The Vicious Circle of Evaluation Transparency – An Ignition Paper

Miloš Jovanović[1]

[1]*milos.jovanovic@int.fraunhofer.de*
Fraunhofer Institute for Technological Trend Analysis, Appelsgarten 2, 53879 Euskirchen (Germany)

Introduction

The present paper introduces a model, which describes different phases that typically occur in situations, in which a researching subject (e. g. an author, an institution, a country etc.) needs to be evaluated and in which some kind of reward (e. g. monetary in the form of a bonus or funding) is based on this evaluation. This model, the present author calls it the "vicious circle of evaluation transparency", will be underlined by giving examples for each of its phases. In order to be able to observe a process that is described by this model, there first needs to be something that is to be evaluated, for example a research group at a university. Such a need normally comes up, when money is to be divided among different groups or focused on one. The problem of evaluation and rewarding is at the core of the model (see Figure 1).

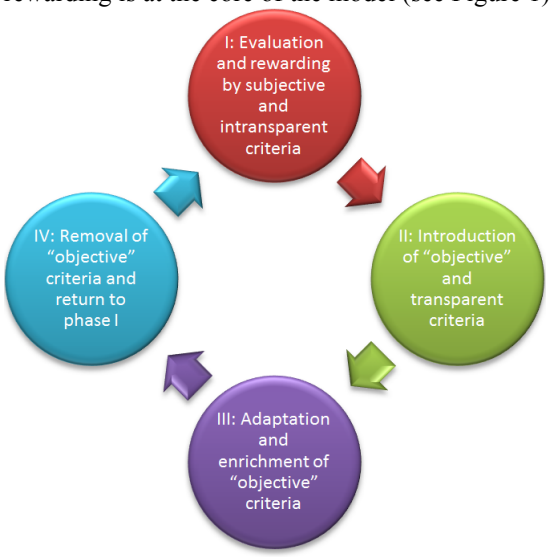

Figure 1. The "vicious circle of evaluation transparency"-model.

Phase I – Evaluation and rewarding by subjective and intransparent criteria

The first question that might come up in such a situation is the question of how to evaluate a research group. In hierarchically organized universities the leader of a department will decide whether or not and how this group is evaluated. Very often, this person is also the one that conducts the evaluation and, based on this, determines the type and amount of a reward or funding (or some kind of penalty, if the evaluation is negative). In today's world of vast amounts of digital data, it might be hard for only one person to do such an evaluation. Naturally, having one person alone evaluate a group's performance and decide on rewards will lead to a number of persons feeling unfairly evaluated, because the evaluator might not know about their achievements or their work in detail. This criticism might be alleviated in part by expanding the number of evaluators, for example by having a board of evaluators. Another possibility is to improve the transparency of the evaluation by documenting and publishing certain evaluation criteria by which the evaluated subjects can read about the evaluations and try to strive to get a better evaluation. These evaluation criteria are a first step towards phase II of the model.

Phase II – Introduction of "objective" and transparent criteria

These evaluation criteria might be subjective. For example "Quality of work" can be a criterion that is evaluated differently by different people. In order to make evaluation criteria comparable and independent of the evaluating person, "objective" criteria are often introduced. The reason why the word is put into quotation marks is due to the fact that very often these "objective" criteria are not objective at all. The introduction of "objective" and transparent criteria is a simplification of reality, an attempt to put parts of reality into some kind of a score in order to compare them with each other. Bibliometric indicators are one example of such a simplification. In many countries, different kinds of "objective" and subjective evaluation criteria have been introduced, for example in Italy (Abbott, 2009). Normally, these "objective" evaluation criteria (often in the form of different kinds of indicators) are communicated transparently. And while transparency is an important factor for these evaluations, it also leads to one problem in this phase: the fact that the evaluated subjects, in our example researchers at universities, react to the evaluation by starting to change their behavior, in order to maximize their scores in the evaluation. Of course, one reason behind evaluation is to positively influence the behavior of the evaluated researchers. But in Germany, for example, this has led to authors aiming to publish more in internationally known journals that have a US publisher and which are more general in their scope (Michels & Schmoch, 2013). This underlines the fact that authors do not base the decision in which journal they wish to publish in on scientific reasons

alone and constitutes a negative change of behavior. Also, some of the evaluated subjects might complain that the evaluation criteria do not reflect their work adequately and need to be refined. This leads to the next phase.

Phase III – Adaptation and enrichment of "objective" criteria

The need to fairly represent and evaluate researchers' work in the evaluation criteria and to adapt these in order to not allure unwanted change of behaviour leads to reforms in the evaluation system, e.g. new or a mix of indicators are proposed. The current discussion on alternative metrics is an example for phase III (e.g. in Haustein et al., 2014). The problem here is, that phase III is actually reintroducing parts of the simplification of reality, which was conducted in phase II. The evaluation criteria become more complicated again. A country example for this phase is the Czech Republic, which introduced performance-based research funding (phase II). A study by Vanacek (2014) found that the number of publications increased very quickly. He shows that in comparison to the quickly growing number of publications the quality seems to have stagnated and recommends reworking the procedure of evaluation and performance-based funding in order to increase not only the number of publications but also their quality (phase III). But for some research communities, the adaptation and enrichment of the "objective" criteria is no option. Instead, these criteria are rejected. For example, there is an ongoing discussion in the mathematical community. Authors note that bibliometric data lose "crucial information that is essential for the assessment of research". It is pointed out that bibliometric indicators can be manipulated and lead to undesirable publishing practices (Adler, Ewing, & Taylor, 2009). The authors also dismiss reputation, as determined by surveys as a possible way of measuring the quality of a journal. The evaluation of journal editorial processes is not seen as a good way of ranking journals either. Instead, the authors recommend an "honest, careful rating of journals based on the judgment of expert mathematicians", which is the point, where phase IV starts.

Phase IV – Removal of "objective" criteria and return to phase I

Concretely, the IMU recommends that a rating committee of 16-24 experienced and respected mathematicians should be appointed. Without going into too much detail, this committee (via various panels) is then supposed to rate the different journals and assign them to tiers (ranging from tier 1 = high quality journal to tier 4 = low-class journal) (Journal Working Group, 2011). This system is similar to the peer review process. Introducing evaluation by a committee of experts, either by rejecting "objective" evaluation criteria or because the evaluation system has become too complicated, brings the model full circle. The evaluation has reached phase I again. One should note that in phase II of this new cycle, the criteria probably will not be the same as in the first cycle. Newly developed and more sophisticated criteria will take their place.

Conclusion

It is this author's personal opinion that the above described model of evaluation transparency not only describes a typical process in which bibliometric indicators are involved but rather evaluation processes in general. If this is the case, one may discuss possibilities to change this, since a cycle like this is not an optimal solution. An option might be the introduction of diametrically opposed evaluation criteria so that an evaluated subject could not be good in all criteria. Another idea that might serve to fan the discussion on this topic would be the introduction of a changing system of criteria, akin to the disciplines at Olympic Games. The criteria could be published a year before the evaluation takes place and would change each year. This would be a transparent system, while the evaluated researchers would not need to change their behavior in a negative way because the next year the criteria would be different. Whatever changes might be introduced, it is this author's opinion that the vicious circle has to be stopped and replaced by a different system that leads to the desired goal: a fair evaluation of research.

References

Abbott, A. (2009). Italy introduces performance-related funding. *Nature, 460*, 559.

Adler, R., Ewing, J. & Taylor, P. (2009). Citation Statistics: A Report from the IMU in Cooperation with the ICIAM and the IMS. *Statistical Science,* 1-14.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., &Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics, 101*, 1145-1163.

Journal Working Group (2011). *Report of the IMU/ICIAM Working Group on Journal Ranking.* Retrieved March 10, 2015 from: http://www.mathunion.org/fileadmin/IMU/Report/WG_JRP_Report_01.pdf

Michels, C. & Schmoch, U. (2013). Impact of bibliometric studies on the publication behaviour of authors. *Scientometrics, 98*, 369-385.

Vanacek, J. (2014). The effect of performance-based research funding on output of R&D results in the Czech Republic. *Scientometrics, 98*, 657-681.