

### Necessary changes in the evaluation of graduate programs in Brazil

Rita Barradas Barata<sup>(a)</sup>

<[rita.barradasbarata@gmail.com](mailto:rita.barradasbarata@gmail.com)> 

<sup>(a)</sup> Faculdade de Ciências Médicas, Santa Casa de São Paulo. Rua Dr. Cesário Motta Jr., 61. São Paulo, SP, Brasil. 01221-020. Ex-Diretora da Avaliação da CAPES.

The contribution made by the graduate program evaluation system implemented by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES, acronym in Portuguese) to the development of graduate training and education and scientific production in the country is undeniable. However, as with any evaluation process, after almost 20 years since its introduction, during which time small incremental and operational changes have been made, the current model needs to be reviewed.

The need to review the assumptions and evaluation instruments results from the observation of a number of unintended consequences by area coordinators, evaluation consultants, and CAPES directors, including: the artificial multiplication of programs in the same institution demonstrating a high level of fragmentation between academic groups, which weakens and impairs training and scientific production, leading towards extreme specialization; difficulty in tailoring criteria to the wide range of courses on offer; the predominant focus on scientific production to the detriment of aspects of training



and education; the increasingly acritical use quantitative indicators; the tendency to evaluate program performance based on the individual performance of faculty members rather than considering the program itself as the as unit of analysis; and the increasing rigidity of programs driven primarily by meeting criteria rather than the continuing pursuit of quality.

The first aspect that deserves reflection on the part of the academic community is the conception of evaluation underlying the adopted criteria. Currently, the evaluation compares the performance of different programs, creating the program rankings. Although there is no pre-established cut-off score, as many seem to believe, basing evaluation on comparisons makes the system all the more unpredictable, insofar as a program's score depends not only on the program itself but on the performance of all programs as a whole. This means that the score of a program that has improved its performance does not change if the other programs have improved more.

As an alternative to the ranking system, the evaluation could be based on whether the program meets the typical requirements of each scored item to make the evaluation more predictable and dependent solely on the performance of the program itself. Although this alternative may seem more suitable in terms of program management, it could discourage renovation and efforts to continuously improve performance, resulting in the stagnation of the system as a whole. This way, evaluation would tend to become more conservative, running the risk of creating unattainable "types".

Another important consideration for reformulating the evaluation process is the need to strike a balance between quantitative and qualitative approaches. Evaluation entails assigning a value to that which is being analyzed. One cannot escape this task by assigning quantitative indicators a position of neutrality. Given the sheer size of the National Graduate System (SNPG, acronym in Portuguese), it is inevitable that researchers will resort to quantitative indicators; however qualitative evaluations are also necessary to differentiate between situations that generate similar quantitative indicators from conditions that are very different and, therefore, have distinct meanings.

This consideration is intertwined with the need to distinguish between the different types and modalities of courses, requiring evaluations to be tailored to their particular features. Currently, for example, very large programs with over 100 faculty members are evaluated using the same parameters applied to very small programs with only 10 to 12 professors. This situation creates distortions that are not capable of correctly reflecting performance of individual programs.

Among the documents elaborated to aid the National Graduate Plan (PNPG, acronym in Portuguese)<sup>1</sup> National Monitoring Committee in developing a proposal for a new evaluation system, the document forwarded by the Brazilian Academy of Science (ABC, acronym in Portuguese) brings various interesting ideas, one of which is closely related to the current heterogeneity in the SNPG. Academics have proposed that programs should be able to define their vocations themselves, thus enabling the creation of "clusters" of similar programs to which more tailored evaluation criteria would be applied. The following groups were suggested: basic research, basic research on strategic themes, applied research in the social area, and applied research in the area of technology.

The final document of the National Monitoring Committee, recently approved by CAPES' Superior Council (10/10/2018)<sup>2</sup> proposes that the organization adopts a multidimensional evaluation model that considers different aspects without the need to condense them into one single score. The multidimensional approach is able to accommodate the diverse range of vocations and program objectives. Suggested dimensions include: human resources education and training, internationalization, scientific production, innovation and knowledge transfer, and impact and economic and social relevance.

There are at least three crucial aspects that are currently absent or undervalued in the evaluation process that should be part the proposal for the new evaluation process, each of which have been addressed by the National Monitoring Committee: the issue of education and training, self-evaluation, and social and economic impact.

The current evaluation process is firmly shaped around the intellectual production of programs, which corresponds to practically 70% of the score (30% to 40% of the score allocated to the intellectual production of academic staff and 30% to 40% to student production). Issues related to education and training are underevaluated because the program proposal does not directly contribute to the score, despite the fact that they can hinder the obtainment of certain scores.

Education and training is indirectly evaluated via the publication of the final work results in the form of articles or books and the immediate production (three to five years) of program graduates. As a result of this "undervaluation", programs have decreased subject load and other education and training activities overtime, meaning that qualifications are often awarded without students attaining the desired level of autonomy and intellectual leadership. The idea that the training and education of researchers can be based almost exclusively on the activities of research groups needs to be reviewed. Current programs have practically no room for broad theoretical and methodological education and training, meaning that students are limited to the more technical aspects of research. Academic staff development has also been largely relegated to the background, despite the fact that studies show that the majority of graduates will work in teaching in higher education.

One point that has long been debated internationally by various academies of science is the need to aggregate components from the arts and humanities to science courses, reinstating the question of intellectual formation beyond the simple technical preparation of scientists. Linked to this need to promote broader education and training is the diversity of backgrounds of faculty members, which is currently practically precluded from the evaluation criteria.

The new evaluation process will need to consider the backgrounds of faculty members, researchers, and professionals in the widest possible sense, rather than being restricted solely to scientific or technological production. In this respect, the evaluation of the background of graduates has assumed central importance for effective program evaluation, although how this should be done has yet to be defined.

Systematic program self-evaluation should also assume a more prominent role in the final evaluation process. Both academics and students should be required to conduct a self-evaluation, whose results should be shared with the consultants to the evaluation committee and allow the program and its trajectory to be contextualized. It will fall on the program to outline its virtues and eventual weaknesses. Who



were the best students during the period? Which theses and dissertations warrant specific mention? Which was the most relevant intellectual work produced by the program? What critical contributions has the program made to the region, country, or internationally?

Indicating the best work produced by the program helps committees conduct a qualitative evaluation of the products and graduates rather than simply considering overall production, whose volume has made it impossible for committees to assess the real contribution of the SNPG to the development of the country and the relative position of programs beyond quantitative indicators.

Likewise, the impact and social and economic relevance of programs are currently underevaluated when it comes to social inclusion. The current evaluation instruments do not allow committees to effectively assess the impact that the knowledge produced and graduates have had at local, regional, national, and international level.

These improvements require changes in the operational aspects of the evaluation process and in the evaluation instruments.

Depending on the consensus built with the academic community around conceptual aspects, it will be necessary to redesign instruments and redefine procedures.

One aspect that requires change is the scoring scale. Although scores currently range between one and seven, in practice the majority of programs come down to just three points (3, 4 or 5), since 1 and 2 mean the disaccreditation of the program and 6 and 7 denote excellence. An alternative would be to adopt a five-point scale excluding disaccreditation and excellence. The wider range of this scale would make it easier for evaluators to assess the advances made by the program.

Another point, upon which consensus has been reached in practically all of the evaluation process discussion forums, is the inadequacy of the Evaluation Form, whose items and weighting are very tied, hindering a more qualitative evaluation.

Changes need to be made to data collection, the main instrument used by committees to record information about the program. Various modules need to be replaced or reformulated to make them more useful. Urgent reformulation is required of the modules that refer to the proposal, which should be tailored to the different modalities (academic or professional), and types of courses offered (isolated program, associated, or network), while the books and technical production modules should contain basic information to enable the global assessment of production and the evaluation of professional programs, respectively.

Graduate follow-up should be maintained and improved to allow for a more effective evaluation of the education and training process. Currently, surveys are conducted by the Center for Strategic Studies and Management (CGEE, acronym in Portuguese) using linkage between the graduate database and the Annual Social Information Report (RAIS, acronym in Portuguese) produced by the Ministry of Labor. However, coverage is relatively low for some areas of evaluation because the RAIS only includes information on individuals who are formally employed in this country. Furthermore, the information only allows for the evaluation of employability, income generation, and spatial mobility between areas of knowledge across Masters' and Doctoral programs.



Finally, it is necessary to redefine the Qualis Journals, the system used by CAPES to classify the scientific production of graduate programs, to make this tool more homogenous across areas and capable of signaling the general characteristics of program production. Currently, the criteria adopted for each area are incommensurable, which hampers integration between the areas and results in lack of communication between editors, authors, research support agencies, and institutional policies in higher education institutions.

If the journals in which production is being published are being used by the committees as a substitute for the peer review evaluation process, there is no reason why the same journal receives classifications that are so disparate across areas.

One way of ranking scientific production, taking into account the particularities of each area of knowledge and between the areas of evaluation, would be to build a *Qualis* in which each journal is allocated a single area of evaluation according to the scope of the publication, with each area of evaluation classifying only the journals that come within its scope. The second step would be to establish common bibliometric indicators such as the h-index, the impact factor, or Source Normalized Impact per Paper (SNIP). The ideal solution would be combine metrics that assess distinct aspects of impact. The cut-off points (percentiles) for each indicator should be defined based on an exhaustive list of publications in each area rather than on program production in order to correctly position national production vis-à-vis world production. For publications in the fields of humanities and social sciences, the Google Scholar could provide an adequate substitute for other indexing systems in which production is underrepresented.

In short, any new evaluation process must: be more focused on important aspects of education and training as opposed to just knowledge production; combine quantitative and qualitative indicators; enable better contextualization of programs; include self-evaluation; focus on better quality production rather than global production, whether of graduates or knowledge; value the social and economic relevance of programs; and avoid individual analysis by faculty members, prioritizing the program as a whole.

This is the challenge that the academic community must face in the coming years in order to reformulate the evaluation system and continue contributing to the development of the country.

## Copyright

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, BY type (<https://creativecommons.org/licenses/by/4.0/deed.en>).





## References

1. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Plano Nacional de Pós-Graduação - PNPg 2011-2020 [Internet]. 2018 [citado 10 Nov 2018]. Disponível em: <http://www.capes.gov.br/plano-nacional-de-pos-graduacao>
2. Brasil. Ministério da Educação. Proposta de Aprimoramento do Modelo de Avaliação da PG. Documento Final da Comissão Nacional de Acompanhamento do PNPg 2011-2020 [Internet]. Brasília, DF: MEC; 2018 [citado 10 Nov 2018]. Disponível em: [https://www.capes.gov.br/images/stories/download/conselho-superior/18102018\\_PNPg\\_CS\\_Avaliacao\\_Final\\_CS\\_FINAL\\_17\\_55.pdf](https://www.capes.gov.br/images/stories/download/conselho-superior/18102018_PNPg_CS_Avaliacao_Final_CS_FINAL_17_55.pdf)

Submitted on 11/12/2018.

Approved on 11/12/2018.