Recommendation on the evaluation of individual researchers in the mathematical sciences

endorsed by the IMU General Assembly on August 11, 2014

I. Introduction

The question of how to evaluate and rank the work of academics or scientists has been a recurrent theme since the early days of universities. This issue is closely entwined with questions about the role of scientists in society and the role of education, so that it is difficult to discuss it without considering the wider contexts as well. A number of the aspects of an evaluation process are not specific to mathematics or mathematicians: Each scientific discipline faces similar issues. Hence, interdisciplinary working groups have been producing documents/guidelines on these questions for many years. The present document deliberately does not focus on the rather timeless and important issues that have been often and thoroughly discussed (rules about conflict of interests etc.), nor what is common to all sciences. Rather its goal is to address the following two specific questions:

* What aspects of evaluation are specific to mathematics? A reason for focusing on this issue is that mathematics often needs to be treated somewhat differently from those sciences where teamwork and funding play significantly different roles. Mathematicians have often found it difficult to make this point on interdisciplinary panels, with unfortunate outcomes from their scientific standpoint. The present document aims to help with discussions involving colleagues from other disciplines.

* What is new, i.e., what are the important side-effects of recent developments (internet, internationalization, the growing scientific community, economic constraints, generalization of audit rules) on the way in which mathematicians are evaluated, and which have had strong negative effects that need to be corrected or counterbalanced.

When stressing the specificities of mathematics, one should not forget that a large and important part of the mathematical community is, for obvious and good reasons, working a little differently from the more academically inclined mathematician. In such cases, the “standard” (but specific) evaluation criteria that we will discuss (mostly based on a detailed study of research papers) in the next section of the present document have to be adapted again. For instance, for applied mathematicians involved in projects with confidentiality clauses and/or industrial applications and/or software development, for mathematicians involved in interdisciplinary work or in mathematical education, one needs a different perspective. This document is therefore divided into two parts: The first part, which deals with “generic” academic mathematicians, and a second part commenting on aspects of evaluation in several other important cases.

An important preliminary statement: The evaluation of the performance of an academic is used for many purposes (hiring, promotion, grants etc.) and in most cases the assessment of research activity is only one of many parameters. Many other aspects are essential for a well-functioning academic environment, and are important factors to take into account when such decisions are made. **The present document is commenting only on the part of the evaluation dealing with the research activity; in the appendix, we give some examples of other criteria that can be taken into account in the academic evaluation of individuals.**

II. The case of the "generic" academic mathematician.

Generalities, specificities of mathematics, dangers of semi-automatized evaluations:

Assessment criteria are not universal and uniformity of evaluation criteria is not necessarily a goal. For example:

* University systems in different countries are different, and the variety of individuals with different academic backgrounds is part of the richness for our international community.
* The evaluation of the work of mathematicians when choosing the recipient of an international prize is not the same as when deciding whom to hire for a junior faculty position.
* National communities may seek to take special measures, for example where they perceive potential weaknesses (possibly, a lack of innovation and originality in some areas or a lack of rigour and clarity in others) that they wish to correct, and therefore take them into account when making decisions.

It is standard nowadays for an evaluation committee to examine:

* an individual's publication list (including the names of journals; co-authors; the number of published pages);
* a research statement in which the individual describes the research in a more general context;
* one or several evaluation letters written by specialists, who are supposed to have read the individual's papers.

These specialists may, or may not, belong to the committee, and their contribution is essential. Indeed, a proper evaluation of the significance of research papers requires a close examination by an expert who is actually able to understand them. The use of semi-automatized quantitative evaluations based on journal factors can easily lead to mistakes. At first glance these methods look objective, scientific, and not subject to manipulation or controversy. However this is not so, and they can have some very negative side-effects:

* High-level research is driven by originality, invention and risks (one starts an ambitious project without any guarantee of success). All these aspects would be penalized by a standardized evaluation based on bibliometric data alone.

* Bibliometric evaluation leads to an increase in the number of published papers, because it favours publication of series of papers where results are improved step by step. For the sake of mathematics research in general, it is more important that papers are well-written and in final form. One highly innovative paper is usually more important in the long run for our community than ten technical but routine papers, regardless of the journal in which they are published. In fact, prepublication servers should make it possible nowadays to post prepublications that are not submitted to publications, but will be incorporated in a longer/cleaner/more definite paper that will be published later.

* Impact factors: It is not uninteresting to look at the data that measures how much a given paper has been cited by other papers (such data is made available by the mathematical community itself, for instance by the AMS in MathSciNet), but a lot of care is needed when handling it. First, some fields of mathematics tend to publish many more papers than others, so that one cannot compare such data for a person working in one field (say, category theory) with someone working in another (e.g., biostatistics). Second, it is very easy to artificially create a blow-up of bibliometric data (for instance by cross-referencing etc.) and to manipulate impact factors.

It is therefore important to encourage mathematicians who serve on panels to explain to scientists of other disciplines that bibliometric evaluation is particularly inappropriate for mathematicians. We hope that the present document can help in making this point. It is worth stressing that mathematicians are not advocating that other sciences should change their specific evaluation criteria; IMU does not claim that it knows the best way to evaluate chemists or economists. The conclusion of this paragraph is the following somewhat obvious statement, which is the core of the present document:

Nothing (and in particular no semi-automatized pseudo-scientific evaluation that involves numbers or data) can replace evaluation by an individual who actually understands what he/she is evaluating. Furthermore, tools such as impact factors are clearly not helpful or relevant in the context of mathematical research.

It might look tempting to produce alternative bibliometric tools (keeping in mind that most impact factors are produced by commercial companies for whom it is a business), but this is not something that IMU wishes to be involved with, given the intrinsic negative side-effects of such tools.

The “audit” philosophy and science, explosion of evaluation activities.

The role of “evaluation” has become more and more important in recent decades. The concept of “auditing”, probably first developed in a business context (accounting and then management), has now permeated many parts of Western societies. It is based on the belief that uniform, comparable, objective, evaluations of almost anything, people, organizations, companies, products etc., are possible. In particular, many funding bodies are now so convinced of the importance and universality of the evaluation of scientific activities that they tend to insist on using their evaluation rules, often based on semi-automatized “objective” criteria such as Key Performance Indicators (KPI), even though most experts agree on the fact that these methods are not well-adapted to science in general, and to mathematics in particular.

Another negative side effect of this "generalized audit philosophy" is the proliferation of evaluation activities, which arise because each layer of decision-making wishes to perform its own evaluation. While it is clear that some level of evaluation activity is useful, and that every active mathematician could in principle devote some of his/her time to evaluation of others' work (this starts with the most important and essential part, which is the refereeing of research papers submitted for publication), it is also essential that they keep as much time as possible free for their own research. The proliferation of evaluation activities is a real danger. Moreover it induces a change in the perspective of scientists themselves, i.e., in the way they do and present their own research. The primary goal of research is not to get a good grading at an evaluation, but to simply make progress in understanding things. Shifting these goals would have again very negative consequences.

Smaller scientific communities.

How can all scientific communities get access to a sound and sensible evaluation procedure, and in particular to appropriate evaluators? Smaller countries, or those with very heterogeneous research activities, can find it very difficult to obtain reliable and objective information about the level and quality of their research output. The alternative often seems to be a choice between two poor options: rely on the local community (with the obvious danger of self-evaluation and conflict of interests -- clearly to be prohibited), or a semi-automatized bibliometric system as discussed above, which cannot be viewed as a positive long-term way of dealing with this issue. Another solution is certainly desirable.

There may be a case for ***creating a supra-national structure*** to help in such evaluation activities and it is reasonable to ask whether it is IMU's role to implement such an idea. Arguments in favor include the question "who else?". The main argument against it is that IMU's main goal should be to bring mathematicians together, and not to be a source of tensions that such evaluation activities inevitably create. The dangers that could arise if the IMU gets directly involved in such activities seem to outweigh the benefits.

III. Additional specific comments

Multi-disciplinary and industrial mathematics.

As stated above, the assessment of mathematicians should be based on careful evaluation of their scientific work and not on semi-automatic KPI of any kind. In this paragraph, we draw the attention to special issues arising in the evaluation of mathematicians who are strongly involved in multidisciplinary projects, either in academia or in industry.

Attention is restricted here to mathematicians who have developed novel mathematics *and* used it to solve an applied problem, motivated by challenges from other sciences and industrial (or other) applications. This workflow, modelling-research-development-application, that is of major importance is of a somewhat different nature than the one discussed above. We stress again that work that only involves direct applications of already existing mathematical tools or techniques is not discussed in the present paragraph (this latter type of work can be assessed by the criteria relevant to the applications area only).

Because of the extreme diversity of publication cultures in multidisciplinary projects and in industry, it is even more crucial to base an assessment of this type of research activity on expert evaluation, which can be a very demanding but necessary one. Given the importance of such activities, both in terms of applications as well as for mathematics itself, it is of particular importance to perform this difficult task well.

While the previous general remarks about the evaluation of the mathematical novelty remain true, additional criteria should be used to recognise some additional and specific challenges:

* The benefit of the mathematical perspective to the community of the "problem owner" is very important. Therefore it is allowable that some lack of complete mathematical details or theoretical importance (not to be confused with lack in rigour or novelty) is compensated for by relevance to the "partner" community, in which other indicators can be significant. For example, in publications in other sciences, the first nominated author has a strong meaning, while the alphabetical order is the tradition in mathematics papers. So to conclude this paragraph: Assessment in such activities can include criteria used in other sciences, but **in addition, not as a substitute,** to the relevance on mathematics itself.

* Additional issues arise for mathematicians working in industry or in industrial projects. Here special restrictions may prohibit full publication of the scientific work, either by intellectual property restrictions or (more often) by a lack of time to develop full detailed proofs. Panels or evaluators have of course to take this also into account.

* A related issue is the fact that the "end-product" of such research and development activity is not necessarily a research paper: It can be for instance a software, the development and implementation of which is a very fundamental and time-consuming aspect that can be also of mathematical nature. This example illustrates the variety of possible important contributions that should be taken into account when performing an evaluation.

Mathematics Education.

What follows are some brief comments on the evaluation of researchers in mathematics education. This is a community that is organised very differently from one country to another. For instance, its members are sometimes part of the formal academic/university community, sometimes affiliated to teacher's associations, and sometimes part of the Ministry of Education. Moreover it is a field with a great diversity in aims, foci, methodologies and programmes, ranging from the epistemological analysis of parts of mathematical knowledge to be taught (usually in an academic context), to the design and analysis of a short term classroom experiment (typically involving teachers), to the design and analysis of teacher education programmes, or to large research studies carried out in collaboration with schools. Sensitivities to different categories of students (from low achievers to gifted), or different social and cultural backgrounds, are also factors and require inputs from cognitive and social sciences. All these disciplines are necessary for achieving the ultimate aim, which is to improve the quality of mathematics teaching and learning at all levels.

This area therefore involves some mathematics, but has very significant inputs from all the above-mentioned fields. Hence, evaluating contributions on this topic requires a blend of criteria that are suited to each of these fields.

IV. Appendix: A non-exhaustive list of aspects that can or should be used in evaluation:

All these familiar aspects of academic life are essential and our community needs them to be performed properly. They can also contribute in indirect, but significant ways to high-level research. Note that this is quite a long list, and that no individual is supposed to tick all boxes (writing computer software does for instance only concern a fraction of the mathematical community, not all researchers have teaching duties, etc.).

* Research articles in international journals;
* Research monographs, textbooks, classroom notes;
* Applications, production of software, programming code;
* Special programs organized, especially in institutions where such activities are selected through competitive evaluation of proposals;
* Conferences and seminars organized, especially in institutions where such activities are selected through competitive evaluation of proposals;
* Courses taught, new courses created, teaching awards;
* Refereeing;
* Academic awards;
* Supervision of students: PhDs, masters, undergraduates, future teachers;
* Elected membership in learned societies and other academic institutions;
* Advisory activities, including editorial work for international journals;
* Outreach activities: popularization articles, public lectures, lectures or competitions in schools, role in teachers associations, etc.
* Administrative duties;
* Efficiency and reliability.

This document was prepared by a committee set up by the Executive Committee of the International Mathematical Union, composed of the following mathematicians:

* *Mariolina Bartolini Bussi (nominated by the International Commission on Mathematical Instruction (ICMI))*
* *Carlos Cabrelli (nominated by the Commission for Developing Countries (CDC))*
* *Andreas Schuppert (nominated by the International Council for Industrial and Applied Mathematics (ICIAM))*

and chaired by

* *Wendelin Werner (IMU Executive Committee).*