An Acad Bras Cienc (2022) 94(Suppl. 3): e20210962 DOI 10.1590/0001-3765202220210962
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

ENGINEERING SCIENCES

Metric indicators for the evaluation of graduate programs in Brazil: from Qualis to multi-criteria

PEDRO HENRIQUE G.R.P. CASTRO & DELLY OLIVEIRA FILHO

Abstract: Brazilian scientific production has been mostly developed (90%) by Brazilian universities, mainly graduate programs, which must be assessed and ranked by the Brazilian government for their maintenance. The Qualis system is used for this classification by valuing the scientific production of graduate programs, stratifying journals and assigning grades. Several Brazilian researchers affirm that the Qualis system is inaccurate and subjective because it is carried out by a group of researchers. This work aimed to propose methods to evaluate Brazilian scientific production in order to improve the assessment of graduate programs through multi-criteria methods in addition to Qualis. The application of different metrics presented results significantly different from Qualis, including the over valuation of journals with a low international impact factor. The proposed metrics considered: (i) web of science impact factor of the journals; (ii) Citations from articles; (iii) Citations of citations, i.e., a new metric; and (iv) H-Index for researchers. It was proposed the multi-criteria method, composed of the mentioned criteria, in addition to Qualis. For the sample of researchers, it was demonstrated that multi-criteria methods can assess scientific production more accurately. Therefore, they are more adequate to assess the Brazilian graduate programs, considering several internationally accepted criteria.

Key words: Brazilian scientific production, Qualis, secondary citations.

INTRODUCTION

Between 2007 and 2011, Brazil contributed 2.59% (147,503) of the world's total scientific papers, placing the country as 13th in the global ranking (Leta et al. 2013). Even in constant growth, the Brazilian scientific production remained the 13th largest in the world, considering the production of international papers between the years 2011 and 2016. There has been growth for more than 20 years in the number of Brazilian papers indexed on the Web of Science (WoS) platform, which indicates the expansion of the academic research production (Cross et al. 2018).

Between 2011 and 2016, Brazilian researchers had an outstanding participation in papers involving 123 other countries, through co-authoring with international researchers. About a third of all Brazilian papers in this database from the same period had at least one international collaborator (Cross et al. 2018). It demonstrates the participation and contribution of Brazilian research in the international context and highlights the growth trend found by Cross et al. (2018), who reveal, through time series data from 2011 to 2016, the increasing percentage of papers produced by Brazilian researchers, with international collaboration.

In addition, it was evaluated the impact of Brazilian scientific output, measured in terms of citations per paper, which increased by 18% between 2011 and 2016. As a result, the Brazilian normalized impact factor soared from 0.73 to 0.86, considering that the unit normalized impact factor represents the world average for the period. Brazil produced some highly cited researchers and obtained approximately 1% of the papers among the top 1% most cited in the world for the year 2016. From different metrics, Brazil ranked third among the BRICS countries (Brazil, Russia, India, China and South Africa) in: (i) number of papers, (ii) citations and (iii) papers per capita. The numerical results of the scientific output from BRICS countries are shown in Table I (Cross et al. 2018).

The production of Brazilian scientific papers comes, for the most part, from universities, which are the main institutions involved in the paper production, since most of the authors are faculty and/or students related to graduate studies. The production of papers from public universities accounted for about 90% of the total in the 2011 to 2016 period (Cross et al. 2018).

Graduate programs in Brazil undergo a rigorous evaluation process. They are graded from zero to seven, which is the main criterion for the maintenance and allocation of funding. For such evaluation, the Brazilian public institution Coordination of Improvement of Higher Level Personnel (CAPES) created the Qualis system, a method used for classifying the quality of the output from scientific journals from all graduate programs. This classification is indirect, carried out through the analysis of the quality of the journals publishing the papers (CAPES 2013). The Qualis system classifies the journals into the following descending order strata, according to their score: A1, A2, B1, B2, B3, B4 and B5. The A1, A2 strata are considered to be of international level (Oliveira et al. 2015).

Thus, researchers are encouraged to publish in better-evaluated journals, or increase the number of papers according to Qualis, so as to improve the classification of Brazilian graduate programs (Gabardo et al. 2018). It is important to emphasize that the evaluation of scientific papers has been the most important parameter to assess the performance of a graduate program.

In the beginning, the classification of journals by means of the Qualis system considered their impact factor, but the evaluation committee, which determines the stratification of the journals, observed that the great majority of Brazilian journals would be negatively affected, as they have an impact factor of zero or almost zero. Thus, the journal evaluation committee maintained several Brazilian journals, with impact factor zero or close to zero, among the upper strata of Qualis (Kellner 2017). It seems that the Qualis management committee assigns a classification to the strata A1, A2, B1 to B5 that is not consistent with citations on an international basis, but rather with Brazilian or Latin American ones and/or based on judgment of value at the discretion of the committee. Therefore, the Qualis system evaluation has become imprecise, which generated certain criticisms in the academic

Table I. Papers published, citations normalized and papers per thousand inhabitants of brics countries, between 2011 and 2016.

Country	Papers	Normalized Citations[1]	Papers per Thousand Inhabitants
Brazil	250680	0.78	1.19
Russia	194126	0.63	1.37
India	347293	0.78	0.26
China	1402689	1.00	1.01
South Africa	73663	1.11	1.32

[1]World Average = 1.
Source: (Cross et al. 2018, U.S. Census Bureau 2019).

environment, to a certain extent, because of the arbitrary classification of journals by the commission, which demonstrates a high degree of subjectivity of the system.

The Qualis system uses a rather questionable rule stating that the number of A1 publications must be lower than the number of A2 journals, and the sum of the two may not exceed 26% of the total, whereas the sum of A1 + A2 + B1 may not exceed 50% of the total (Fernandes & Manchini 2019). Thus, the classification of journals in higher strata depends on the existence of a certain number of journals in strata of lesser value rather than the number of citations obtained per published paper, which is the case of international databases.

Kellner (2017) states that the Qualis evaluation system became decisive for authors in their selection of a journal to which their manuscripts would be submitted. Therefore, some journals, which occupy lower positions in the ranks, receive very few submissions, which seriously impacts their chances of funding. Furthermore, there are few prospects for these journals, as Qualis system rules do not allow better ranking of all.

Gabardo et al. (2018) criticizes the Qualis system, stating that the model of the evaluative system causes researchers to be less concerned with publishing in journals of greater international or area relevance, but rather in the journals ranked among the upper strata of the Qualis system. According to the authors, this has led to discrepancy between the metrics of the international scientific evaluation and those of Brazil.

In contrast to the Qualis system, several metrics are used worldwide to evaluate scientific output without necessarily analyzing its content (Cainelli et al. 2015, Abramo et al. 2013). Considering the number of publications as the only method used to classify research institutions or researchers may have significant limitations (Duffy et al. 2011), as this method does not consider the case of the number of authors or the quality of the publications (Wu et al. 2015). Consequently, other objective metrics that take into account other factors, such as quality of publications, are essential to enable a more accurate assessment of the scientific output of individuals or institutions.

Therefore, new metrics for scientific output have been proposed around the world, primarily based on the number of citations of a paper. Citing the research of someone else provides the necessary information, ideas and key points to disseminate knowledge scientifically (Mingers & Leydesdorff 2015). The number of citations a paper receives reflects the impact it has had on subsequent research, and it may be said that a paper (or a collection of papers) with a higher number of citations tends to reflect higher quality. However, the number of citations also depends on the area surveyed and the period when the paper was produced (Cross et al. 2018).

Due to the limitations of citations, other aspects, which together are more representative of the productivity of a researcher, should be considered, including (i) number of publications, (ii) quality of the publications and (iii) scientific contribution of the papers (Duffy et al. 2011). Therefore, it is recommended that the procedures for evaluating productivity should be based primarily on citations of scientific papers.

Traditional evaluation methods, such as counting number of publications and number of citations, have presented some disadvantages, which explains the suggestion of a wide range of new methods for evaluating the productivity of researchers. Several quantitative metrics are used worldwide to investigate and compare academic productivity in the different academic areas, such as Journal Impact Factor (JIF) and H-Index, among others that are also primarily calculated based on the number of citations (Merigó et al. 2015, Garner et al. 2017).

JIF has become an important tool for assessing the quality of scientific journals, calculated by the Scientific Division of Thomson Reuters and published annually in the Journal Citation Reports (JCR). According to Bornmann & Marx (2016), although the metric was designed to evaluate journals as a whole, JIF's availability made it a common tool for evaluating papers and researchers, even though this indicator alone has its limitations, since it is not representative of individual statistics for each paper.

JIF is an indicator calculated for each journal, based on the ratio between the number of citations received in a year and the number of papers published in the previous two years. Some criticisms of the process can be considered, because, in the numerator, citations of all types of publications of the journal are considered, while in the denominator, only the papers are considered (Bornmann & Marx 2016). According to Sugimoto et al. (2013), there is an increasing relationship between JIF and number of citations, mainly in journals with greater selectivity to accept manuscripts.

The H-Index, which was not designed to evaluate research but to help researchers rank the literature more effectively, has become an important metric for evaluating the productivity of researchers (Mingers & Leydesdorff 2015). The H-Index is an easy to apply method that presents certain advantages and is very robust for situations in which there is extreme or scarce data. To calculate the H-Index, only the papers that have at least h citation are counted rather than the number of citations they have. In addition, the method values both the number of citations received and the number of papers cited (Harzing & Van der Wal 2009). Due to the characteristics of the H-Index calculation itself, some limitations can be observed, such as the disregard of papers with few citations and the counting of cited papers with the same weight. Clearly, because it depends on the number of citations, the H-Index is changeable over time.

Tüselmann et al. (2015) has shown that, when the journal is classified according to the H-Index, results similar to those of other metrics are obtained, which evidences this metric is in consonance with others used for classifying journals, such as citations and impact factor. This relation is explainable, as these indicators also depend on the citations. It is evident that, even when using citations as the basis for determining different metrics, they return similar but not necessarily equal results, so the use of multi-criteria may be a more accurate method to evaluate the scientific output of a group of researchers. However, the H-Index is mostly used to classify groups of researchers and no longer to classify journals.

Data sources are necessary for applying metrics. For such, certain databases can be consulted. Google Scholar, WoS and Scopus are among the most widespread databases in the world. The Scientific Electronic Library On-Line (SciELO) is one of the most important bases for Latin America and the Caribbean (Almeida & Grácio 2018). Despite the great relevance in these regions, according to Freitas et al. (2017), the use of SciELO may have limitations because it is a less comprehensive database. Thus, the international databases stand out due to their greater coverage. Mingers et al. (2012) recommend Google Scholar as a data source for the H-Index compared to WoS and Scopus. Despite the limited data quality, Google Scholar has the advantage of being one of the largest databases (Mingers et al. 2012). Each of these databases has its peculiarities.

Google Scholar stands out as a free access database, capable of creating opportunities for academia, as it enables citations and other metrics to be consulted in a very simplified way, with worldwide coverage (Adriaanse & Rensleigh 2013). Amara & Landry (2012) provided evidence that the data compiled with the Google Scholar present better coverage than those compiled with the WoS and Scopus for the business and administration area. However, it is observed that the availability of data in Google Scholar depends on the researcher who is being registered and the availability of the data for consultation. On the other hand, other international bases, such as WoS and Scopus, may imply considerable costs.

Some recent studies point to limitations of the Qualis system, that is a Brazilian system, in contrast, how internationally used metrics can be an alternative to improve the evaluation system of Brazilian graduate programs. In light of this, the objective of this paper was to evaluate the Qualis system against internationally accepted metrics for a group of researchers and to propose an improvement in the evaluation system, through multi-criteria, including Qualis.

MATERIALS AND METHODS

Certain traditional metrics were selected to evaluate scientific output and classify the researchers involved. For the purpose of exemplifying the effect of these metrics, a group of professors, who are also researchers, was selected from a Brazilian graduate course in Agricultural Engineering, evaluated with a score 6 out of 7 by CAPES, from the area of Agrarian Sciences I. The professors were selected by the criterion of being registered in the Google Scholar (GS) database. In this case, 14 were selected, out of 26 permanent professors who are the faculty of the graduate program studied as an example of the proposed methodology.

Through the GS platform, all papers published by these professors were consulted in scientific journals from 2015 to 2018. The H-Index of each professor was also obtained directly from the GS database, considering scientific output up to 2018. The number of citations of each of the published papers was also obtained.

As the GS database includes other publications, such as congress papers, dissertations, theses, books and other forms of scientific output, these were disregarded, since CAPES only counts papers in scientific journals. Duplicated papers and incoherent attribution of papers to a given researcher were excluded. For the cases of multiple authors, only the first was considered.

Besides GS, two other platforms were used to evaluate the score of each of the papers published by the group of professors, namely, Qualis and Journal of Citation Reports. For the former, the CAPES Sucupira platform was consulted with data referring to the most recent available four-year period, 2013-2016 (CAPES 2019). For the latter, the journal impact factor from 2017 was used through the Journal Citation Reports (or JIF) index, which is available through the ISI Web of Knowledge platform (UC San Diego 2019).

The Qualis, impact factor and number of citations were determined for each of the papers considered. In order to improve the evaluation of professors, a new metric was proposed to reduce the inconsistencies obtained using citations alone (Hutchins et al. 2011, Duffy et al. 2011). The metric known as secondary citations, i.e. citations of citations, was also used. This new metric was employed to reduce possible inconsistencies arising from the use of traditional citations only, as this is obtained by counting the citations of the papers, which cited the papers being evaluated. Secondary citations would be a kind of “second derivative of citations” when making an analogy with the Taylor series (Berry et al. 2015).

The Taylor series is represented by Eq. 1. This equation defines the relation between the discrete time values of the time function f(t) sampled at t = kT, $f_0^{(n)}$ denotes the value of the nth derivative of f at t=0 (Khan & Ohba 1999).

$$f_k = f_o + kTf_0^{(1)} + \frac{(kT)^2}{2!} f_0^{(2)} + \ldots + \frac{(kT)^n}{n!} f_0^{(n)} \qquad (1)$$

Taylor series approximations accurately describe the behavior of functions (Wang et al. 2016). Thus, the Secondary citations metric proposition is an analogy with the Second-Order Taylor-Series Expansion (Li & Deng 2016). In analogy with the Taylor Series, we are proposing that the valuation of scientific production does not depend on citations alone, which would be in analogy with the Taylor series, the first derivative of scientific production. We are proposing that citations of citations be included as an additional indicator that allows the in-depth evaluation of scientific production.

With Secondary citations, we are proposing a new metric to estimate the quality of citations, since the papers that gave rise to the most cited citations tend to be more relevant. Thus, the calculation procedure for estimating secondary citations consists of adding up all citations obtained by the papers that cited each of the researcher's articles investigated.

In order to exemplify the relevance of the new metric citations of citation, we are proposing an analysis of a simple case: Professor "x" had a paper with one citation and that citation had one citation. On the other hand, professor "y" had a paper as well, with only one citation, but that citation had 100 citations. It seems that professor's "y" paper tend to be more relevant due to the fact pointed by that new metric, the citations of citation

To evaluate the Qualis system, the score obtained by each professor was initially calculated with the following metrics:

i. Qualis score;
ii. JIF score;
iii. Citation score;
iv. Secondary citation score; and
v. H-Index score.

The Qualis score (s_Q) was calculated by summing the score of each paper, assigning scores to the seven strata of the Qualis system, that is 1.00 point for journals classified as A1, 0.85 for A2, 0.70 for B1, 0.55 for B2, 0.40 for B3, 0.25 for B4 and 0.10 for B5 (CAPES 2017).

The JIF score (s_{JIF}) was defined as the sum of the impact factors of the 2017 journals corresponding to each of the published papers (Bar-Ilan 2009). Citation score (s_c) was calculated as the sum of the number of citations of the published papers (Thelwall & Kousha 2017). Secondary citation score (s_{cs}) was obtained by summing the citations of papers' citation of each of the papers published by the professors, according to the GS database. The H-Index score (s_{HI}) assigned to the researchers was equal to their H-Index (Harzing & Van der Wal 2009).

Figure 1 illustrates the procedure adopted to obtain the score of each professor through the five metrics proposed. The three databases consulted are in red; in yellow, we find the data obtained after consulting these databases for each paper referring to a certain professor. The mathematical process of the sum for assigning the score is shown in green. Finally, the scores referring to: JIF score, Qualis score, Secondary citations score, Citations score and H-Index score are in blue.

Initially, the evaluation of the scientific output was carried out using the five metrics that can also be denominated as mono criteria, shown in blue, in Figure 1. Each of the five metrics was applied individually, and the scores were attributed for each professor, which normalized all the results. Therefore, the maximum score for each of the metrics was set to be equal to 100%.

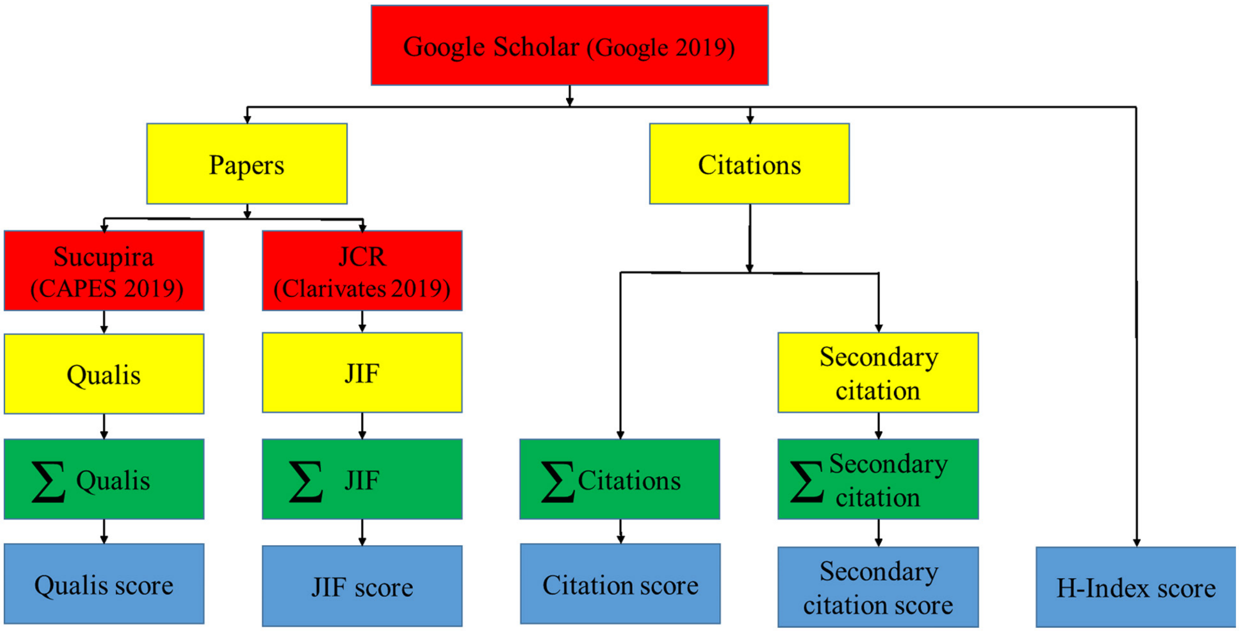

Figure 1. Flowchart of the data collection process for the five metrics proposed for evaluating Brazilian graduate programs.

Next, multi-criteria methods were applied based on the five metrics described. For such, three combinations of these criteria were created, with different weights (w_i). The score for each professor (w_{multi}) is defined by:

$$w_{multi} = w_Q \frac{s_Q}{s_Q^{max}} + w_{JCR} \frac{s_{JIF}}{s_{JIF}^{max}} + w_{ci} \frac{s_{ci}}{s_{ci}^{max}} + w_{fH} \frac{s_{HI}}{s_{HI}^{max}} + w_{cs} \frac{s_{cs}}{s_{cs}^{max}} \quad (2)$$

When s_Q^{max}, s_{JIF}^{max}, s_{ci}^{max}, s_{HI}^{max} and s_{cs}^{max} are the maximum scores obtained for the professor with the best evaluation in each criterion, for example, s_Q^{max} is the score for the professor with the highest Qualis Score. Three multi-criteria were defined based on Eq. 2. The three have the five metrics in common, and different weights were adopted for each of the metrics to create the three methods. The sum of the weights W_Q, W_{JCR}, W_{ci}, W_{HI} and W_{CS} was limited to the unit.

Finally, the correlation among the scores obtained for each of the mono criteria and multi-criteria used was calculated. Correlation was calculated among the matrices expressing the scores obtained for each professor. Correlation $c_{i,j}$ was calculated using the expression below:

$$c_{i,j} = \frac{\sum (x_i - \underline{x}_i)(y_j - \underline{y}_j)}{\sqrt{\sum (x_i - \underline{x}_i)^2 \sum (y_j - \underline{y}_j)^2}} \quad (3)$$

Where represents the criterion correlated with. Besides, $\underline{x}_i \wedge \underline{y}_j$ indicates the means of the matrices. As five mono criteria and three multi-criteria were evaluated, which totals 8 criteria,we have: $i,j \in \frac{N}{[1.8]}$, with $i \neq j$ i.e., we evaluated 56 correlations.

RESULTS AND DISCUSSION

To exemplify the effect of proposing new metrics for improving the Qualis system of journal classification, papers by 14 professors in a Brazilian graduate program were consulted in the Google Scholar (GS) database, for the period from 2015 to 2018. This basis was consulted because it allows easy individual access to scientific production through the profile of each professor in GS. In the results, only papers in scientific journals were considered. We found 286 papers.

We also verified the authorship of the papers and the existence of duplicates. Therefore, the inconsistencies of the GS platform, listed in the works of Adriaanse & Rensleigh (2013) and Haddaway et al. (2015), were overcome. It should be emphasized that a different database could have been chosen, which would provide similar results.

The input data history for the scoring metrics for each professor is represented as $p_{i,}$. Where p_1 indicates the professor with the highest Qualis score, and p_{14} indicates the professor with the lowest score. The cumulative results for the score obtained from 2015 to 2018 following the rules of Qualis, JIF, citations (Cit.) and secondary citations (Cit. Sec.) can be found in Table II. Table II also shows the H-Index up to the year of the sample studied.

Based on the data in Table II and the H-Index, the scores for each of the five metrics, calculated and standardized at 100% are shown in Figure 2. This methodology was proposed and applied to a study sample to evaluate any discrepancies among the results obtained according to each criterion, which facilitates the comparison among the results of the different metrics.

In Figure 2, the discrepancy between the scores obtained using the Qualis and JIF criteria is evident for the case in which professor p_1 scores more than p_2 in relation to Qualis, while in terms of JIF, professor p_2 obtains practically twice the p_1 score. This discrepancy leads us to conclude that the Qualis system overestimates journals with low or even no International impact factor. The fact that professor p_2 obtains higher scores for all other criteria further reinforces our hypothesis. The existence of zero scores is also evident in three criteria for three professors in the sample.

Table III was prepared to better explain the distribution of the output scores of the 14 professors in the period analyzed by four of the mono criteria in the seven Qualis strata.

In Table III, the higher Qualis total scores for papers classified as B1 is evident, i.e. about 62.67%, but when we analyze the same stratum in terms of JIF score, we corroborate the overvaluation of

Table II. Input data for professor metrics from graduate program rated 6 out of 7 from 2015 to 2018, cumulative score history and H-Index until 2018.

	Qualis Score	JIF Score	Cit. Score	Sec. Cit. Score	H-Index
p_1	31.75	39.74	72	88	14
p_2	29.35	73.56	198	613	26
p_3	23.05	6.46	32	11	11
p_4	22.30	15.10	69	139	13
p_5	20.60	14.79	80	96	19
p_6	19.35	46.97	41	113	14
p_7	10.20	15.38	52	183	12
p_8	7.55	10.05	22	67	13
p_9	5.45	1.29	6	0	2
p_{10}	4.70	3.38	0	0	20
p_{11}	3.00	0.77	5	2	3
p_{12}	3.10	0.62	4	1	2
p_{13}	2.50	5.11	7	23	10
p_{14}	1.40	0.00	2	0	1

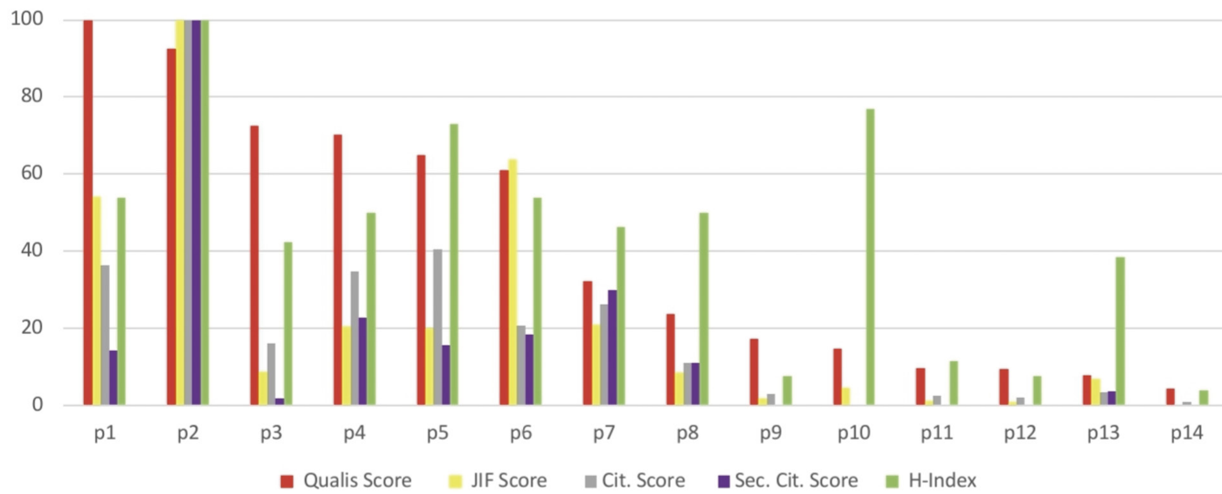

Figure 2. **Scores obtained using the Qualis, JIF, Citation, Secondary Citations and H-Index criteria for evaluating Brazilian postgraduate program.**

Table III. **Total Qualis score, JIF, Citations and Secondary Citations for the sample for the different Qualis strata, referring to the 2015 to 2018 period.**

	Qualis Stratum	**A1**	**A2**	**B1**	**B2**	**B3**	**B4**	**B5**
Qualis	Unit values	1.00	0.85	0.70	0.55	0.40	0.25	0.10
	Number of papers	29	18	165	15	32	5	22
	Total score (Score %)	29.00 (15.74)	15.30 (8.30)	115.50 (62.67)	8.25 (4.48)	12.80 (6.95)	1.25 (0.68)	2.20 (1.19)
JIF	Total score (Score %)	124.98 (54.46)	26.58 (11.58)	69.12 (30.12)	7.63 (3.32)	0.39 (0.17)	0.23 (0.10)	0.56 (0.25)
	% of "0" JIF score	0.00	5.56	29.09	53.33	96.88	80.00	90.91
	Mean JIF/paper	4.31	1.48	0.42	0.51	0.01	0.05	0.03
Citations	Total score (Score %)	86.00 (15.01)	67.00 (11.69)	339.00 (59.16)	24.00 (4.19)	20.00 (3.49)	9.00 (1.57)	28.00 (4.89)
	% of "0" Citations	41.38	44.44	55.15	46.67	68.75	20.00	54.55
	Citations /paper	2.97	3.72	2.05	1.60	0.63	1.80	1.27
Secondary Citations	Total score (Score %)	151.00 (11.38)	209.00 (15.75)	856.00 (64.51)	37.00 (2.79)	4.00 (0.30)	1.00 (0.08)	69.00 (5.20)
	% of "0" Sec. Cit.	51.72	61.11	71.52	46.67	93.75	80.00	90.91
	Sec. Cit. /paper	5.21	3.76	5.19	2.46	0.13	0.20	6.90

papers classified in lower Qualis strata, which is 30.12%. Other discrepancies are more evident in the lower strata of the Qualis system. The percentage of papers with zero JIF score rises drastically, which indicates low relevance of the journals classified in these strata. The stratum that receives the highest score for citations and secondary citations is that of papers classified as B1, but it is emphasized that this phenomenon occurs in case of citations by the amount of papers, since the number of citations per paper is decreasing in the first five strata. For secondary citations, the highest score per paper was obtained in stratum B5, which is a contradiction. Again, the need for multi-criteria is emphasized in order to prevent incorrect assessment.

Further in-depth analysis of the data shown in Table III reveals that the use of individual mono criteria may limit the classification of programs and/or researchers or lead to serious evaluation errors. Thus, this study also assessed the use of multi-criteria presented along with the five previous mono criteria in Table IV.

Table IV summarizes the five mono criteria and the three multi-criteria. The three multi-criteria weights 80%, 60% and 20% for Qualis, and the remaining weights are equally distributed for the other four mono criteria considered. These weights were used throughout the simulations proposed in the study. The different weights proposed were thus determined as an alternative to provide a transition from Qualis alone as mono criteria to a multi-criteria system.

The first two multi-criteria shown in Table V were proposed to enable a transition from the current evaluation system of graduate programs based only on the Qualis Score toward a more comprehensive evaluation, according to the five mono criteria.

This transition could take place over a period, say eight years, with four years for each of the initial multi-criteria to facilitate accommodation of the scientific output strategies of the professors involved. With this transition, characterized by the reduction of weight and consequent increase of the others, the evaluation criterion of graduate programs is expected to become fairer and more efficient over time, since it is based on a multi-criteria system, without abandoning, however, the system in effect, i.e. Qualis.

After this period, the third multi-criteria would enter into effect, which proposes a 20% weighted evaluation for each mono criterion. It is also believed that Qualis has its value because many Latin American journals, primarily published in Portuguese and Spanish, obtained regional relevance, despite their coverage area.

Of course, there are other mono criteria that could be counted in the metrics for classifying a group of professors, including: (i) the G-index, calculated based on the distribution of citations received by an author (Egghe 2006) and (ii) the i10-index, in which i refers to the number of papers with at least 10 citations (López-Cózar et al. 2014). It is believed that the use of the four metrics to evaluate graduate programs according to citations, plus Qualis, already considers the G-index and i10-index score. Thus, we consider that the dimensions of the evaluation considered are simple, comprehensive and direct enough to indicate quality classification with lower probability of error.

Table IV. **Weights for classifying productivity for mono criteria and multicriteria.**

	Mono criterion					Multicriteria		
Name	**Qualis**	**JIF**	**Cit.**	**Sec. Cit.**	**H-index**	**Qualis 80%**	**Qualis 60%**	**All 20%**
W_q	**1.00**	0.00	0.00	0.00	0.00	**0.80**	**0.60**	**0.20**
W_{JIF}	0.00	**1.00**	0.00	0.00	0.00	**0.05**	**0.10**	**0.20**
W_{ci}	0.00	0.00	**1.00**	0.00	0.00	**0.05**	**0.10**	**0.20**
W_{sc}	0.00	0.00	0.00	**1.00**	0.00	**0.05**	**0.10**	**0.20**
W_{fH}	0.00	0.00	0.00	0.00	**1.00**	**0.05**	**0.10**	**0.20**
Total	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table V. Productivity ratings for the study sample, according to the 8 established criteria.

Qualis	JIF	Citation	Sec. Cit.	H-index	Qualis 80%	Qualis 60%	All 20%
p_1	p_2	p_2	p_2	p_2	p_2	p_2	p_2
p_2	p_6	p_5	p_7	p_{10}	p_1	p_1	p_1
p_3	p_1	p_1	p_4	p_5	p_4	p_4	p_6
p_4	p_7	p_4	p_6	p_6	p_3	p_5	p_5
p_5	p_4	p_7	p_5	p_1	p_5	p_6	p_4
p_6	p_5	p_6	p_1	p_4	p_6	p_3	p_7
p_7	p_3	p_3	p_8	p_8	p_7	p_7	p_3
p_8	p_8	p_8	p_{13}	p_7	p_8	p_8	p_8
p_9	p_{13}	p_{13}	p_3	p_3	p_{10}	p_{10}	p_{10}
p_{10}	p_{10}	p_9	p_{11}	p_{13}	p_9	p_9	p_{13}
p_{11}	p_9	p_{11}	p_{12}	p_{11}	p_{13}	p_{13}	p_9
p_{12}	p_{11}	p_{12}	p_9	p_{12}	p_{11}	p_{11}	p_{11}
p_{13}	p_{12}	p_{14}	p_{14}	p_9	p_{12}	p_{12}	p_{12}
p_{14}	p_{14}	p_{10}	p_{10}	p_{14}	p_{14}	p_{14}	p_{14}

Table V presents the results of the classification of the group of professors studied for each of the eight criteria. In this case,it is clear that those professors who had research published in journals of greater international relevance were more valued, that is, they occupied higher positions, as their scores were calculated using the proposed multi-criteria. The results found are in line with the academic environment and recent literature, as the Qualis system has been strongly criticized for being a subjective method of journal evaluation. In other words, subjectivity results in a significant discrepancy with international metrics (Kellner 2017, Gabardo et al. 2018).

Highlighted in blue in Table V, we have professor p_2, the best evaluated in all proposed mono criteria methods, except Qualis, where he occupies the second position. Thus, by applying the multi-criteria methods, the academic performance of this professor can be better evaluated. The same process occurs for p_7 and p_{13}, highlighted in yellow and red, since in these cases, their evaluations were improved by multi-criteria methods. In the case of professors such as p_3, the application of multi-criteria undervalues their assessment because they have fewer scores in 4 out of the five proposed mono criteria.

Table VI presents the results of the correlations between the evaluation criteria considered, calculated by Eq. 3. The variation of the correlations between the metrics ranged from 0.584 to 0.994. In Table VI, it can also be seen that the means of the correlations of the Qualis metric with the others occupies an intermediate value among the other metrics. Therefore, it can be said that there is a correlation between Qualis and the proposed metrics. As citations are a metric that serves as the basis for other metric indicators, it is evident that the mean correlation of this metric is the highest among the mono-criteria cases. However, similarly to other metrics, citations have some limitations, and can be even subdivided into positive, neutral or negative (Tahamtan & Bornmann 2018). It can

Table VI. **Linear correlation between the ranking lists obtained by the 8 classification criteria.**

Criterion	Qualis	JIF	Cit.	Sec. Cit.	H-index	Qualis 80%	Qualis 60%	All 20%
Qualis	-	0.763	0.782	0.584	0.646	0.994	0.973	0.858
JIF	0.763	-	0.866	0.851	0.680	0.816	0.866	0.929
Cit.	0.782	0.866	-	0.944	0.733	0.838	0.892	0.962
Cit. Sec.	0.584	0.851	0.944	-	0.677	0.662	0.745	0.894
H-index	0.646	0.680	0.733	0.677	-	0.700	0.754	0.833
Qualis 80%	0.994	0.816	0.838	0.662	0.700	-	0.992	0.908
Qualis 60%	0.973	0.866	0.892	0.745	0.754	0.992	-	0.953
All 20%	0.858	0.929	0.962	0.894	0.833	0.908	0.953	-
Mean	**0.800**	**0.824**	**0.860**	**0.765**	**0.718**	**0.844**	**0.882**	**0.905**

be concluded that the correlation between the three multi-criteria methods decreases as the participation of the Qualis system is reduced, and in the case with the lowest correlation, all 20%, it is observed that the correlation is higher, if compared with any mono criterion methods, i.e. 0.905.

In Table VII, JIF was considered as the basis for calculating the Qualis for the sample studied. Thus, it is evident that the Qualis system overestimated the papers with the worst classification in its stratum. For example, to fit the Qualis system score in relation to JIF, we would have to reduce the scores of strata B3 and B4 by approximately 133 and 25 times, respectively. We have the discrepancies between the scores by the Qualis and JIF systems, which indicates that, according to the JIF, the B2 stratum should be better valued than B1, and that strata B4 and B5 should be better valued than B3 for the sample studied.

CONCLUSIONS

In the results presented, the subjectivity of the classification of the Qualis system was demonstrated for the sample of supervisors and professors of a graduate program in Agricultural Sciences I, evaluated with a score of 6 out of 7. This can be proved by the overvaluation of journals with impact factor (JIF) equal to zero. It was also demonstrated that the publications classified in the two lower strata of Qualis, i.e. B4 and B5 are more relevant than B3, which contradicts what the Qualis system establishes, in view of JIF.

All mono criteria used have their qualities and limitations. Using the number of citations to evaluate the quality of a paper is a good indicator, but it has two basic limitations, namely, the age of the paper, since the number of citations depends on the date of the publication, and the fact that papers may be cited not for their relevance, but because they may have a disputed concept or vision of a given subject. Using JIF as a metric for evaluating researchers is very important because it is an objective global indicator. In other words, unlike Qualis, JIF can be obtained through a mathematical equation. There are some flaws in using JIF as an indicator to evaluate papers because the system considers journals rather than papers individually. The H-index is a good indicator of individual productivity, but it can also be questioned because it depends on the length of the researcher’s career.

A new metric was proposed, secondary citations, i.e. citation of the citations of the papers investigated, as a method of mitigating the possible disadvantages of the previous metrics. We can emphasize an advantage of the

Table VII. Ratio between the Qualis score and JIF score for the sample.

Classification	A1	A2	B1	B2	B3	B4	B5
Current Score, Qualis	1.000	0.850	0.700	0.550	0.400	0.250	0.100
Proposed Score, JIF	1.000	0.340	0.090	**0.120**	0.003	**0.010**	**0.006**
Ratio	1.000	2.500	7.778	4.583	133.333	25.000	16.667

metric, which is that it allows the assessment of the impact of the papers that originally cited the evaluated paper, that is, it allows the assessment of the real impact or contribution to the scientific community. Thus, when analyzing citations of citations, the intention is to deepen the results obtained by evaluating the chain of citations received by a paper, that is to evaluate not only the citations received, but the citation of citations received. It must be pointed out that a limitation of the secondary citations, as well as the citations, is the time of publication of the paper, since recent studies generally have fewer citations. Still about the importance of using the criteria citations and secondary citations is the analogy of the Taylor series to explain one given mathematical expression, where they represent the first and second derivative.

It was proposed that the assessment of Brazilian graduate programs, traditionally carried out using the Qualis system, should be performed through the multi-criteria system proposed in this study. The use of multi-criteria methods allows fairer evaluation by considering several metrics together and minimizing any failures or questionable points of the individual application of a mono criterion alone. Therefore, the Qualis system evaluates scientific output only through the classification of the journal, while it is also proposed the evaluation of the indicators of the paper itself and the researcher, through citations and the H-Index.

Finally, there is a growing need for the improvement of the Qualis system, the tool used for the assessment of the productivity of graduate programs, which produce most of the Brazilian scientific output. Thus, it is proposed that this system give rise to a multi-criteria evaluation process that includes Qualis, JIF, citation, secondary citation and H-Index. It is also suggested that this transition be made over several years to objectify metrics that are fairer to the Brazilian scientific community.

Databases other than those consulted in this study could be used with similar results. The Google Scholar database was consulted as the primary source of data for evaluating the scientific output of a group of professors, due to its operational simplicity, comprehensiveness and for being an open platform. Thus, the results could even undergo significant changes if other databases were adopted. The present paper presented a methodology aimed at improving the evaluation system of Brazilian graduate programs. The results obtained for the sample used exemplify the deficiencies of the current system, which is exclusively based on Qualis.

The results of this study have some limitations, as a group of professors was used as an example of application of the proposed methodology. The replication of the methodology for other groups of professors of other graduate programs could support the improvements to the proposed method. As a final limitation, inherent to any metric indicators, which do not individually evaluate the content of each paper, inconsistencies are likely to occur, since scientific output is not considered in its content but in indirect measures, here known as metrics.

Acknowledgments

The authors are thankful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for their financial support to the development of this research.

REFERENCES

ABRAMO G, D'ANGELO CA & ROSATI F. 2013. Measuring institutional research productivity for the life sciences: the importance of accounting for the order of authors in the byline. Scientometrics 97: 779-795.

ADRIAANSE LS & RENSLEIGH C. 2013. Web of Science, Scopus and Google Scholar. Electron Libr 31: 727-744.

ALMEIDA CC & GRÁCIO MCC. 2018. Produção científica brasileira sobre o indicador "Fator de Impacto": um estudo nas bases SciELO, Scopus e Web of Science. Revista Eletrônica de Biblioteconomia e Ciência da Informação 24: 62-77.

AMARA N & LANDRY R. 2012. Counting citations in the field of business and management: why use Google Scholar rather than the Web of Science. Scientometrics 93: 553-581.

BAR-ILAN J. 2009. Rankings of information and library science journals by JIF and by h-type indices. J Informetr 4: 141-147.

BERRY DW, CHILDS AM, RICHARD C, KOTHARI R & SOMMA RD. 2015. Simulating Hamiltonian Dynamics with a Truncated Taylor Series. Phys Rev Lett 114: 327-345.

BORNMANN L & MARX W. 2016. The Journal Impact Factor and alternative metrics: A variety of bibliometric measures has been developed to supplant the Impact Factor to better assess the impact of individual research papers. J Sci & Soci 17: 1094-1097.

CAINELLI C, MAGGIONI AM, UBERTI TE & FELICE A. 2015. The strength of strong ties: How co-authorship affect productivity of academic economists? Scientometrics 102: 673-699.

CAPES - COORDENAÇÃO DE APERFEIÇOAMENTO PESSOAL DE NÍVEL SUPERIOR. 2013. Qualis Periódicos. Resource Document. http://www.capes.gov.br/acessoainformacao/perguntas-frequentes-/avaliacao-da-pos-graduacao/7422. Accessed 25 July 2019.

CAPES - COORDENAÇÃO DE APERFEIÇOAMENTO PESSOAL DE NÍVEL SUPERIOR. 2017. Qualis Periódicos. Resource Document. http://www.capes.gov.br/images/stories/download/-avaliacao/relatoriosfinais-quadrienal-2017/20122017-CIENCIAS-AGRARIAS-I-quadr-ienal-.pdf. Accessed 25 July 2019.

CAPES - COORDENAÇÃO DE APERFEIÇOAMENTO PESSOAL DE NÍVEL SUPERIOR. 2019. Sucupira platform. Online database. https://sucupira.capes.gov.br/sucupira/. Accessed 25 July 2019.

CROSS D, THOMSON S & SINCLAIR A. 2018. Research in Brazil A report for CAPES by clarivate analytics. Resource Document. Capes Clarivate report. https://www.capes.gov.br/images/stories/download/diversos/17012018-CAPES-InCitesReport-Final.pdf. Accessed 25 July 2019.

DUFFY RD, JADIDIAN A, WEBSTER GD & SANDELL KJ. 2011. The research productivity of academic psychologists: assessment, trends, and best practice recommendations. Scientometrics 89: 207-227.

EGGHE L. 2006. Theory and practise of the g-index. Scientometrics 69: 131-152.

FERNANDES GAAL & MANCHINI LO. 2019. How QUALIS CAPES influences Brazilian academic production? A stimulus or a barrier for advancement? Braz J Political Econ 39: 285-305.

FREITAS JL, ROSAS FS & MIGUEL S. 2017. Estudos métricos da informação em periódicos do Portal SciELO: visibilidade e impacto na Scopus e Web of Science. (Metric studies of information in journals indexed at SciELO: visibility and impact at Scopus and Web of Science). In Portuguese. Palabra Clave 6(2): e021.

GABARDO E, HACHEM DW & HAMADA G. 2018. Sistema Qualis: análise crítica da política de avaliação de periódicos científicos no Brasil. RDUNISC 0(54): 144-185.

GARNER RM, HIRSCH J, ALBUQUERQUE FC & FARGEN KM. Bibliometric indices: defining academic productivity and citation rates of researchers, departments and journals. J Neurointerv Surg 10: 102-106.

HADDAWAY NR, COLLINS AM, COUGHLIN D & KIRK S. 2015. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLoS ONE 10: 237-247.

HARZING AW & VAN DER WAL R. 2009. A Google Scholar h-Index for Journals: An Alternative Metric to Measure Journal Impact in Economics and Business. J Am Soc Inf Sci Technol 60: 41-46.

HUTCHINS BI, YUAN X, ANDERSON JM & SANTANGELO GM. 2016. Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level. PLoS Biol 14: 541-554.

KELLNER AWA. 2017. The Qualis system: a perspective from a multidisciplinary journal. An Acad Bras Cienc 89: 1339-1342.

KHAN IR & OHBA R. 1999. Closed-form expressions for the finite difference approximations of first and higher derivatives based on Taylor series. J Comput Appl Math 107: 179-193.

LETA J, THIJS BGW & GLÄNZEL W. 2013. Um estudo de nível macro da ciência no Brasil: sete anos depois. Encontros Bibli 18(36): 51-66.

LI X & DENG Z. 2016. Identification of Dynamic Loads Based on Second-Order Taylor-Series Expansion Method. Shock Vib 2016: 1-9.

LÓPEZ-CÓZAR ED, ROBINSON-GARCÍA N & TORRES-SALINAS D. 2014. The Google scholar experiment: How to index false papers and manipulate bibliometric indicators. J Assoc Inf Sci Technol 65: 446-454.

MERIGÓ JM, MAS-TUR A, ROIIG-TIERNO N & RIBEIRO-SORIANO D. 2015. A bibliometric overview of the Journal of Business Research between 1973 and 2014. J Bus Res 68: 2645-2653.

MINGERS J & LEYDESDORFF L. 2015. A review of theory and practice in scientometrics. Eur J Oper Res 246: 1-19.

MINGERS J, MACRI F & PETROVICI D. 2012. Using the h-index to measure the quality of journals in the field of Business and Management. Inf Process Manag 48: 234-241.

OLIVEIRA AB, RODRIGUES RS, BLATTMANN U & PINTO AL. 2015. Comparação entre o Qualis/Capes e os índices H e G: O caso do portal de periódicos UFSC. Rev Inf Inf 20: 70-85.

SUGIMOTO CR, LARIVIÈRE VNC & CRONIN B. 2013. Journal acceptance rates: A cross-disciplinary analysis of variability and relationships with journal measures. J Informetr 7: 897-906.

TAHAMTAN I & BORNMANN L. 2018. What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. Scientometrics 121: 1635-1684.

THELWALL M & KOUSHA K. 2017. ResearchGate versus Google Scholar: Which finds more early citations? Scientometrics 112: 1125-1131.

TÜSELMANN H, SINKOVICS RR & PISHCHULOV G. 2015. Towards a consolidation of worldwide journal rankings – A classification using random forests and aggregate rating via data envelopment analysis. Omega 51: 11-23.

U.S. CENSUS BUREAU. 2019. U.S. and World Population Clock. Online database. https://www.census.gov/popclock/world. Accessed 25 July 2019.

UC SAN DIEGO. 2019. Measuring your Research Impact: Journal Impact. Resource Document. https://ucsd.libguides.com/ResearchImpact/JIF. Accessed 25 July 2019.

WANG Z, YE N, MALEKIAN R, XIAO F & WANG R. 2016. TrackT: accurate tracking of RFID tags with mm-level accuracy using first-order taylor series approximation. Ad Hoc Networks 53: 132-144.

WU D, LI M, ZHU1 X, SONG H & LI J. 2015. Ranking the research productivity of business and management institutions in Asia–Pacific region: empirical research in leading ABS journals. Scientometrics 105: 1253-1272.

How to cite
CASTRO PHGRP & OLIVEIRA FILHO D. 2022. Metric indicators for the evaluation of graduate programs in Brazil: from Qualis to multi-criteria. An Acad Bras Cienc 94: e20210962. DOI 10.1590/0001-3765202220210962.

Manuscript received on November 3, 2020;
accepted for publication on January 21, 2022

PEDRO HENRIQUE G.R.P. CASTRO
https://orcid.org/0000-0002-1849-0365

DELLY OLIVEIRA FILHO
https://orcid.org/0000-0003-4133-0199

Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, Av. Peter Henry Rolfs, s/n, 36570-900 Viçosa, MG, Brazil

Correspondence to: **Delly Oliveira Filho**
E-mail: delly@ufv.br

Author contributions

Pedro Henrique Gonçalves Rigueira Pinheiro Castro collected data and references and wrote and revised the manuscript. Delly Oliveira Filho guided the work, supervised the data and helped in the definition of the methodology, discussion of the results and final review.