Evaluation of Scholarly Work

Can Scientific Productivity be Measured with a Single Number?

Prof. Haldun M. Özaktaş

Principal Member, TÜBA

haldun@ee.bilkent.edu.tr

Can scientific productivity be measured with a single number? In my opinion, no.

Neither am I of the opinion that it can be properly measured with two or three numbers. A person's intelligence or abilities, the beauty of a work of art or its effect on people or the overall impact of an artist, the emotions of a person in different situations are very difficult, if not impossible, to measure. These are not things we would consider measuring in the first place. Measuring a researcher's contribution to science is likewise difficult. Let alone these rich and difficult to define concepts, even measuring a relatively concrete fact such as the economic output of a country is laden with major difficulties.

At best, these qualities can be measured with a multi-dimensional scale. Such scales have been introduced for the measurement of intelligence. However, multi-dimensional scales do not allow for each and every kind of comparison, since we usually do now know how much weight we should assign to the different dimensions. Reducing multiple dimensions to one by combining different dimensions of ignoring some of them has always been attractive since it allows direct comparison and ranking. Such comparisons has considerable utility: when choosing among two students or scientists for admission or hiring, when granting research funds or awards, the ability to compare individuals significantly facilitates the process.

Beyond this utilitarian aspect, scientists, like all people, compare themselves to others and wish to see how they stand with respect to others. Thus, the unbearable attractiveness of ranking individuals, despite all its weaknesses and faults, contributes to the tendency to measure scientific performance by reducing it to a single number.

This phenomenon is an example of what is known as the "fallacy of the instrument". We are confronted with an entity which, due to its very nature, cannot be reduced to a single number. By nevertheless insisting on trying to measure it with a single number, what we do in effect is to ignore those parts of it which we cannot measure.

Unavoidably, our measure will involve a degree of arbitrariness, as it will depend on what we choose to ignore or deemphasize. This is not much different than Nasreddin Hoca's (a Turkish folk hero) searching for a lost item not in the dark place which he actually lost it, but in another place since it is better illuminated. We do not really measure the thing we set out to measure, but pretend that we are doing so. We sometimes commit an even greater blunder: we start confusing the thing we are trying to measure, with the thing we actually measure.

What we understand from the terms intelligence or scientific productivity or the economic output of a country leave their place to what we are able to measure. These rich concepts become reduced to instrumentally (operationally) defined concepts.

This general critique is applicable not only to measures such as the number of publications or citations of a scientist, but also to the recently proposed h-index. In order for a scientist's h-index to be h, this person must have at least h publications each of which has received h citations.

The total number of publications of a scientist is indeed not a very meaningful indicator, since he or she may have authored many insignificant publications. The h-index is meaningful from this perspective, since publications with few citations are ignored. However, upon closer inspection, it is difficult to find sufficient justification to claim that this index is more meaningful than many other alternative indices, which may be similarly defined. The citation histogram of a scientist reveals important information. This histogram is obtained by ranking all of the publications of a researcher according to the number of citations they receive. An expert examining this histogram can make a reasoned judgement regarding that scientist by looking at the content of the highest cited publications, the number and order of their authors, where the publications have been made, and the distribution of citations over the publications.

It is possible to derive various measures from this histogram. For instance, the total number of citations is the area under the histogram. The sum of the n'th power of the citations is another measure; while such a measure is not used for scientific evaluations; it is a widely used measure of distributions in many areas and may be used for this purpose as well. Beyond these, the many measures of width and mass used to characterize distributions employed in different fields can be applied to this histogram to obtain many different new measures.

Every consumer, when making an expensive investment such as a house, vehicle, or appliance is confronted with similar difficulties. The market contains many products with many different features, making it difficult to compare them and make the right decision.

Some magazines or web sites rank these products to make the consumers' job easier. But one should not forget that the ranking produced highly depends on how much importance has been assigned to the different features, and is, therefore, highly subjective and might even have been influenced by commercial interests.

Depending on how different features have been weighted, very different ranking may emerge. Which one of these rankings is more meaningful is a question that can only be answered through empirical studies, but regarding entities such as scientific productivity and intelligence, even such studies are unlikely to provide us a solid ground for comparison.

To illustrate the arbitrariness of the index in question, let us slightly generalize its definition: Let the h-index with parameter k, denoted h(k), be defined such that a scientist with h(k) has at least h publications each of which has received hk citations (k may be either less than or greater than unity). The h-index already defined is a special case of this index with k=1. In my opinion, there is nothing special or prior about the special case k=1; not only may different k values be more appropriate for different fields, but different k values may result in significantly differen rankings. Under these circumstances, the proposed index would considerably lose its significance. To summarize: If small variations in the definition of an index can change the resulting rankings considerably, the information provided by that index is of limited value. Even if it is actually the case that a significant portion of the information inherent in the citation histogram can be captured meaningfully with a single number, this must be supported with empirical evidence. For instance, despite the difficulties in data collection and evaluation, assume that a ranking of scientists in a given field was based on extensive information and subjective evaluations of a large number of scientific peers.

If the h-index or another easily computed index produced similar results as such a study, did then we could conclude that the use of such a measure is meaningful. But as we have noted before, demonstrating the utility of an index with such a study is not an easy task.

In conclusion, while objective measures seem to make life easier in many ways, if they are leaving out a significant part of what was originally intended to measure, they will have introduced an unacceptable source of systematic error. These objective systematic errors may often be of greater magnitude than the errors introduced by subjective evaluations which we may have been trying to avoid in order to use objective measures.
A healthier attitude will be to accept that it is intrinsically difficult to measure some things in nature, and not to insist on measuring and ranking these things, and furthermore, to be critical of systems and institutions which push us to insist on such measurements and ranking in the first place.

- Hirch, JE. "An index to quantify an individual's scientific research output" http://xxx.arxiv.org/abs/physics/0508025
- Özaktaş, HM. (2002-2003) Ölçülebilen ölçülemeyeni kovar: Ekonomik düşüncenin fakirleşmesi üzerine bir deneme. ("The measurable drives out the unmeasurable: An essay on the impoverishment of economic thought") *Toplum ve Bilim*, no. 95:151-158,216.

Academic Performance Evaluations and Number Games

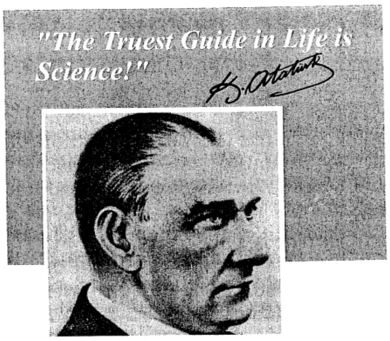
Prof. Ersin Yurtsever
Principal Member, TÜBA
eyurtsev@ku.edu.tr

In Turkey, publishing scientific articles in international journals has a recent history going to the early 1980's. Even though, there were publications in international journals before those years, their numbers were very small. For example, the first article from Turkey in the prominent physics journal (of the American Physical Society) Physical Review belongs to R.B.Kennard who was then teaching at Robert College. Later we see articles from B.Kurşunluoğlu, F.Gürsey, O.Sinanoğlu and T.Erben in that journal. Still, the total number of these articles is only 13 until 1980. In the year 2005, the number of articles from Turkey in the same journal is 53. With the help of the relatively recent developments such as monetary awards to scientific articles and the promotion criteria based on number of articles, the number of papers coming from Turkish institutions has increased almost to 15000 a year. Naturally, as the numbers increase, we are met with the dilemma of quantity vs. quality. Even daily newspapers have started printing comments on this quality issue.

It is an extremely important task to analyze the quality of scientific publications; however, there is no consensus on what quality criteria can be used for such purposes. Famous "publish or perish" concept was born in the US universities first and now its effects are being felt in Turkish institutions albeit in a much weaker sense. In a parallel fashion, the "number games" associated with the above concept have also reached Turkey. The most commonly known number games are a) assigning weights to different academic achievements to make "objective" decisions and b) rank countries, institutions and moreover scientists. Both of these approaches aim to develop black-box algorithms to perform academic evaluations in an effortless manner.

It is important to see the dangers of reducing the quality issue to numbers. In all these number games, there are two essential factors. First of these is based on methods to increase the numbers involved in academic evaluations such as the number of papers, citations, invited talks, and awards. This can be achieved by either producing high quality work or flexing the ethics of research. One of the most commonly used strategies has been named as LPU by E. Garfield, the founder of ISI. LPU stands for the least-publishable-unit and corresponds to the minimum amount of information sufficient to publish an article. Dividing the research results into small articles and then publishing them in different journals is unfortunately not seen as unethical by a large number of people. If the aim of publishing results is to inform the other scientists, then it is much more efficient to put everything into a few articles rather than distributing to many. There are other similar tricks in number games such as introducing other researchers as authors even though they may have little or no contributions. At the extreme, we see highly unethical cases of made-up data and results.

Turkish Academy of Sciences December 2006 No: 4



diary

Owner
Engin Bermek
President of the Turkish Academy of Science

Editor
Metin Heper

Advisory Board
Cengiz M. Dökmeci
Yücel Kanpolat
Bernard Lewis
Saim Özkar
Dinçer Ülkü

Editor (Sorumlu Yazı İşleri Müdürü)
M. Çetin Gülovalı

Managing Editor/ Coordinator
Filiz Çiçek Bil

Presidency&Secreteriat of TÜBA
Piyade Sokak No: 27, 06550
Çankaya- ANKARA
Tel: +90.312.442 29 03 (pbx)
Fax: +90.312.442 23 58
http//www.tuba.gov.tr

Diary (Günce)
Piyade Sokak No: 27, 06550
Çankaya- ANKARA
Tel: +90.312.442 29 03 (pbx)
Fax: +90.312.442 64 91
e-mail: tubagun@tuba.gov.tr
ISSN: 1302-9541

Printer
Yeni Reform Matbaacılık Ltd.Şti
K.Karabekir Cad. No:91/2 İskitler/ Ankara
Tel: +90.312.341 20 92
Fax: +90.312.341 20 93
Diary 4 published 2.500 number.

Global Warming, Physics, Science and Economics Education

During the last half-a-century, the mankind experienced a global warming, which was caused by increased human activity and the consequent need for more energy. In the first article of this issue of Diary, **Professor Sadık Kakaç** takes up the alarming consequences of this phenomenon and the measures that may be adopted for grappling with this threat. Since its formulations several decadas ago, numerous experiments were made to test the various aspects of Einstein's Theory of Gravitation. In our second article, **Professor Metin Gürses** shares his views with us on probable new experiments concerning that theory. In our next article, **Professor Ayşe Erzan** raises the crucial question of how to render Physics more attractive to the public and attract new students, particularly women, to that field of study.

How best to evaluate the performance of scientists continues to be a controversial matter among the scholars and funding agencies. In this issue of *Diary,* **Professor Ekmel Özbay** explains the so called h-index, which is a recently developed instrument for rank scientists among themselves; **Professor Haldun Özaktaş** points to the limitations of that index; while **Professor Ersin Yurtsever** argues that in this regard peer review is a more reliable measure as compared to its quantitative counterparts.

In recent decades, inquiry-based science education, as against rote learning, has become popular in several countries. We continue this issue with **Professor Yücel Kanpolat**'s article, in which he explains this new style of education and familiarizes us with its various patterns and adaptaions in different countries.

In August 2005, **Dr. Ercan Kumcu**, set aside his six columns in Istanbul daily *Hürriyet* and made a critical evaluation of the way Economics is taught at Turkish universities. Here we reproduce those articles and three artaicles in which **Professor Şevket Pamuk, Professor Fikret Şenses,** and **Professor Ercan Uygur** discuss the points made by Dr. Kumcu.

In the final article of this issue of Diary, **Professor Fikret Şenses** provides us with a colourful account of the dynamics of our Academy's foresight study concerning Economics.

With our best wishes.

Editor

In this issue