualis classification states that "there is no reason why the same journal receives classifications that are so disparate across areas" (Barata, 2019, p. 5). Undoubtably, Figure 7.1b shows not only how disparate these classifications could be, but also that they are more a rule than an exception.

7.3 Towards the current Qualis system

In 2015, CAPES started a new effort to reform Qualis, appointing a working group to study the classification and propose changes to its methodology. The WG presented its diagnosis and a series of recommendations in the following year. Some of the main proposals, as reported by Pascutti (2016), were as follows:

i) Encourage dialogue towards more homogeneous evaluation criteria and classifications, so that the disparity of strata across areas would not be so extreme;

ii) Avoid excessive emphasis on journal adherence to the areas, limiting stratum variations that exceed one level (i.e., if the journal quality would grant it the *A1* rank, lack of adherence to the area would warrant an adjustment to *A2*, but not to further strata such as *B5* or *C*);

iii) Adopt reference areas for journals, when possible. That means the stratum attributed by an evaluation area to its main journals would serve as reference for the evaluation in other areas;

iv) Evaluations should be based on qualitative and quantitative criteria, rather than subjective ones;

v) Introduce bibliometric indicators in fields that do not yet adopt them. Even if as a secondary criterion, they could encourage PPG to publish in indexed journals and stimulate the indexing of national journals.

The changes proposed by the working group were not implemented at that time, as the national evaluation for the 2013-2016 quadrennium would be held in 2017, so no change in procedures would be adviseable near the end of the assessment period. However, the debate was resumed shortly after the 2017 evaluation results were made public, and a new working group was appointed, with the aim of continuing the effort and proposing changes to be implemented in the next Quadrennial Evaluation (CAPES, 2018d). The proposed new model would not only align with the main recommendations of the previous working group, but would take them further. The main principles of the new Qualis are described in a technical report (CAPES, 2023b), and can be summarised as follows.

i) Qualis becomes, as originally intended, an instrument for the evaluation of graduate programs through the classification of journals based on quality. Pertinence, relevance, and adherence are no longer part of the analysis, which should rely on objective indicators;

ii) Journals are no longer classified in multiple strata across evaluation areas. Now, each journal has a unique classification;

iii) A journal is classified by the area it aligns with most closely, denominated its "mother area". This designation is determined by the volume of articles published in the journals between 2013 and 2019, although with allowance for shifts as negotiated between evaluation areas;;

iv) A new scale was adopted, expecting a more balanced distribution of percentiles between strata: *A1*, *A1*, *A2*, *A3*, *A4*, *B1*, *B2*, *B3*, *B4*, and *C*, which remained for those journals with zero value for the evaluation.

v) The previous restrictions on the proportion of journals per category were lifted. This means that the evaluation areas were no longer bound by guidelines such as ensuring that the sum of journals rated *A1* and *A2* didn't exceed 25% of the total journals in the area.[2]

[2] As will be discussed later in this study, while the restriction on journals allocated to each strata was removed from the Qualis methodology, the change was somewhat offset by the introduction of CAPES adoption of a suggested classification derived from bibliometric indicators, which allocates 12,5% of the total number of journals to each of the eight strata, from *A1* to *B4*.

A first look at the Qualis classification released in early 2023 (CAPES, 2023d) reveals some positive and negative consequences of the new model. For example, Figure 7.2 shows a distribution matrix of the journals according to the mother areas established in the evaluation and published in CAPES (2023c). The numbers in rows and columns represent the codes of the evaluation areas, according to what has been reported in Brasil (2023b), and the areas are grouped according to the three broad areas in the CAPES classification.

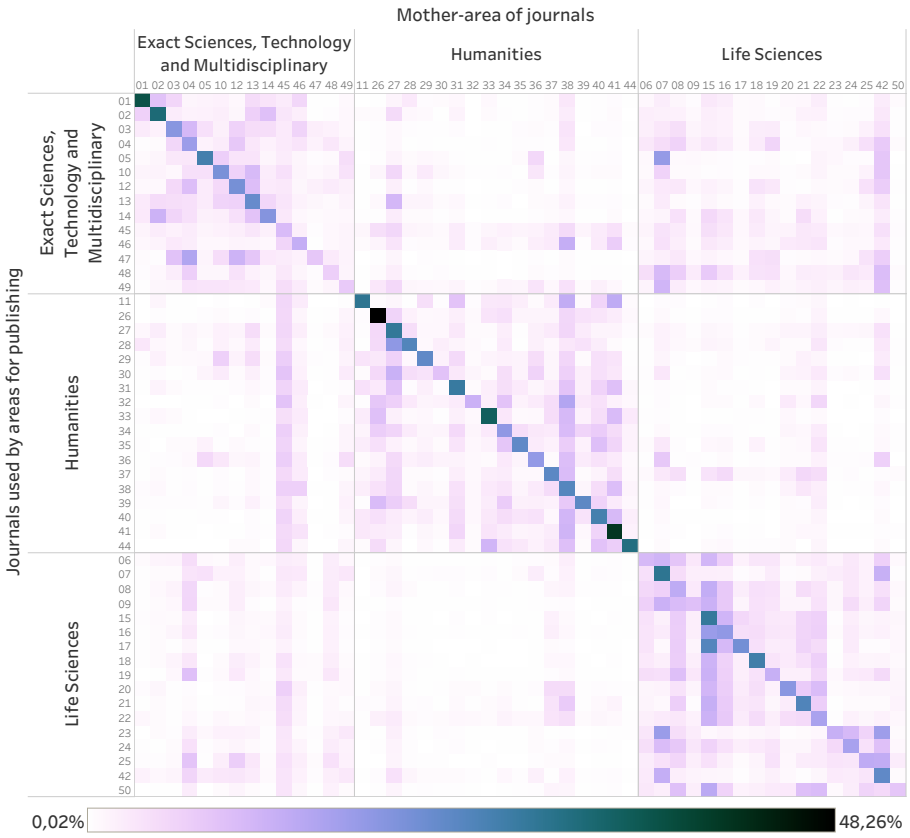


Figure 7.2.: Distribution of journals according to mother-area in the evaluation and publishing (2017-2020)

At the top of Figure 7.2, you see the mother areas, so the journals displayed in the columns represent those that were classified by these areas. On the left you see the areas publishing in journals that were classified by the areas on top. The

darker diagonal, as expected, reflects the number of journals that each area used to publish and classified as the mother areas. For example, Law (26) classified a total of 1418 journals, which represent around 48% of the universe of journals used by its graduate programs during the 2016-2020 period. Law also published in many other journals, for instance, in 153 classified by Education (38), 113 by Sociology (34), 96 by Philosophy and Ethics (33), etc. Navigating on the Law column, it is possible to see that area 39, which is Political Science and International Relations, published in 209 of the journals that were classified by Law as their mother area.

Although Law is the area that classified the highest percentage of journals used by their graduate programs, Biotechnology (48) classified the least, being the mother area for only 6,4% of the journals used by its PPG in the period (267 journals). Evidently, as an interdisciplinary field, biotechnology research can fit many journals within areas such as Agricultural Sciences (42), Biodiversity (07), Chemistry (04), and others. The interactive matrix is available at https://andrebrasil.github.io/viz/qualis.html, where each intersection shown can be explored in detail. Furthermore, it is possible to change the visualisation from journal to published articles, for example, revealing that more than 88% of the articles published by Law were in the journals they classify.

Returning to the example presented in Table 7.1, the mother area of the journal "Evaluation: The Journal of Higher Education Evaluation" in the most recent Qualis was "Education". Thus, as a result of the change in approach that guarantees that a single stratum will be assigned to a journal by its mother area, the journal was granted the *A1* stratum. The classification was then used for the evaluation of the area, but was also applied to 18 other areas with published articles in the journal during the 2017-2020 quadrennium.

However, "mother areas" are not always absolute and independent in these decisions, in part because they can also have "sister areas". This happens if the area with the most publications in a journal does not include at least 50% of the journal's publications in the period of analysis. In that case, the area with the most publications is still the mother, but any decisions regarding the classification of the journal must be agreed with the sisters, which can be up to three other areas with a significant number of publications in the journal.

Furthermore, although some key concepts of the new Qualis were not polemical, such as the new scale, the same did not happen with changes like the

single classification for journals and the methodology, or even the existence, of mother areas. Even during preliminary discussions, CAPES received many manifestations criticising ideas being explored to review Qualis. For example, a letter from the Brazilian Association of Collective Health (ABRASCO, 2019) mentioned that the number of publications was not sufficient to determine the motherhood of a journal and highlighted how interdisciplinary areas would lose much control over the classification of journals relevant to them. After the publication of the classification used during the Quadrennial Evaluation of 2022, numerous groups complained about the changes and results, and most of the criticism focused on the adoption of a single strata for a journal in all areas (Brigatti, 2023; Ferrari, 2023; Yamashita, 2023).

Some manifestations may be important for the evolutionary dynamics of evaluation. For instance, the letter by ABRASCO (2019) pointed out that the initial proposal to introduce "mother areas" considered only two years of publications for the decision on how journals should be distributed, but that a longer period should be adopted. As a result of constructive criticism, CAPES (2023c) described the final methodology that considered seven years of data for distribution. However, some other manifestations should be mainly ignored, for instance, the complaints of some Economics researchers who cannot accept that journals such as "Religion Studies" and "Experimental Dermathology" could be granted the same *A1* that is used for elite Economics journals (Brigatti, 2023). Why would it not be possible to accept that quality exists beyond the borders of one's discipline?

7.3.1 The challenge of an indicator-based Qualis

Right after the conclusion of the Quadrennial Evaluation of 2017, CAPES instituted many working groups to reflect on ways to improve the evaluation model for the next cycle. A seminar series branded "Rethinking Evalution" was organised to broaden the discussion and allow the groups to present preliminary findings and debate them with the evaluation community in Brazil. The second seminar of the series focused on the evaluation of scholarly production, including books, events, technical and technological production, and more (CAPES, 2018b). The presentation by the Working Group on Qualis Journals presented their main findings and the proposal for a new Qualis, which was very close to the final format that was implemented.

Regarding the use of bibliometrics, the group showed that indicators such as Scopus CiteScore and the Journal Impact Factor had a very high correlation with previous evaluations in most areas in "Life Sciences" and "Exact Sciences, Technology, and Multidisciplinary". As described in Brasil (2021b), these areas are better covered in databases such as the Web of Science. Therefore, well-established indicators should indeed correlate with the classifications performed by the areas that traditionally mix metrics with peer review in their analysis.

Based on the findings of the working group, a proposal was presented for a methodology based on the adoption of a common set of indicators for all areas. No agreement could be found as the classifications in a significant share of the areas, particularly within the Social Sciences and Humanities (SSH), did not correlate with the proposed metrics. After months of debate, a new working group was established to deal with the specificities of journal classifications in the SSH (CAPES, 2019c), and the original group was reinstated with its scope limited to "Life Sciences" and "Exact Sciences, Technology, and Multidisciplinary" (CAPES, 2018e).

As the new working groups worked on their proposals for a Qualis classification that included the use of open metrics that supported the evaluation of different areas, CAPES organised midterm seminars with the participation of graduate program directors (GPD) of more than 4500 PPG in the country. During a three-month period, each of the 49 evaluation areas met with their respective GPDs to diagnose the status of their areas and programs midway through the four years of the evaluation cycle (CAPES, 2019b). For that, a provisional Qualis was prepared, using for the first time the concepts of "mother area", unique journal classification, and the use of the mentioned indicators to serve as a reference for area committees in a peer review phase of the process.

The process worked as follows: CAPES would rely on indicators to suggest a stratum for the classification of each journal in the analysis. The evaluation area committees would receive the recommendation, and they could either accept or reject each proposed classification, reclassifying the journals under their purview according to the methods of preference in each area. Figure 7.3 shows how the area committees dealt with the suggested classification, displaying the indicator-recommended classifications at the top of the graph, and bars showing the percentage of recommendations that were validated (highlighted in orange) or reclassified during the peer review phase.

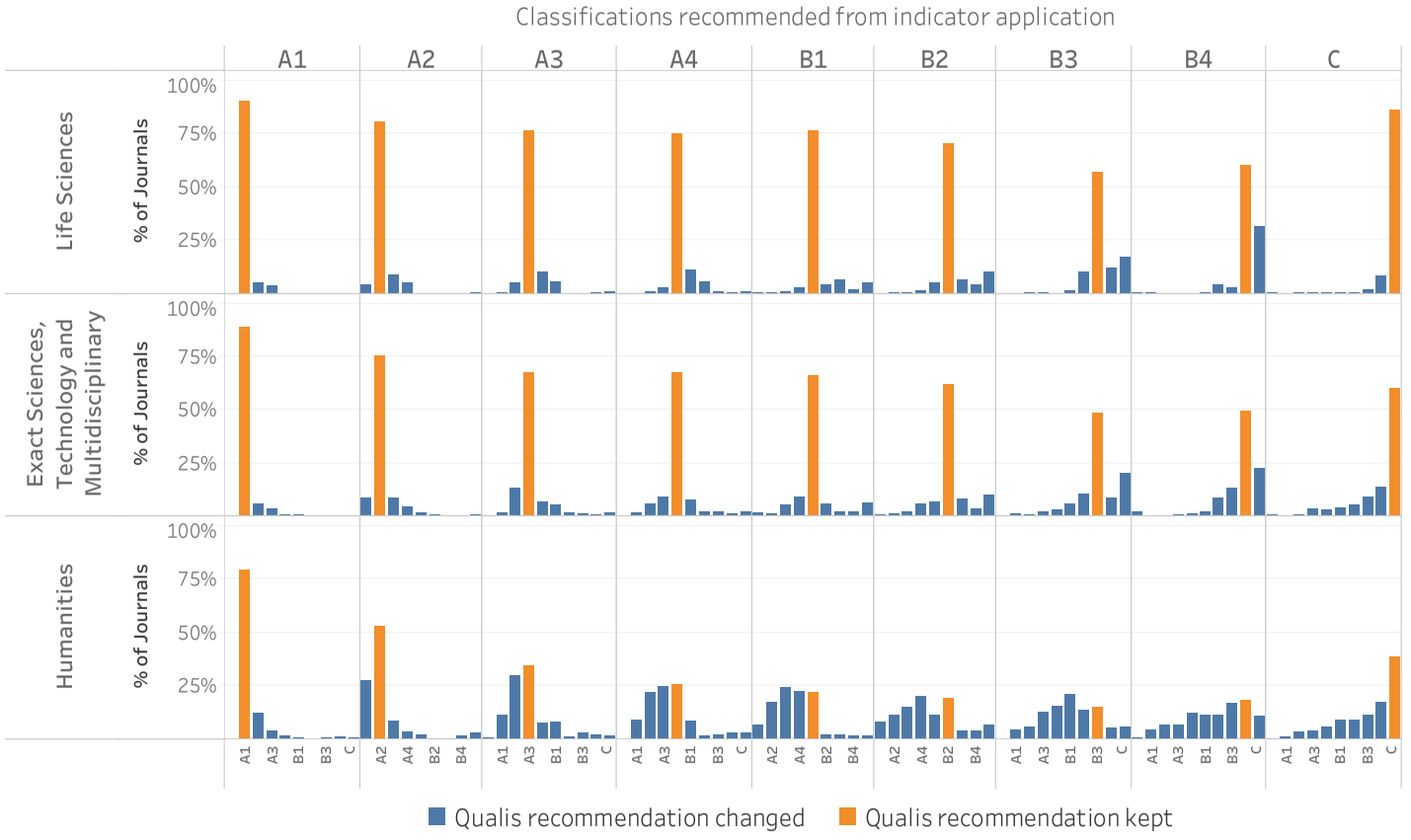


Figure 7.3.: Mid-term reclassification of journals from the indicator-based suggestions to the peer review results, grouped by broad areas

Figure 7.3 shows how the indicators used were able to capture what the committees in "Life Sciences" and "Exact Sciences, Technology, and Multidisciplinary" consider as a measure of the quality of the journals, particularly in the upper strata. Around 90% of the journals with a suggested classification of *A1* were confirmed by the committees as such, and a minority was reclassified mainly as *A2* or *A3*. The correlation continues high in the following strata, slowly decreasing as it reaches the bottom percentiles. Even then, for the B4 and B5 classifications, around 60% of the recommendations are kept in "Life Sciences", and 50% in "Exact Sciences...". In both cases, even the journals listed as *C* were considered to be mostly correct by the evaluation committees.

The panel for the "Humanities", including all of SSH, is quite distinct. Although there seems to be a good correlation with indicators in the top *A1* stratum, with nearly 80% of the indicator-based suggestions kept by the committees, the percentage drops drastically to 52% in *A2*, 34% in *A3*, and then is completely lost as a guide in the next strata. Regarding the stratum *C*, where journals were assigned when no indicators were available, more than 60% of them were reclassified by the "Humanities", some even reaching the stratum *A1*.

What is shown at Figure 7.3 can be considered a good result. Indicators indicate; they do not determine final scores, which are a prerogative of peer review committees. Through the experience and knowledge of the evaluators within their disciplines, the ability to challenge the indicators when necessary is a crucial benefit of the Qualis model as implemented during the mid-term evaluation. However, reclassification within "Humanities" may be considered excessive if two of the core recommendations of the 2015 working group are taken into account: Evaluations should apply qualitative and quantitative criteria, rather than subjective ones; Bibliometric indicators should be used in fields that do not yet adopt them. The results suggest that indicators may be mostly ignored in the process, and subjectivity may also be an issue of concern.

The mid-term Qualis was an exercise to test the methodology under development. With the ongoing work of both Qualis working groups, the exercise could contribute to the improvement of the methods, plus the addition of others useful for the SSH. Together, these approaches could make the final reclassification look more like the one in the "Exact Sciences..." for all three groups.

The final Qualis methodology was approved by CAPES in September 2020, one year before the Quadrennial Evaluation of graduate programs was scheduled to

occur. The new process is detailed in CAPES (2020a), and the model reflected the recommendations of the working groups in the "Humanities" (Amado et al., 2020) and "Life Sciences" and "Exact Sciences..." (Santos et al., 2020). Some key results detailed in the final document relate to the calculation methods, which can be summarised as follows:

i) Journal percentiles are collected from Scopus CiteScore and Clarivate's Journal Citation Reports (JCR). Journals are ranked according to percentiles by comparison, after normalisation of their impact factors within 334 subsubjects (Scopus) or 235 categories (WoS).

ii) Google Scholar's h5 is used to calculate percentiles equivalent to those from Scopus and WoS, in case the journal is not covered by any of those. The imputations are calculated using regression models described in the documentation. Sometimes, a cap is applied to the results to avoid overvaluing journals considered not to be international enough, meaning that they are not indexed by the two preferred databases.

iii) The classification of "Humanities" journals considered:
   - Existence of subareas for normalisation;
   - Indexing by discipline-specific databases (i.e., "Business..." used indicators from the Scientific Periodic Electronic Library – SPELL);
   - Optional use of the impact factor or reputation indicators;
   - Consideration of publication languages in normalisation procedures;
   - Use of h5 or h10, to better reflect citation practises in each area.

iv) The stratum *C* will include all journals that do not have any of the indicators adopted by the model, or that do not meet good publishing practises.

The presented selection of indicators is far from groundbreaking, going against principles from responsible research movements represented by the San Francisco Declaration on Research Assessment – DORA (ASCB, 2012) and the Leiden Manifesto for research metrics (Hicks et al., 2015). Furthermore, the proposal also takes Brazil in the opposite direction of evaluation systems like the one existing in The Netherlands, where indicators such as the Journal Impact Factor and the h-index are banned or restricted (VSNU et al., 2020).

Nonetheless, CAPES implemented the calculations to provide the classification suggestions used for the 2016-2020 Qualis. Figure 7.4 shows how the evaluation committees dealt with the suggestions.

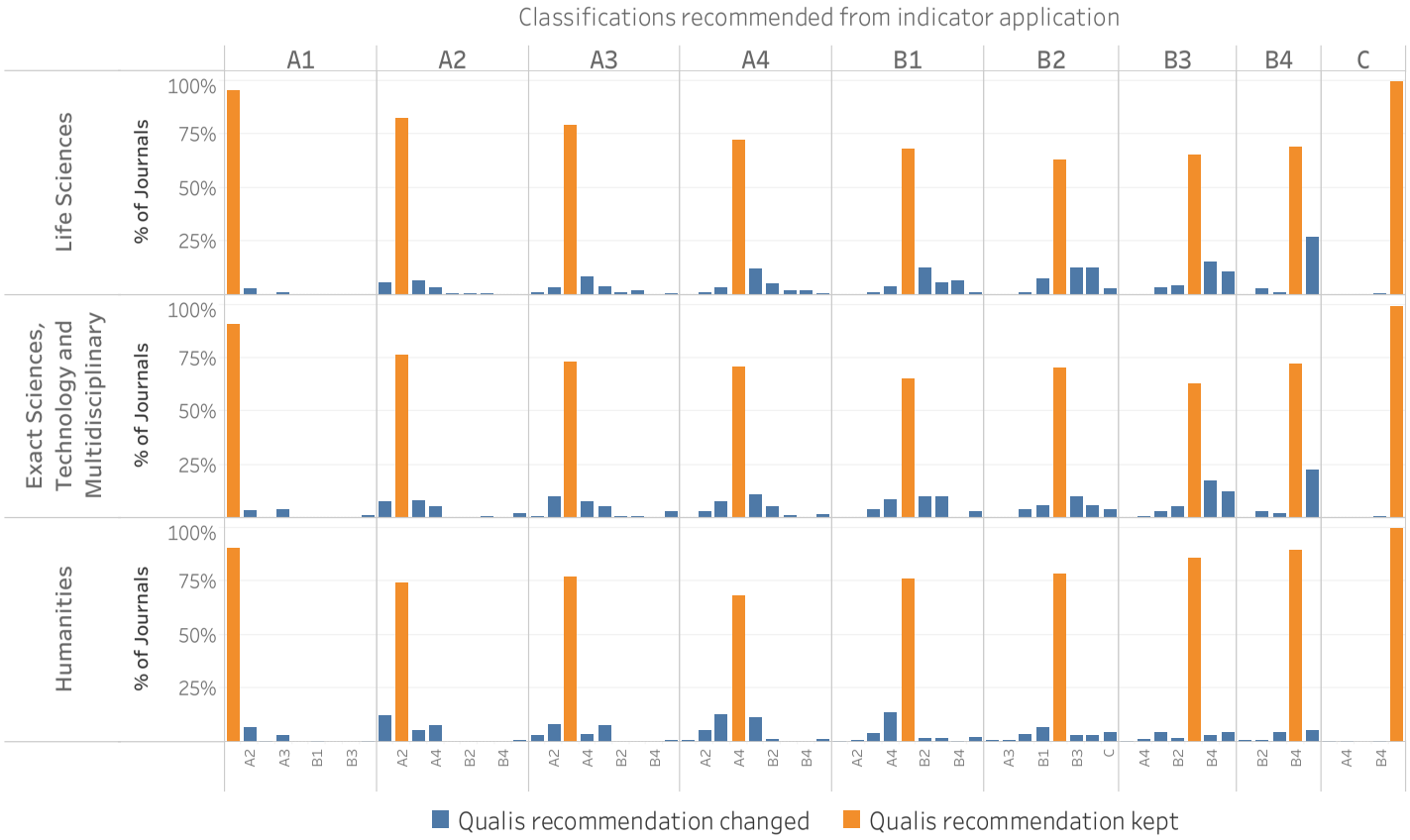


Figure 7.4.: Final reclassification of journals from the indicator-based suggestions to the peer review results, grouped by broad areas

Figure 7.4 is surprising at first glance, especially when compared to Figure 7.3. The chart covering the final 2016-2020 Qualis shows that committees in the three broad areas accepted most of the indicator-generated journal classifications offered by CAPES. Although little has changed for "Life Sciences" and "Exact Sciences...", the "Humanities" went from questioning and reviewing most indicator classifications to a level of agreement even higher than that of the other two groups. Could CAPES have found the perfect journal ranking system that makes expert review nearly irrelevant? A deeper investigation of the methodology described in CAPES (2020a) reveals that this is not the case.

According to CAPES (2020a, p. 10), "Each area can adjust its classification, observing a maximum of 30% of changes to the strata, of which 20% can be made in up to 1 level and 10% in up to 2 levels". Therefore, the apparent alignment between indicator classification and expert analysis seen in Figure 7.4 is more a representation that areas are not free to reclassify journals as needed. The limitations imposed prevent us from seeing the level of agreement between indicators and evaluators in the final Qualis, as was possible in its mid-term version shown in Figure 7.3.

7.4 The persistent need to evolve

There is no doubt that the latest iteration of Qualis has been an advancement. Historical problems derived from multiple and extremely diverse classifications of the same journals have been solved. With the mother area approach, evaluation can focus on quality and the classification ceases to represent a one-stop system where a magic stratum should capture quality, relevance, and pertinence. Finally, the adoption of an indicator-based reference that can be reviewed by experts has the potential to be a step forward, despite the evident flaws observed; that is, better indicators should be considered in future iterations of the classification, and the limited room committees have to question such indicators should be revisited.

However, while the discussion on seeking further evolution is relevant, it is also important to reflect if there is still a need for a Qualis at all. Arguments for the end of Qualis find support from many Brazilian academics, and the idea has been registered by a special committee in charge of monitoring the most

recent National Plan for Research and Graduate Education (PNPG) in Brazil. In their final report, with reflexions on Qualis as part of research evaluation in the future, the committee stated:

> Qualis has played an important role in the qualification of graduate production. Given the new resources and features in IT tools and the consolidation of databases that have emerged in the time between the last triennial evaluation and this moment that we approach the second quadrennial evaluation and the maturity achieved in the SNPG, we have a new context. In this context, the evaluation of production in each evaluation area should be based on established international metrics widely accepted by the community. In this sense, the Committee proposes the extinction of Qualis after the next evaluation cycle (2021-2024) (PNPG Committee, 2020, p. 25).

There are three important points in the recommendation of the committee. The first relates to the available technology and the advances in tools. Indeed, while Qualis evaluation at the journal level has been a manifestation of a lack of resources and data to directly evaluate published papers, now there are not only resources, but also sophisticated indicators that can be used for a quality assessment that is a certain improvement over the limitations of journal evaluations (Polonsky and Whitelaw, 2005).

The second point relates to what is forgotten by the committe when they recommend using established international metrics, and to what they mean when talking about databases. As highlighted by (Brasil, 2021b), databases such as the Web of Science have uneven coverage of Brazilian publishing in evaluation areas, failing to include a significant part of the papers published in Portuguese, which are important for communicating with local audiences on topics of local relevance. It is wrong to believe that only science with an international impact can be of value, and evaluation must be responsible to not push researchers away from local impact, forcing them to chase metrics under penalty to see their work unrecognised and unrewarded. Databases do not need to be limited by Web of Science or Scopus, as there are alternatives with a better coverage, such as Dimensions and OpenAlex, that can be used for more detailed and relevant publishing analyses.

The third point is whether Qualis should be extinguished. If Qualis is seen as a ranking that reduces the complexity of scholarly publishing to a set of simple strata, constantly misused beyond its intended application of helping conduct a complex, multidimensional evaluation of graduate courses; then, yes, it may be time for Qualis to disappear. However, Qualis is a broader strategy to evaluate scientific production. Today there is a Qualis for books, another for events, for artistic and for technical/technological production. As with the Qualis journals, they are all based on applying a combination of quantitative and qualitative methods, combined with expert analysis, to help separate the good from the bad in the challenging process of evaluating graduate programs in Brazil.

Brazil cannot abandon a system that allows the valuation of journal production of national and international relevance, and also of a broader set of products that fit within DORA recommendations for responsible evaluation (ASCB, 2012). Qualis should remain as a means of combining the best available expert review with methods and indicators that can be applied to Brazilian reality, considering the scale of the SNPG and the limitations it imposes on an evaluation system. To abandon that in exchange of the suggested international metrics will probably hurt the country's efforts towards a science system that aims both to expand the frontiers of knowledge and produce societal impact.

7.5 Conclusion

This chapter has delved into the intricacies of the Qualis system, analysing its historical development and implications as a crucial part of evaluating research and graduate education in Brazil. Qualis was established around 25 years ago to incorporate a qualitative dimension into the assessment of the country's expanding scientific production. Considering that resources were limited, an article-level analysis of scholarly production was not feasible, so Qualis has been an imperfect but necessary solution to classify scientific journals as a proxy of the expected quality of articles published by Brazilian researchers.

Although Qualis has experienced numerous changes since its inception, it continues to grapple with the challenge of balancing objectivity and precision with the unique characteristics of scholarly work across various disciplines. In its latest iteration, Qualis addressed some of its notable historical shortcomings.

For instance, different evaluation areas used to classify journals into distinct strata, factoring in not just quality, but also pertinence and relevance as ranking adjustments. While these factors are essential in an evaluation, it is improper to artificially elevate a journal's rank to incentivise more submissions, despite its diminished quality. Similarly, it is unjust for areas to downgrade a journal's ranking merely because it is from a different discipline.

The new Qualis now empowers evaluation areas with the responsibility to classify the journals that belong to them, and the resulting ranking is unique, being adopted in the assessment of all areas. The advance created a new challenge, as the methods adopted for classification in different areas must be comparable and valid across all disciplines. For that, quantitative methods are now applied, relying on indicators such as CiteScore, JCR and Google Scholar's h-index to suggest classifications to expert committees, which can make the necessary adjustments through a peer review process. However, the quantitative approach seems to have stripped the committees of their voice, since only 30% of the calculated rankings can be modified in the peer review phase, with any modifications also being limited in range.

However, the importance of having a framework that enables the valuation of scientific production should not be underestimated. While some call for the extinction of Qualis and subsequent replacement by international metrics from which many developed countries try to escape, the system remains vital to foster an inclusive and equitable academic landscape in Brazil. Although established international metrics offer greater objectivity and acceptance for the evaluation of scientific production, they may fail to capture the unique context and challenges faced by Brazilian researchers and graduate programs. Furthermore, an overemphasis on international metrics can discourage researchers from focussing on locally relevant research, often produced in the local language, which is crucial to address the needs and concerns of the Brazilian society.

Moving forward, it is essential not to abandon, but to refine and enhance the Qualis system by drawing on the best available methods and indicators tailored to the Brazilian context, as well as recognising the importance of expert review. By doing so, we can work toward creating a more inclusive, equitable, and comprehensive evaluation framework that supports and values the diverse forms of scientific production that contribute to the growth and development of the Brazilian graduate system and the broader scientific community.

References

ABRASCO. (2019). Apreciação da proposta de Qualis Periódico Referência. Retrieved March 2023, from https://bit.ly/3mSLm2a

Amado, A. M., Tomanari, G. A. Y., Fernandes, L. M. R., Dalmonte, E. F., Magalhães, J. S. d., Siqueira, V. B. C., ... Oliveira, T. M. (2020). *Relatório do GT do Qualis Periódicos Humanidades*. CAPES. Brasília, DF. Retrieved April 12, 2023, from https://bit.ly/3KSUzAF

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

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

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

Brasil, A. (2020). Building a national system of research and graduate education: How did the university become the house of science in Brazil? *Revista Nupem*, *12*(27), 222–253. https://doi.org/10.33871/nupem.2020.12.27.222-253

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

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

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

CAPES. (2003). *QUALIS: Concepção e diretrizes básicas*. Brasília. Retrieved July 24, 2020, from https://bit.ly/3u77Iuy

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

CAPES. (2018d). *Portaria nº 150, de 4 de julho de 2018. Institui o Grupo de Trabalho (GT) do Qualis Periódicos*. Diário Oficial da União. Brasília, DF.

CAPES. (2018e). *Portaria nº 271, de 14 de dezembro de 2018. Institui o GT do Qualis Periódicos das Humanidades* (Ordinance No. 271/2018). Diário Oficial da União. Brasília, DF. Retrieved April 21, 2023, from https://bit.ly/3mW7bhh

CAPES. (2019b). Meio-Termo: refletir para avançar. Retrieved April 21, 2023, from https://bit.ly/41rui2k

CAPES. (2019c). *Ordinance nº 8, de 31 de janeiro de 2019. Institui o GT do Qualis Periódicos das Ciências da Vida e Ciências Exatas, Tecnológicas e Multidisciplinar* (Ordinance No. 8/2019). Diário Oficial da União. Brasília, DF. Retrieved April 21, 2023, from https://bit.ly/3mUUvXY

CAPES. (2020a). *Metodologia do Qualis Referência: Documento base aprovado no CTC 198, de setembro de 2020*. CAPES. Brasília, DF. Retrieved April 20, 2023, from https://bit.ly/41w96rW

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

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

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

CAPES. (2023d). Qualis Periódicos. Brasília, DF: CAPES. Retrieved from http://qualis.capes.gov.br

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

Comissão Especial de Acompanhamento do PNPG 2011-2020. (2020). *Proposta de aprimoramento da avaliação da pós-graduação brasileira para o quadriênio 2021-2024: Modelo Multidimensional*. CAPES. Brasília, DF. Retrieved from https://bit.ly/3wyR097

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

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

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193. https://doi.org/10.1023/a:1014848323806

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

Hortale, V. A. (2003). Modelo de avaliação CAPES: desejável e necessário, porém, incompleto. *Cadernos de Saúde Pública*, *19*(6), 1837–1840. https://doi.org/10.1590/s0102-311x2003000600027

Leite, M. F. B., Viana, A. B. N., & Pereira, G. G. (2010). Métodos quantitativos na avaliação da CAPES: Uma pesquisa bibliográfica. *FACEF Pesquisa - Desenvolvimento e Gestão*, *9*(2), 166–174. Retrieved from https://bit.ly/2ZEYg3X

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

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

Pascutti, P. (2016). Entendendo o Qualis Periódicos CAPES. Campinas, SP. Retrieved April 2022, from https://bit.ly/40leSuZ

Polonsky, M. J., & Whitelaw, P. (2005). What Are We Measuring When We Evaluate Journals? *Journal of Marketing Education*, *27*(2), 189–201. https://doi.org/10.1177/0273475305276842

Reategui, E., Pires, A. d. S., Carniato, M., & Franco, S. R. K. (2020). Evaluation of Brazilian research output in education: confronting international and national contexts. *Scientometrics*, *125*(1), 427–444. https://doi.org/10.1007/s11192-020-03617-z

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

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

Santos, P. J. P. d., Oliveira, T. M. d., Paixão, F. d. A. J., Pascutti, P., Amado, A. M., Oliveira, D. d. H., ... Guterres, S. S. (2020). *Relatório do GT Qualis Periódicos*. CAPES. Brasília, DF.

Soma, N. Y., Alves, A. D., & Yanasse, H. H. (2016). O Qualis Periódicos e sua utilização nas avaliações. *Revista Brasileira de Pós-Graduação*, *13*(30), 45–61. https://doi.org/10.21713/2358-2332.2016.v13.1128

Spagnolo, F., & Souza, V. C. (2004). O que mudar na avaliação da Capes? *Revista Brasileira de Pós-Graduação*, *1*(2), 8–34. https://doi.org/10.21713/2358-2332.2004.v1.38

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

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