

Draft

EDITORIAL

Published as:

Mulder, M. (2015). Editorial: Research Assessment. *Journal of Agricultural Education and Extension*, (21),3, 187-200.

Kindly refer to the journal version for exact quotes because of detailed corrections and precise page numbers

Dear reader,

If you are in the business of doing research and publishing, you will be subject to a regular assessment. Be it as a PhD candidate, researcher or professor at the university, head of department or dean. As PhD candidate the object of assessment is your progress in the programme and for the final dissertation. As researcher or professor it is your publications, and as head of department or dean, it is the collective research performance of your department or school.

In this editorial I would like to focus on the research assessment of a department, because I am in the middle of this process. My department is part of the Social Sciences Group of Wageningen University, and because Wageningen University is a world-leading university in agricultural and environmental sciences, the standards for the research assessments are high. The process is rigorous, however I am happy to share information about this process for if you are involved in a similar assessment process it can be worthwhile to compare your approach with the one being used in Wageningen. Or if you are not involved, you may be soon, as departmental research assessments are being implemented in more and more countries.

The research assessment approach of Wageningen is in fact not typical for Wageningen. The approach is kind of given by the national organizations which work on quality control and management, the Netherlands Association of Cooperating Universities (VSNU), the Royal Academy of Sciences (KNAW), and the Netherlands Organisation for Scientific Research (NWO). They have agreed upon a national assessment protocol, called the Standard Evaluation Protocol. This protocol, which is currently set for the period 2015-2021, is quite extensive, and departments (or chair groups as they are called in Wageningen) have to produce a Self-Assessment Report based on a given format. Then a committee visits the Department, takes interviews with the general management of the school, the Department head and some faculty members. Based on the extensive documentation about the research and the interviews the committee prepares a

narrative report and gives scores on three assessment criteria. Based on that the ranking of the Department within the group of Departments is determined, This very often has funding consequences. At least, the status of the groups is determined for the next six years, which is the research assessment cycle in the Netherlands.

Let me elaborate three elements here: the assessment criteria and scale which are being used, the Self-Assessment Report and the committee issue.

As said, there are three assessment criteria. These are: research quality, relevance to society and viability. The assessment scale is: 4 = unsatisfactory; 3 = good; 2 = very good; 1 = excellent. Score 2 means that the group does internationally recognized research. Score 1 means the group is among the best research groups in the world in its field. Departments have to give extensive evidence of their research performance so that the assessment is really substantiated.

I will shortly describe the contents of a Self-Assessment Report (from the Self Evaluation Protocol of the organizations mentioned above. The Self-Assessment report contains 7 chapters. Chapter 1 is on the objectives and research area. It specifies the vision, mission and objective of the programme, the research area, the strategy of the group and the research environment and the way the research is embedded in the organization and the field of research. Chapter 2 describes the means with which the research programme has been carried out: the resources, facilities, the researchers and the research funding. Chapter 3 is about research quality. It contains demonstrable products which are meant for other researchers in the field and which are being used by peers. This is based on data about the research output, the various types of publications by publication year, and the bibliometric analysis of the publications of the group. I will go a bit deeper into this below. The chapter also presents an overview of the marks of recognition by fellow researchers. These marks need to be demonstrable. Chapter 4 is about the relevance of the research to society. It also has to show demonstrable products and other products of research which are aimed at teaching target groups in society. Not only have these products to be presented, the evidence of their use in practice by societal groups has to also be provided. Together with that, the chapter presents marks of recognition by groups in society which can be showed. So it is not sufficient to just perform activities and publish them for these groups, the actual use and impact of the results has also to be shown. Chapter 5 is about viability. This is described by a benchmark and a SWOT-analysis (Strengths-Weaknesses-Opportunities-Threats). I will come back to the benchmarking. The chapter has to present the SWOT-analysis, and what measures are taken or envisaged to cope with the internal (weaknesses) and external (threats) challenges. Chapter 6 gives a reflection on the SWOT-analysis and a description of the strategic priorities which are set by the group. The chapter also contains a comparison of the current state of play in the research programme with the situation 6 years ago at the time of

the previous assessment. It concludes with a description of the viability of the research programme. Questions here are: Is the group big enough and sufficiently equipped to realise the objectives of the program? Is the program situated in a domain which is growing? Is the scientific and societal demand for the research program big enough? Are there sufficient research funds to finance the research team? Finally, in chapter 7 a description is given of how the group establishes research integrity. In appendices the precise volume of the research staff is presented (full-time faculty are counted for 0,4 fte research time) and mini-CV's of the core research staff (one page per person).

Now two things will be elaborated a bit further regarding the Self-Assessment Report. The bibliometric analysis and the benchmarking.

The bibliometric analysis is complicated if you are not familiar with metrics which are being used to measure the impact of scientific publications. In a former Editorial I have been explaining this a bit in reference to the impact of the JAEE, which is developing fine by the way. Next to Publish and Perish and Google Citations which are both freely available on the internet, universities (can) hold subscriptions to the Web of Science (WoS), owned by Thomson Reuters, and Scopus, owned by Elsevier. The WoS covers the Science Citation Index (SCI) and the Social Science Citation Index (SSCI). The SSCI is most commonly used in the social sciences. It covers a large range of journals, which are categorized by subject. Subject categories for the JAEE are for instance Education & Educational Research (E&ER), Planning & Development and Geography. Every Subject category includes a certain number of journals. Currently, there are 216 journals listed in the WoS in the subject category E&ER. Each journal has an Impact Factor (IF). The IF is related to the number of times articles in the journal are being cited by publications in other journals within the WoS. An IF of 1 means that on average the articles in the journal are being cited once per year. All citations are divided by the articles, which gives the average citation score for a certain period of time. That average is the IF. The relative position of a journal in the journal listed ranked by IF varies by subject category because the length of these subject-specific lists varies; furthermore, the maximum IF differs by subject category: the maximum IF within the category E&ER is 5, while the IF of the journal Science is over 30. The list of a certain category is divided in 4 quartiles, which gives a Quartile number to each journal. If a researcher is not aware of the range of the impact factors of a certain subject category, it is informative to know in which Quartile a journal is. Knowing that a certain journal has a Q1 status immediately conveys the message that this is a good journal in the field.

Now, coming back to the bibliometric analysis, there is a whole range of indicators that are being used in Self-Assessments. First of all, there is the number of publications in each year, and the average number of publications over the review period. But that does not say much about the average

productivity by FTE research input. That can be calculated if the total output is divided by the number of FTE. But in terms of research use it is more interesting to know what the total number is of citations of the publications. Or more precisely, the average number of citations per publication. That can also be compared to the average of the world number of citations in the field. Then, maybe the most important indicator of the citations to the articles of the group is the Relative Impact (RI). This is the average citation score which is normalized and based on the articles of the same age and within the same subject category in the journal. This means that the publications are compared with the global average of citations to publications which are published in the same subject category.

An average RI score of 1 means that a research group's publications are being cited more than 66% of the publications of the same subject category or the same age. A RI of score means that it is better than 95% of the other publications.

Now, regarding the benchmarking, Scopus has an intelligent option to do this. The difference between the WoS and Scopus is that Scopus is more inclusive. It contains publications in peer reviewed niche-journals, which is very relevant for many social sciences. To perform a benchmarking of groups, the team which conducts the benchmarking has to define researchers and groups and to determine which group members will be included in the benchmark. This can easily be done by taking the websites of the benchmark groups and to decide which persons will be included in the group of the given research unit. However, there are some practical considerations to take into account: how can the user determine for how many FTEs a research is part of the program? What to do with research assistants and coordinators? What to do with colleagues who did not publish in journals which are included in Scopus in the given time frame? What to do with retired faculty who are still on the list, but maybe not so active anymore in publishing? For my research assessment I decided to not correct for FTEs, hoping that that would middle out because in my own group there are also part-time faculty. I excluded research assistants, coordinators and non-publishing colleagues. I included retired faculty if there were still publishing. In some cases, I was making groups and defining researchers, but certain researchers cannot be found in Scopus. Apparently they have not been publishing in Scopus journals in the given review period, and therefore left out from the counting.

Based on the definition of researchers and groups, interesting graphs and tables can be produced which show the relative position of one's own group. Whereas WoS tables can be generated by subject category, the tables within SciVI can do comparable things, but doing this is quite cumbersome, because all subject fields have to be done one by one. As a caveat, it has to be mentioned that the tricky part is the inclusion of researchers in benchmark groups. That is the most vulnerable part of the whole exercise. However, selection of researchers can be made transparent by naming them in a footnote of the benchmark figures or

tables. This will prevent fancy compositions of research groups, and enable dialogue with the other groups about the question as to whether the benchmark can be seen as trustworthy.

Finally the committee issues. The fate of the assessment of research groups at the end of the day is in the hands of the peer review committee. The committee members act like a jury, and they have to be convinced by the Self-Assessment Report and the presentation of the research group for and the discussion with the committee . There are various issues here. Are the assessors independent? They have to be, as this is a criterion for selecting them in the committee. But it is a small world, and many scholars have different palettes, and certain groups have already certain images. Is the process powerful enough to enable making second impressions? Does a group, which did not perform very well in the previous review period, have the chance to be rated in a fair way? Or, does a group which has a solid reputation get lower scores based on a fairly mediocre research performance during the current review period? These are questions which are hard to answer. At the end of the day, the assessment results are based on human decisions which are based on the inputs given to them.

If you are engaged in a research assessment process, or if you are anticipating one, do not hesitate to contact us for more information, or let us know what your experiences are. Because for sure, it is impacting our research work to a growing extent.

MARTIN MULDER, Editor-in -chief