Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

Round Table: The Use of Metrics in Evaluating Research

J.M. Ball*

The use of metrics for evaluating research is a hotly debated issue. The IMU/ICIAM/IMS report on Citation Statistics [1] highlighted the dangers of uncritical use of impact factors, which play an increasing role in funding, promotions and library purchases. Are impact factors and other such indices good measures of journal quality, and should they be used to evaluate research and individuals? What can be done about unethical practices like impact factor manipulation? Is there a role for metrics in evaluating research? Are there better alternatives?

These were the topics of discussion at the ICM 2010 Round Table on Thursday, 26 August, between 6 and 8 p.m. It was chaired by John Ball, and organized by IMU's Committee on Electronic Information and Communication (CEIC).

This record of the Round Table consists of edited and shortened versions of the presentations by the panellists, together with excerpts from some of the contributions by participants in the discussion. A complete video is available at the IMU website `http://www.mathunion.org/publications/historic-material`.

Introduction of the Panellists

John Ball. Good evening. I'm substituting for the IMU President, László Lovász, who is actually here but has some problem with his eyes that make it difficult to be in front of bright lights. This round table is a sequel to the 2008 Citation Statistics Report, which was a joint report of the International Mathematical Union, the International Council for Industrial and Applied Mathematics and the Institute of Mathematical Statistics. The writing group for that report was chaired by John Ewing, who was then Executive Director of the American Mathematical Society. This report had a very good reception and

*Mathematical Institute, University of Oxford

it drew attention to the dangers of uncritical use of the impact factor as a statistical measure of journal quality. We have a very interesting panel:

Doug Arnold is Professor of Mathematics at the University of Minnesota in Minneapolis and currently is President of SIAM.

Malcolm MacCallum is the Director of the Heilbronn Institute at the University of Bristol, and was a consultant on the United Kingdom Research Excellence Framework, which is going to be the next evaluation of research in the UK.

José Antonio de la Peňa was Director of the Mathematical Institute at the National University of Mexico and is a former President of UMALCA, the Mathematical Union of Latin America and the Caribbean, and he is currently Deputy General Director for Science at the National Council for Science and Technology, Mexico.

Frank Pacard is Professor of Mathematics at the Université Paris Est-Creteil, and is Scientific Advisor of Mathematics in the French Ministry of Higher Education and Research.

Presentations by the Panellists

Doug Arnold. I will focus mostly on one research metric: the Impact Factor (IF), which is simply the average number of citations made in a given year to a journal's papers from the preceding two years. It is intended as an easily used journal quality measure, but, as I will demonstrate, it is fatally flawed.

The Citations Statistics report found many failings in the IF design as a proxy for journal quality, but I am going to focus on something else: Goodhart's law and IF manipulation. Goodhart's law states that: 'When a measure becomes a target, it ceases to be a good measure'. An example used in economics is that if a nail factory in a centralized economy is judged on the number of nails produced, pretty soon they will figure out they should make lots and lots of tiny nails. If it is judged on the weight of the output, they will start making very big nails. The metric ceases to be an accurate proxy for the more complex attribute, say productivity, which was intended.

How do people manipulate the IF? One way was demonstrated by an editor of Journal of Gerontology A. Every January, he would write a review article citing all the articles of the preceding two years, and so acquire 200 impact factor citations, more than most math journals get altogether. Another approach is that 'the editor cultivates a cadre of regulars, who can be relied upon to cite themselves and cite the journal shamelessly'. Such a bargain between authors and editors is difficult to detect. Authors are often under citation pressure, but the editors of the Balkan Journal of Geometry and Applications put it in their instructions to authors: '[it] is advisable for each accepted paper to contain citations to articles published during 2006-2008 in our journals'.

In order to determine to what extent such manipulation is actually damaging the IF, I compared it to expert opinion, for which I used a journal ranking carried out with broad and careful expert consultation as part of an Australian research assessment exercise. This study [2] demonstrates that many of the bottom class, B and C, journals have higher IF than a significant proportion of the journals that are judged by experts to be the best in their subfield. The grossest anomaly is The International Journal of Nonlinear Sciences and Numerical Simulations (IJNSNS), which has had the highest IF in all of applied mathematics by a large margin for the last four years running, although as a B-rated journal there are roughly a hundred journals in front of it according to the Australian rating. Working with librarian Kristine Fowler, I studied this case in detail.

Which authors gave IJNSNS all those citations? It turns out that 30% of the citations were from just three authors, and these were the Editor-in-Chief, who cited his own journal 243 times in the IF window, and two other editors. (For control we looked at high reputation journals in applied mathematics, and found it is rare to have more than a few citations come from a single author). As a second approach, I looked at the highest citing journals for IJNSNS. First place is a single issue of the Journal of Physics Conference series, which provided 294 citations. This was the proceedings of a conference that the IJNSNS Editor-in-Chief organised and controlled the peer review for. The next highest citer was a special issue of a different journal that was again organised by the Editor-in-Chief of IJNSNS. Similar issues arose with other highly citing journals, so that more than 70% of the citations were under the immediate control of the IJNSNS editorial board. A different sort of check is to look at the citations outside the IF window. With IJNSNS, 72% of their citations are in the two years that count for the IF and only 28% in all the other years. With SIAM Review, for example, it is the very opposite: only 8% fall in the IF window.

Although I have been mainly concerned with journals, the people who make the IF say their citation database 'can rank top countries, journals, scientists, papers and institutions'. Who do they think is the top mathematician? Ji-Huan He, the Editor-in-Chief of IJNSNS! He was named by them as a 'Rising Star' in Computer Science; he had a 'New Hot Paper' in Physics, another one in Mathematics; a 'Fast Breaking Paper' in Engineering. And then in 2007-2008, they named 13 scientists in all of science as 'Hottest Researchers of the Year', and he was the only mathematician, a performance he repeated the next year.

To conclude, there is little doubt that IF is highly flawed as an indicator of journal quality. I showed how a journal which is roughly number 100 in applied mathematics moved itself up to number one. There are certainly many other cases in which journals manipulate the IF more subtly, moving themselves up (and so moving more honest journals down) five or ten places. We cannot expect an easy formulaic fix. If we agree to judge quality by counting citations, Goodhart's law indicates that we will fail. However, there is a need, e.g. for library purchase decisions, for an easily consulted indicator of journal quality.

The IMU and ICIAM have discussed this and taken a big step forward this month by resolving to develop a plan for a joint ICIAM/IMU method of rating journals, based on expert opinion. This has the potential of providing truly useful information to those who need it, while returning the process of judgement to us, the experts.

Malcolm MacCallum. I think a lot of the discussion is going to centre on impact factors and citation indices. I want first to draw your attention to the other sorts of metric used, in particular in the UK Research Assessment Exercise (RAE). It had three headings: 'Outputs', 'Environment' and 'Esteem'. 'Outputs', essentially papers, and 'Esteem' were assessed by peer judgement. In judging Environment, we had about 20 metrics presented to us, for example the number of Research Assistants per full time equivalent members of staff. There was no sane way to use them all.

Some of them were really input measures, and it is very hard to establish how effectively they had created output or knowledge transfer. My own suspicion is that the less income you have, the better you use it. Some are outside institutional control. Some are historical: you may be very attracted to where, say, Hardy worked although Hardy died long ago. In fact, I think too many of them are self perpetuating, rather than reacting to current research quality. Even if you accept them as valid, there are still various ways to use them. For example, in considering the total research income per person against the size of departments, do you reward the department that earned most or the one spectacularly effective with the number of people they have? Kenna and Berche [3] found that in almost all disciplines there is a critical size above which the research quality tails off. Unfortunately this isn't a very useful message for this assembly because while true for applied, it is not true for pure mathematics.

In the UK, they plan to replace 'Esteem' by 'Impact', meaning economic, social or cultural but not scientific impact. That has to do with why a government should fund research at all, which is a very fair question. But I think the specific way that they are intending to answer it is not the right one. The Royal Astronomical Society and the UK Institute of Physics, concluded 'we can't do it' and 'we don't think it's doable' Fabian [4].

Now I want to come back to bibliometric measures. There has been a lot of research on citation data, and the many problems it has, such as consistency, coverage, nationality and gender biases, indexing, 'obliteration', discipline size and citation practice etc. (see e.g. Blustin [5], and for fun [6]). In RAE we specifically did not use bibliometric data. But after I had read and assessed each paper, I looked up its citations. That caused me to change my opinion on only two or three of the 400 papers read. So citation information can be useful, but it has to be interpreted with a knowledge of the sociology of the discipline and an understanding of the mathematical content. For the Expert Advisory Group on the replacement for RAE, there was a pilot of looking at citations of individual papers. The resulting data was given to us to compare with our

actual assessments. There was general agreement across all subjects that the bibliometric data could not have been used without some serious injustices.

As a journal editor I find impact factors a useful measure of how we are doing against the competition. But I do not believe one can judge a paper by where it appears: thus I do not agree with Professor Arnold's proposals.

In summary, I have two messages.

1. To bureaucrats: no metric is safe for use without human interpretation. You have to be very careful to realise that correlation does not imply causation. One of my colleagues claimed that the UK ranking of institutions was very tightly correlated with the number of gardeners they employed!

2. To those entirely opposed to metrics: they can be a useful sanity check, providing you don't try to use too many or make them too complex.

Frank Pacard. I wanted to say something about the situation in France concerning the use of citations and metrics to evaluate mathematical research, either by the government or by the universities. First of all, there have been some changes in the French higher education and research system and, to understand how citations and metrics are used, it is very important to understand how the money supporting research is now distributed. In France almost all the money for mathematics comes from the Ministry of Higher Education and Research but it travels through many different channels before it reaches mathematicians. As far as the assessment of research is concerned, the government has created some evaluation agency to this effect. So far, the evaluations from this agency are not based on the use of metrics and complicated impact factors, there is though a definition of an 'active researcher' which depends on the number of publications. Therefore, everything seems to be going smoothly in France with a very limited use of statistics in the assessment of research.

However, looking closer you find that there is also an institution whose work is to provide statistics based on the number of publications and citation. Even though these statistics are not used officially to evaluate a research department, they are becoming more and more popular to measure for example the strength, weakness and evolution of the different fields in a given part of France (for example, all sciences in the south west of France). These data are also available to all actors of the research system. These statistics can be very precise and can cover very different scales : at a scale of a whole country up to the scale of a research department.

For example, in my own university, statistics about the number of publications of the mathematics department (which is a small department) are received and, as you can imagine, interpretation of the data can be rather controversial at such a small scale. French universities are now autonomous and have more freedom in their scientific policy. In particular, to some extent, they can decide to give more support to department A rather than to department B and the government does not provide them with any guide on how to distribute the money among departments. As a consequence, there is more and more pressure

to make use of metrics in order to distribute the money as best as possible, using possibly some very complicated mathematical formula.

Even though French mathematics is very strong, it is fair to say it only corresponds to a very tiny subset of the French research system. What is true at a national level is also true at the level of a university where mathematics departments are now in direct competition with other departments of other sciences whose weights are much bigger and for which the use of metrics seems more natural. This is where I see that there is some danger for mathematics in France. My experience shows that there is a strong temptation to use metrics not necessarily coming from the top of the research evaluation system but also coming from the bottom of the evaluation system, because metrics are a rather quick and convenient way to compare people or departments from different fields!

On the other hand, the use of metrics at a large scale (say the scale of a country like France) is probably worth considering and, carefully analysed and complemented, can give some interesting insight on the strength and weaknesses of a given field. For example, the relative share of publications of French mathematicians in the world has decreased over the past years slightly faster than expected. This is an interesting piece of information but unfortunately, since there is no further analysis of this information, it might be improperly used. Also, people in charge of building the statistics based on publications are well aware that some indices used are not adapted to mathematics (for example, the number of citations in the two years after publication is not very meaningful in mathematics) and they would be very interested in having some more meaningful formula.

To conclude, I would say that the situation concerning the use of metrics in France is still not completely clear. There is some pressure to use them and we have to be very careful in the next years to protect ourselves from improper intensive use.

José Antonio de la Peňa. Citation indices, originally designed for information retrieval purposes, are increasingly used for research evaluation. The concern that the consideration of these indices is distorting the evaluation of the individual work has passed, in the last few years, from corridors to main stream journals.

In the developed countries, at least since the second half of the 20^{th} century, science is accepted as a social, cultural and economic asset. Although the relevance of scientific work has been evaluated from decades back, current evaluation practices have a recent history that respond not only to academic needs, but to conceptual changes of political, economic and social character.

In evaluating scientific work, the criteria used are expected to have universal validity (as much geographic, as thematically), to be objective, to be simple to measure and to determine, as far as possible, the quality of the work. The criteria used so far show many limitations and misinterpretations. Notably, the use of impact factor of journals as a measure of the quality of the science

published and, still worse, the quality of the individual papers published in those journals, is an extended practice without a solid support. Even Eugene Garfield has warned against some abuses: 'It is absurd to make comparisons between specialist journals and multi-disciplinary general journals like Nature'.

To check the evaluation practices in Latin American countries, we asked friends from Argentina, Brazil, Colombia, Chile, Mexico and Venezuela. Here I quote just a few answers to illustrate the discussion:

Q1. *Are indices (such as number of papers, number of citations, impact factor of journals, h-number, etc) used for the evaluation of mathematicians in your country? If yes, which indices are prefered?*

Chile: In general no. Up to now the committees of mathematics agree on the quality of the journals to evaluate the research projects or CV. Sometimes they use, as complementary information in the analysis, some citation indices.

Colombia: In the public universities, the salary of the professors depends on the numbers of papers.

Venezuela: Yes, in some cases. At research institutions, the tendency is to use all those indices to evaluate researchers, but not so much at universities.

Q2. *Who promotes the use of these indices (the administration, scientists in general, mathematicians in particular)?*

Everybody: the administration, in first place; scientists of other fields, as second.

Q3. *Is it considered that the use of indices provides a more: efficient, scientific, fair, objective way of evaluation? Who thinks so?*

Most: I guess that some groups of scientists look for efficiency and some kind of 'fairness'.

Q4. *In your opinion, what is the effect of the use of these indices?*

Most: I believe they do add value to the evaluation, if used carefully and in combination with other parameters.

Argentina: the use of indices is helpful to discriminate between real scientists and those who pretend to do scientific work but have no impact whatsoever.

Chile: I do not know the effect for all areas, perhaps in some of them the systematic use of indices could be useful (but, at the end the prevalence of indices would mean that the work of specialists is not necessary). A systematic use of indices in mathematics will constitute a big catastrophe for its development (an enormous deformation that could affect quality for a long time).

Q5. *Could you give an idea of the general feeling of (dis)satisfaction concerning evaluation among the scientists (in particular, mathematicians) in your country?*

Brazil: The general feeling is actually very positive, among mathematicians and among scientists in general. This is perhaps because the scientific community itself is directly in charge of the evaluation.

Chile: People that have been part of the local evaluation committees say that there is mutual dissatisfaction between mathematicians and other groups of scientists.

Comparing the use of impact factors to measure quality of research with the story of the measuring human intelligence by means of the IQ, we point out the misunderstanding of thinking that a person is intelligent because they have a high IQ. Similarly, we are pushed to believe that a scientific paper is good because it is published in a journal of high impact factor. This is my last argument: I would call it the mismeasure of science, to keep the parallelism with the situation described by Stephen Jay Gould. It is a complete misconception to transfer the value, whatever the impact factor measures, from journals to articles. It should be made in the converse way, after all, a journal is not more than a collection of papers. The only meaningful definition for the impact factor of a journal is the mean value of the impact factor of the papers it publishes. If this is so, it is the impact of a scientific article which should be discussed: is it possible to give a sound definition?

General Discussion

Doug Arnold. While we're waiting for someone to pluck up their courage, let me respond to just one misimpression which may have arisen from Malcolm's talk. He said one cannot judge a paper by where it appears and for that reason didn't like my proposal. So I want to make clear that I agree 100% with Malcolm that one cannot and should not judge a paper by where it appears. In fact in some cases it might be wise to choose a lower impact journal for an excellent paper, for example to help strengthen the journal. My proposal to rate journals is in no way aimed at judging individual papers, and any report that comes out of it would clearly state that. It is a way to get a sense of a quality of a journal for reasons like library purchase decisions, helping the editorial board to know how their work is going and so forth.

George Andrews, Penn State University, USA. I'd like to ask Prof MacCallum, since you say you do not accept Doug Arnold's proposals, I wonder if you are not disturbed by, not the manipulations and outliers, that were in the graph, but the discrepancy that he described between the top level journals, as people assess them, having a lower impact factor than really badly ranked journals.

Any solution is going to have problems, but aren't the problems mitigated somewhat by Doug's proposal?

Malcolm MacCallum. I think that there are certain problems that would be mitigated but what worries me are the ways in which this is likely to be used, and the degree to which it seems to be going along with the idea that you can make judgements by where something appears. I think we should simply be opposing use of data on journals for this kind of purpose. What was shown in the comparison you refer to doesn't surprise me because different journals appeal to different subcommunities or accept papers with a different kind of angle or approach.

Doug Arnold. So I just want to repeat again that there was never any suggestion that one should use the journal quality, no matter how carefully measured and determined, as a way to rate papers, or what you call products of research. I know you have been very involved with rating products of research and you may think that is what this proposal is for. The proposal is to rate roughly, to give a rough idea of what we all know as mathematicians, to put down what we all know about the quality of journals.

Why do we want to do this? We want to do this, for instance, because people must make a decision on which journal their libraries are going to subscribe to. If they don't have enough local expertise in the area then the library must make a decision based on data. Right now they are making such decisions based on seriously flawed data, and we were hoping to replace that with reasonable data which reflects the expert opinions of the people who look carefully at the journals. You can say that people might misuse that, but in fact people are misusing a highly flawed database. We can create one that is less flawed and with clear instructions of what it can be used for and what its limitations are. The fact that somebody might refuse to honor those or do something foolish, is not a reason not to do anything, particularly because what is being done now is much worse.

László Lovász, Budapest, Hungary. So first of all thank you, John, for being out there instead of me. The second remark is that I am a bit envious of Prof Arnold that he lives in a country where it's still the librarians who decide which journal to subscribe to; in many countries it is by bulk subscription by some government agency for all universities in that country, especially for the electronic versions. This is a situation which is a separate question but I just wanted to mention that this is also a very serious concern as far as I can see. My next remark is that I like very much Malcolm's remarks, essentially that the peer review system and numerical data should complement each other. In case there is a discrepancy then it should probably be more carefully looked at. We all know examples where the numerical data gives an entirely false impression, but I have also seen the peer review system run amock, with somebody who was by personality not so well liked or had one enemy in the system, and it

has produced very very strange results. So I think in that case numerical data should have corrected the procedure at some point. So I think the question to look at is which numerical data and how can we use it? Now I am talking about evaluating people not about evaluating journals, these are two different issues.

R. C. Cowsik, Mumbai, India. In India we have journals which publish only to the writers of papers in that journal – no other copies are ever sold. And we also have departments where everybody works in the same subject, a narrow part of mathematics. They quote each other so the citations would be large for them. We have a journal called Annals of Mathematics, India, and India is in small print!

Daya-Nand Verma, formerly at TIFR, Mumbai, India. My question to the entire panel is, isn't there some sort of a parallel between the life of research papers and life of individuals? Educationalists know that all children are not equal, in the same way as you have been pointing out that all research papers are not equal. So sometimes some research paper goes unnoticed, or maybe with very, very few exceptional references by a few people, and has not been referred to for 40 years, 100 years perhaps. Is there a way of devising a system which can pick up these exceptional, high calibre youngsters, so by that I mean the exceptional papers which go unnoticed, just as many high calibre children go not only unnoticed but get punished by the system.

Malcolm MacCallum. As mathematicians we like to have absolute objective truth. One area where there will not be an objective truth is in assessment of papers. It is a human activity and we're inevitably going to make mistakes. I don't think we can do anything but accept that and try to minimise its extent.

Doug Arnold. I would add that I certainly agree with what Malcolm just said. The most we can do is try to be careful when it come to assessing and the way you assess a paper is to read it. Counting the citations, no matter how carefully you count them, is not very helpful. You brought up the very good point that great papers in mathematics often go uncited for a long period. One of the wonderful facts about mathematics is you often see papers that are very highly cited many years after they are written. And another point is that citations come from all sorts of reasons. If a paper has a mistake and there are criticisms and retractions published, those cite the paper and boost its quality according to a foolish, citation-counting viewpoint.

Malcolm MacCallum. In fact I would say if you really want to be highly cited quickly the best way to do it is to write a paper that is just subtly wrong, so that lots of people pitch in to tell you why.

Garth Dales, Leeds, U.K. I would like to ask about possible political action, perhaps particularly addressed to Prof Arnold. I share your doubt about citation indices and I entirely agree that they are seriously flawed, but I see a lot of use in them, and it seems that the IMU and mathematicians don't like

this and they are inclined to try to protest against this or do something. But I regret to say that political realities are that mathematicians are a small group in the overall scheme of things, and my experience is that however cogent and powerful our arguments are that impress us, they have very limited impact on our government and agencies and so on. And I wonder what your assessment is. It seems to be that the only possibility of changing the culture in this particular respect is to find allies in the much bigger subjects of engineering, biology, physics and chemistry. Unless we have allies and friends in these subject areas, we'll have no impact whatsoever on the governments and agencies, or in particular private publishers that make money out of publishing these statistics. So what is your assessment of our chances of finding allies among these subject areas?

Doug Arnold. Well I think that's a very good point and one that has to be raised and thought about quite a lot. I'll make a couple of comments. First of all my comments are limited to impact factor as a journal quality proxy. I am not taking on the bigger question of an individual or departments. If we limit ourselves to pointing out, as many have pointed out, and many will continue to point out, that impact factor is highly flawed, we will go unheard. That has already been done and is basically a proven proposition. It is not only mathematicians who are complaining about this. Many, many groups are complaining about it. I feel that – because we are a fairly small community with a great devotion to our literature and some coherence – that by providing an alternative we have a realistic chance to say: 'Well you know there is an alternative that you can use instead. It is much, much better but just as easy to use. It has the imprimatur of the major math organisations in the world and there is all this evidence that it is better.'

This won't be used for comparing mathematics journals to say geophysics journals, which is meaningless, but for the purposes where you need to make an evaluation and judgement on journals of mathematics. I think this has a chance to come about. I think there is a possibility that people will say 'you know these mathematicians have some integrity and they really are doing this right, and maybe we should see about doing something like this.' As far as building up allies, recently I travelled to Singapore, to the World Conference on Research Integrity. There were 350 delegates including people from ministries of science and so forth. Out of the 350 delegates only I was a mathematician. I spoke a little bit about this proposal and I saw lots of allies and got lots of support. People are actually looking forward to seeing what we are going to be able to do in this area.

José Antonio de la Peňa. Well I think it's important that mathematicians take a position with respect to the indices, and maybe propose new ways to measure the impact of journals. But even what is done now, which is very bad, very flawed for mathematics, like measuring the impact factor of journal using this two years window which is completely nonsignificant for mathematics, could be

changed. For example, why not calculate the impact factors not using the two years window but using the full history of the journal? Just simply that. That can be much more significant for all sciences: why is this not done? I had an opportunity to speak with some high-ranking person from Thomson Reuters and the answer was 'of course we calculate this, we don't publish these results but we do calculate them'. So this means there is a completely different agenda, there's a hidden agenda why they calculate the indices in this way: maybe it is an economic agenda.

Chandan Dalawat, Harish-Chandra Research Institute, Allahabad, India. I just want to know if this new measure or classification on the quality of journal that's been proposed, has it actually been tested and could we look at the results that it gives?

Doug Arnold. No. The situation is the following. First of all, I am the President of SIAM which publishes these journals, so it is not my place to personally set down the mechanics of rating the journals. The proposal, which is brand new, just passed by the IMU General Assembly, is to establish a committee to try to design the best possible system, and then consider the question of how difficult it will be to implement. I can give you just a rough idea of at least what I have in mind, although other people may well change this. This is something akin to the program committee and panels that chose the invited speakers of this congress. That is many people, between 100 or 200, that were carefully chosen to cover many areas of mathematics. There will be a fairly small number of rating tiers, a few tiers or, perhaps, a matrix with separate tiers for journals that are tightly concentrated on one subdiscipline and broad journals, and so forth. Then these experts would review the journals and try to determine where they place them. Maybe there would be a time for public comment. There would be some rule against conflict of interest. Once they present the results, we will get the opportunity to test them. They will need to be renewed every 4 years or something like that. That's what I have in mind.

John Ball. To amplify that a bit, the committee would consider what would be the best way to create such a ranking system, then decide whether to implement that system, and in particular consider some of the issues surrounding such a system, maybe legal implictions, whether there would be the involvement from the community to sustain such a system, and what the knock on affect of such a system would be.

Zhiming Ma, China. Several years ago in China this problem was really very serious. For example in China if you apply for a promotion or for a prize you have to submit a document with citations. You maybe have to pay money to an agency or a library and then the agency (library) will type the citations, and then you submit it. This was several years ago; now the situation is getting better because many people complained about this. In China we mathematicians say that maybe people in other disciplines such as biologists will use this but

for mathematics it's not the case. We always ask the agencies or government to distinguish between subjects, so in this way we get some improvement. Now in China (at least in CAS) when mathematicians apply for a promotion or a prize, we will not follow the general rule of metrics. In this sense we are improving.

Martin Grötschel, Berlin, Germany. Somebody said before that we have no influence. This is absolutely not true; I think mathematicians are heard. Here is an example. The 2002 IMU General Assembly endorsed a document about best practices of journal publishing, advice to authors and so on, and open access in particular. This document was taken up in 2003 by the Max-Planck-Gesellschaft in Germany, Germany's top research organisation. MPG and other institutions finally formulated what was then called the 'Berlin Declaration' on open access. IMU's influence was clearly visible in this activity. Hundreds of research organisations worldwide signed this declaration, and mathematicians were the forerunners of this effort.

One can come up with many ways of classifying journals. Of course, targets have to be formulated together with reasons why we want to classify, why we want to sort journals, or people, or departments by quality. Even if we have reasonable arguments for the organization of the system of our journals, we must not only provide information about scientific quality but also about the way authors are handled, the turnover times and all the things that are important for journal publishing. Making available a broad spectrum of relevant information may be an alternative to just addressing the current crude measurements.

The panel addressed totally different targets, for example, whether we rank a paper, a journal, a department, or an individual, or how we compare mathematics to other sciences. We can't handle all these issues in the same way. I personally think that we mathematicians have to simply declare how we would like us and our work be judged; we then have to discuss the evaluation system with our peers in science and in administration. After that we can negotiate with them the way we are in fact judged. Most of the ideas presented here today are good, and our task is to find a reasonable combination of these measurements. My main field is optimisation and what we see in front of us is a multi-objective optimisation problem. There is something like a Pareto set that we have to target for, and which point on the Pareto curve is chosen will depend on local circumstances. We should simply be aware of this fact and spell it out.

Something I was really puzzling about is one of Frank Pacard's arguments. Everyone is happy about being free to make decisions. Now the French government seems to give financial support to the universities and the freedom to distribute it. I think that everywhere in the world you would be happy to have such a situation: you just have to elect a good president and good deans. They ought to have good insight and will determine who is doing good research. Do you really want the bureaucracy to give rules? I think it is better to have good people with good judgement distributing the money.

Frank Pacard. I agree with you, but in France we are passing from a system in which everything was decided at the top to a system in which a lot is decided at a local level. This takes time. Assessment of research is not an easy thing to do at the level of a university. Also, I think that the importance of the use of metrics really depends on how the money supporting research is distributed and this differs from one country to the other. In France, for example, one of the problems we are already confronted with in mathematics is that departments now have to fight against each other inside each university, to get research funds. And, so far, universities have no real way to decide how much support they should give to a given department. Beside the question of research support, there is also the problem of the evaluation of individuals. French universities now have to compare mathematicians with biologists, chemists or lawyers and panels performing these evaluations do not necessarily have mathematicians, biologists or lawyers on them. In this case, as you can imagine, metrics turn out to have a great impact on discussions. One can hope that the system will probably evolve towards a better equilibrium between the use of metrics and peer review, but in French universities I'm not so sure that the system has already reached this equilibrium.

Cheryl Praeger, University of Western Australia. I thought I would say a little bit about the Australian experience. The mathematical scientists in Australia did not choose, that is, did not set out, to rank journals. It was the Australian government that decided that all journals would be ranked. The government dictated the proportion of A*, A, B and C journals. So the mathematical scientists decided that we would prefer to make the ranking rather than have the government do it for us. We ended up having to do it three times; in our first run through we decided to rank as many journals as we could, so we would have more A* and A journals, since we had a fixed proportion available for them. The government did not accept this and we were given a limit on the number of journals we were allowed to rank. Even our second attempt was not accepted and we had to make a third attempt.

We are not terribly happy with it but it is something which has had the support under pressure of the whole Mathematical Sciences community, the pure, applied, the statisticians. Everyone joined together to try and do as good a job as we could. It has not been used yet but it is going to be used in a research assessment exercise, which is happening in the next year. We fear it will be used for other purposes. Already it is being used in an unfortunate way; for example my university proposes to measure research activity of individual staff members by the number of journal papers they publish in A* and A rated journals only, which comprise the top 20% or 30% of the journals according to an imperfect ranking. All other publications will be ignored.

John Ball. Am I not correct in saying that there is also a ranking of conferences, because I saw a listing of this on the Australian website (see `http://www.`

arc.gov.au/era/era_journal_list.htm). So I wondered whether you weren't allowed to go to a conference unless it was an A rated conference.

Malcolm MacCallum. That would have particular relevance in Computer Science where a lot of the best papers come out in refereed conference proceedings.

Hamidou Toure, Burkina Faso. We are a small community of mathematicians in Africa and the administrations are trying to use these different indices. Since the evaluation of publications in journals is done normally by peer review, it will be good that the International Mathematical Union make a peer evaluation of the ranking of different journals. It will be very useful for us.

Jean Lubuma, University of Pretoria, South Africa. I would like to say something about the system which we have in South Africa. I think the colleague from Australia (Cheryl Praeger) said something which is a bit similar. The system in South Africa is such that when you publish a paper, the South African Ministry of Education allocates directly an amount of about 20,000 dollars, which is paid to the university where the research work was done for papers published in the so-called accredited journals. For the moment those are the journals which are in the ISI list. We as mathematicians in the South African Mathematical Societies fought to show the government that this ISI list is not a system which is effective and which is definitely not in favour of mathematicians. The government said 'look, we want a simple method for us to decide' and so far the method which has been suggested came mostly from our colleagues from medicine and biology, because probably that is where all these ideas of the ISI lists originated. So this is the situation which we have at present in South Africa, and unfortunately we tried to fight but it didn't work. So I don't agree with what was said by the Secretary earlier, that mathematicians are powerful. I think I would rather agree with our colleague from England, that we are a very small group and it is not always easy to try and convince our colleagues from biology etc. who publish almost every day.

Jorge Soto-Andrade, University of Chile. I would like to point out that in our country we have some Chilean analogue of NSF and mathematicians have had some word to say concerning assessment of research, but most of the funding for research comes from the government, not directly through the universities. To some extent we have been able to make the point that mathematics is specific, compared with other domains like biology or economics and so on. One of the points is that journals which count for funding for reports are those which you find in this list of ISI or Thomson Reuters. Many people in the government agencies had the idea that ISI was something like IAS or some scientific institute. They didn't realize that it was just a private enterprise with commercial criteria, like Microsoft, Thomson being analogous to Bill Gates.

I would say that the International Mathematical Union is a rather small community but is quite homogeneous and has taken stance in a very significant way in the past, and if we can cite a report or some work of the IMU concerning

these points this will strengthen our position. I would like to recall the report by Figa Talamanca [7] who was very keen from the systemic viewpoint concerning ISI and perhaps some sort of update of this report would be very helpful. It pointed to the fact that the systemic role of ISI in science in the world was a very interesting subject in sociology, and there is a complex dynamic interaction of ISI with big American libraries, with publishers and so on.

One interesting point also is that in our country, which is somewhat free-market oriented, the government had the idea to give rewards to papers and so if your paper is in ISI, you'll get perhaps 1000 dollars, and if it's not there you'll perhaps get just a symbolic reward. One important thing I think is that IMU may have some alternative to ISI. If one looks a little bit, one finds very impressive examples of flaws in ISI reports and listings. For instance you have a list of highly cited mathematicians, highly cited researchers in ISI, and I realized perhaps one or two years ago that no Fields Medallist is a highly cited mathematician.

There is a field in which the situation is even worse than mathematics, which is mathematics education and there perhaps the best journals are not listed in ISI, and there was some reaction which was very positive from IMU and I think this should be pursued. Concerning other scientific communities, I also work with biochemists, biologists and other researchers in cognitive science, fields whose dynamics are quite different from ours, where updated reports from the IMU concerning this issue may help us a lot.

Michel Hébert, Cairo, Egypt. A few years ago the American Mathematical Society has started publishing their own impact factor in MathSciNet. I think it was a result of their own long study. I didn't read the report in detail at that time but remember it was precisely to respond to all these wrong ways for mathematics such as the two year window. So I'm a bit surprised also that there doesn't seem to be any collaboration, or there has been no result. Don't the IMU and AMS know what each other is doing?

Ali Ulas Ozgur Kisisel, Middle East Technical University, Turkey. So in my university our struggle is usually with the university administration, which rarely consists of any mathematicians; however in the mathematics department we have quite a good idea about what should be brought up, and what should be kept down. Maybe I should give some specifics. For instance all the hiring procedures and appointments to posts are based on the number of papers in science core index journals and for instance in order to be an Associate Professor in the Maths department it should be at least 7, and that is a fairly low number, and as you could expect it has drastically different effects if you are studying Applied Mathematics or Modular Forms. So I asked some friend who was working in Ex-Soviet Union how did it happen there? And it was very easy – Kolmogorov decided everything, so no problem! But of course in today's world I guess this is out of the question. But something that we could use, and IMU or global organisations could do, would be to bring forward some experiences from

prestigious universities, like interviews with deans, interviews with department chairs so we could use them in our struggle with our administrations.

Gholamreza Khosroshahi, IPM and University of Tehran, Iran. I work in an institute called the Institute for Research in Fundamental Sciences. In the beginning we were just theoretical physicists and mathematicians and the fight from the beginning started in the committees and the councils about evaluation of mathematics and physics. Physicists usually dominated the issue because of citations and these kind of things and later on other schools like computer science, theoretical computer science, neuroscience, nanoscience were added to our institute. The fight was widened and there are two problems which are always there, one – inside the mathematics council you have to fight – suppose I'm a combinatorialist, at the beginning those who didn't do any research about 20 years ago said what? Combinatorics? And they were saying that it is easy to publish in combinatorics and it is very difficult to publish say in algebraic geometry etc. So this fight gradually subsided because gradually they had to publish and they couldn't publish. Then outside of mathematics, physicists used to say always 'what is the citation on this'? 'This paper has 100 citations' and so forth? This fight still is going on, but I agree with Prof Grötschel that mathematicians should be tough fighters and they should handle these hard situations. We do that and we have succeeded.

One more thing is that we have to prove to others that every discipline has its own culture: culture in mathematics is quite different from culture in computer science or physics.

Gerhard Paseman, USA. There are a number of communities online (such as mathoverflow.net) that are doing rankings of various things, anything from individuals to pizzas. In particular there are some communities forming, scientific communities that exchange information and they do ranking based on reputation, and it seems to me that they are models of some of the things to look at, as examples of what might be a good form of metric, and there are also some obvious mistakes in some of these models, that could probably be avoided by forming a metric. I'm curious to see how metrics for journals, for professional mathematicians, for scientists will actually reflect some of their activity online.

References

[1] *Citation Statistics*, IMU/ICIAM/IMS report, `http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf`, 2008.

[2] D. Arnold and K. Fowler, *Nefarious Numbers*, `http://arxiv.org/pdf/1010.0278.pdf`, 2010.

[3] R. Kenna and B. Berche, *The extensive nature of group quality*, EPL (Europhysics Letters), June 2010, 58002.

[4] A. Fabian, *The impact of astronomy*, Astronomy & Geophysics 51 (2010) 3.253.30.

[5] A. Blustin, *Publication and citation statistics for UK astronomers*, Astronomy & Geophysics, 48(2007)6.326.35.

[6] J. Zuntz, T. G. Zlosnik, C. Zunckel, J. T. L. Zwart, *Orthographic Correlations in Astrophysics*, arXiv:1003.6064v1.

[7] A. Figa-Talamanca *Strengths and weaknesses of citation indices and impact factors*, in *Quality Assessment for Higher Education*, Alessandro Cavalli ed., Chapter 8, Portland Press, 2007 `https://www.portlandpress.com/pp/books/online/QAHEE/`.