New Index for Quantifying an Individual’s Scientific Research Output

Mahmoud Abdel-Aty1,2,3
1. Mathematics Department, Faculty of Science, Sohag University, Sohag 82524, Egypt
2. Scientific Publishing Center, University of Bahrain, 32038 Kingdom of Bahrain
3. University of Science and Technology, Zewail City, Sheikh Zayed District, Giza 12588, Egypt

Abstract: Classifying researchers’ work according to the quality of their publications rather than the quantity of their publications is an important issue. To this end we introduce a new measure, the “percentage range” or $A$-index, which provides a qualitative evaluation of a researchers’ productivity. The percentage range depends to a great extent on the number of single-author published papers and their citations. It is to be a new index to be considered along with the $h$-index. The combined factors have the advantage of making clearer the innovation of the individual authors. The resultant percentage range gives a reduced impact on its numerical value for authors who gain citations by adding their names on multi-author papers. It is shown that various dimensions of ethical integrity and originality are clarified by the new index. The important scenarios arising from this analysis are demonstrated with examples. The great differences between the new percentage range and old $h$-index come from the percentage range’s emphasis on considering the whole work of an author, including the significance of the author’s single-author papers as opposed to multiple-author contributions. This emphasis is demonstrated.

Key words: $h$-index, impact factor, research output.

1. Introduction

We would like to start this paper with the following question: Are we judging researchers’ productivity by correct standards? Although qualitative measures such as the $h$-index are used to judge productivity and have increasingly grown in institutional use, confusion often exists [1-8]. How is it that for different types of contributions authors receive the same $h$-index? How can the quality of the publication record be assessed when authors of multi-author papers are compared to authors of single-author papers [3]. These topics are addressed in this paper.

The most widely used and highly regarded quantification of research productivity was given in 2005 when Hirsch [9] proposed the $h$-index as an indicator of research productivity in different fields [10-13]. This is defined as the number of papers for which the number of citations is greater than or equal to $h$. The scientific output of a researcher is of index $h$ if $h$ of his/her $Np$ papers each have at least $h$ citations, and the other $(Np - h)$ papers each have no more than $h$ citations. This index has been used to characterize the scientific output of researchers in a very good way. In particular, the application of the $h$-index treats the total number of papers published over a certain period of years, considering the total number of citations for each paper. By its nature, however, it weighs single and multiple author contributions the same so that it does not address the issue we raise. In addition, with the $h$-index and all other indices derived from publication rates the question of what constitutes a journal publication must be addressed.
Before addressing the issue of single versus multiple authors in the development of a productivity index, the question of what constitutes an acceptable journal publication is a most important issue which factors into the determination of any author productivity index. This has been independently addressed in a number of ways by different organizations and is a well settled question [2]. As to what constitutes a journal publication: The ISI (Institute of Scientific Information) Web of Knowledge now indexes more than 11,000 science and social science journals [12] and uses the impact factor to report on ranking the journals. As another source, Elsevier, as an international publisher, has also established the independent and international Scopus [14] content selection and advisory board. This is meant by Elsevier to avoid any conflict of interest in the choice of the journals to be considered for the inclusion in the database and to maintain an open and transparent content coverage policy. Scopus covers nearly 20,500 titles from over 5,000 international publishers, of which 19,500 are peer-reviewed journals in the scientific, technical, medical, and social sciences [6, 7, 14].

These are the commonly used major sources for recognizing peer-reviewed journals, and are used in the $h$-index as well as other formulations which we now briefly summarize and the new formulation we propose below.

Another, less used, measure of productivity has been suggested by Egghe [5]. It is called the g-index. This again quantifies scientific productivity based on publication records and is calculated through the distribution of citations received by a researcher’s publications. Given a set of articles, ranked in decreasing order of their number of citations, the g-index is the (unique) largest number of citations that the top $g$ articles together receive. This accounts, then, for at least $g^2$ citations. The g-index is a number which for the same author gives, in general, a ranking close to the $h$-index. It provides an indicator which agrees well with the $h$-index.

2. Model

More recently, another index based on an application of a modified version of the $h$-index and journal impact factor was proposed [1]. This new version was originally applied to estimate the impact of journals published in the Arabic Language as well as to quantify the cumulative research contributions of Arabic scientists. Specifically, in order to achieve the best indicators for a journal or researcher, another model system was developed yielding an effective numerical index. Consider for this model the case of the productivity of an individual researcher. The modified index proposed is based on the Hirsch index, $h$, and is defined by the equation:

$$AsF = \left( \frac{h}{h_i + 1} \right) \times 100.$$  \hspace{1cm} (1)

From this equation it is seen that the index $AsF$ is an increasing number according to the increasing increment of the $h$-index. However, it does not exceed 100. In particular, to consider some examples, assume that the researcher has an $h$-index of 1, this corresponds to $AsF = 50\%$ while an $h$-index 5 corresponds to $AsF = 83.3\%$ and so on with this increasing trend with increasing $h$-index. A feature of the $AsF$, however, is that as the $h$-index increases the difference with respect to the $AsF$ gets smaller. This decrease in the incremental increase with increasing $h$ offsets some of the effects of multiple authored papers.

To develop an effective new productivity index, we are interested in researchers quality and productivity indices, taking into account the relevant magnitudes of the number of single-author papers, citations rates, and the $h$-index [9]. This is most conveniently accomplished in a mathematical formalism involving a single quantifying equation. In this regards, some related treatments by us based on both $h$-index and number of citations and discussing authors ranks have already been presented in Ref. [1]. What is present below is essentially an extension of these of our past
efforts to the most general form of a complete indexing equation.

One may ask the following question: Is there any difference between the $h$-indices of two authors, each publishing the same number of papers and receiving the same number of citations, but one on single-author papers and the other on multi-author papers having very limited input from him/her? The answer according to the present $h$-index and the related $g$-index and rates of citations is no. They both receive the same ratings. This is clearly a poor rating of the two authors’ research productivity, and we need to add some factors to the current indices to avoid such a misleading result.

It is debated in the literature whether the $h$-index’s assessment of qualitative output should parallel the results obtained from counting citations or be quite different from this count [10, 11]. In fact the quality of published papers comes from the work presented and the author’s benefit from this presentation should depend on his relative contribution. This requires a revision of the current factors used to quantify these relative contributions. We suggest a new algorithm which is clear and more precise. The algorithm reflects the impact of different variables which cannot be avoided when speaking about research quality and it generates a new quantifying index number. The numerical value of the index takes into account the total number of published papers, number of citations, and $h$-index for both single-author and multiple-author papers. Its value is defined as follows:

$$A = \sum_{i=1}^{6} \mu(\Omega_i)$$

where

$$\mu(1) = \frac{N_2}{N_{h,1} + 1} \times \Omega_{A,1}$$

$$\mu(2) = \Omega_2 \times \frac{\max(0, \frac{N_2 - 70}{50})}{1 + \max(0, \frac{N_2 - 70}{50})} \times \Omega_{A,2}$$

$$\mu(3) = \frac{N_{2,4}}{N_{h,4} + 1} \times \Omega_{A,4}$$

Here $N_1$ is the total number of published papers by the author, including the multi-author papers;

$N_2$ is the modified total number of the citations which have been received for all published papers by the author, including the multi-author papers;

$N_3$ is the $h$-index of the author for all his work;

$N_4$ is the number of single-author papers published in ISI journals by the author;

$N_5$ is the modified number of citations received for single-author papers only;

$N_6$ is the author’s $h$-index considering single-author papers only.

The remaining factors in Eq. (2) are weighs, and we suggest weighting values given by $\Omega_1 = 20$, $\Omega_2 = 10$, $\Omega_{A,4} = 14$, $\Omega_8 = 12$ and $\Omega_9 = 30$.

### 3. Results and Discussion

From now on we will call the above numerical index the “percentage range” or $A$-index. It offers the opportunity to allow more factors to play their necessary roles in determining the final index quantifying the individual’s scientific research output. Also, the weighting of each factor can be changed according to the purpose of the evaluation.

For example, we consider the following hypothetical for the application of the $A$-index: Two authors each have published 20 papers in ISI journals, 11 of them have been cited 11 times or more, but the authors have different single paper numbers and citations. The relevant statistics are summarized in Table 1.

According to Table 1, we see that the two authors have the same $h$-index. The first author did not publish any single-author papers, and the work of the second author was published mostly as well recognized...
New Index for Quantifying an Individual’s Scientific Research Output

Table 1  Comparison between two authors using $A$-index.

<table>
<thead>
<tr>
<th></th>
<th>$N_1$</th>
<th>$N_2$</th>
<th>$h$</th>
<th>$N_3$</th>
<th>$N_4$</th>
<th>$N_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>20</td>
<td>300</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$A_2$</td>
<td>20</td>
<td>300</td>
<td>11</td>
<td>15</td>
<td>300</td>
<td>8</td>
</tr>
</tbody>
</table>

where, $A_1$ refers to author 1 and $A_2$ refers to author 2; $N_1$ is the total number of the published papers; $N_2$ is the total number of the citations; $N_3$ is number of single-author papers; $N_4$ is the number of citations for single author-papers and $N_5$ is $h$-index of the single-author papers.

An independent (single-author papers) work, receiving 300 citations. Accordingly the second author has an $h$-index for single-author papers of 8 while the first author has $h$-index of zero. As we shall now see, these facts can be used to lead to a much more effective accounting of the real research productivity of the two authors. Specifically, according to our equation for the $A$-index, the index of the first author is 42.06 out of 100 while the second author obtains 97.05 out of 100. This means that the number of single-author papers and their citations play an important role in the $A$-index evaluation of the quality of the researchers’ outcome.

A recent study has shown that Web of Science, Scopus, and Google Scholar produce citation results for journal published articles which are quantitatively and qualitatively different [13]. The study noted that different citation data bases (i.e., including the Web of Science, Scopus, and Google Scholar) use methods unique to them to record and count citations. In addition, differences have been found between these data bases in the citation counts involving a number of different fields [8]. With this in mind, our $A$-index equation gives a reasonable, more stable, and acceptable index for all cases in the measure of research productivity as it is less sensitive to the exact citation and paper count.

To obtain a deeper understanding of the properties of the $A$-index equation it is useful to consider some examples of how the values of the $A$-index change under variation of some of the publication factors it considers. We will consider three different examples meant to show the effects of a change in a single variable when the other variables are kept constant.

Consider the hypothetical results for four different authors that are presented in Table 2. The same parameters of Table 1 are used here in Table 2. For the case presented in Table 2 the statistics of all four authors are the same except that the number of citations is different between the four authors. The resulting $A$-index for each of the four authors is presented in Table 3 and also plotted in Fig. 1.

The $A$-index shows an increase with the increase in the total number of citations but this is modified by the constant rate of citations for the single author papers. The $A$-index benefits the authors whose work is popularly cited. Now let us fix the total number of citations in Table 1 of hypotheticals presented to be fixed at 200 for the four different authors and change the total $h$-index values from 11 for each author to 3, 7, 12 and 16 for each author, respectively. This means that all of the authors are equally good at single author publications but they vary in the success of the totality of their works. The resulting $A$-index for each of the

<table>
<thead>
<tr>
<th>Authors</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
<th>$A_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_2$</td>
<td>200</td>
<td>400</td>
<td>800</td>
<td>3,000</td>
</tr>
<tr>
<td>$A$-index</td>
<td>73.131</td>
<td>75.797</td>
<td>77.575</td>
<td>79.152</td>
</tr>
</tbody>
</table>

Fig. 1  Plot of the $A$-index for different values of the number of total citations.
four authors are presented in Table 4 and again presented graphically in Fig. 2.

Again the authors are rewarded for their increased citations even though the single author papers receive equally effective research citations. Now let us fix the total number of citations Table 1 of hypotheticals presented to be fixed at 200 for the four different authors and, instead of changing the total author $h$-index, let us change the single author $h$-index from 3 for each author to 0, 1, 4, and 5 for the four different authors. Everything is the same but the single author publications vary in their recognition. The corresponding $A$-index for each author are given in Table 5 and presented graphically in Fig. 3.

The importance of the single author work is seen to be validated by the $A$-index. Taken as a whole, Figs. 2 and 3 show that any small change of the involved factors lead to a dramatically change of the $A$-index. The suggested equation is very sensitive to the single-author papers $h$-index factor, while giving a good mix of the other factors to be considered in the assessment.

More specifically, the above results show that small changes in the single-author papers $h$-index induce large changes in the $A$-index, while big changes in the number of citations give less change in the $A$-index. The $A$-index, then, gives quantitative and qualitative indications of the research output, giving ample credit for single-author papers. While, inevitably, diverse perspectives which use citations methods only and their differing views on how people rank different researchers should be studied and account taken for disagreement and controversy over how quality should be evaluated. We think that, despite this, it is important to develop the present formula and to use a common criterion allowing for the evaluation of the quality of published work.

One of the biggest challenges may be described as follows: How do we select the main author of the multi-author papers and assure the quality and trustworthiness of his/her contribution? The researchers’

Table 4 Comparison between four authors according to their $h$-index only using $A$-index.

<table>
<thead>
<tr>
<th>Authors</th>
<th>$A1$</th>
<th>$A2$</th>
<th>$A3$</th>
<th>$A4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$-index</td>
<td>3</td>
<td>7</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>$A$-index</td>
<td>70.797</td>
<td>72.547</td>
<td>73.220</td>
<td>73.474</td>
</tr>
</tbody>
</table>

Fig. 2 Plot of $A$-index for different values of the total $h$-index as 3, 7, 12 and 16.

Table 5 Comparison between four authors according to their single autho’s $h$-index only using $A$-index.

<table>
<thead>
<tr>
<th>Authors</th>
<th>$A1$</th>
<th>$A2$</th>
<th>$A3$</th>
<th>$A4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N5$</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$A$-index</td>
<td>50.631</td>
<td>65.631</td>
<td>74.631</td>
<td>75.631</td>
</tr>
</tbody>
</table>

Fig. 3 plot of $A$-index for different values of the single-author papers $h$-index as 0, 1, 4 and 5.

index needs to reflect this quality, avoiding the above discussed points of criticisms, in order to avoid misleading results.

One may think of solving the above mentioned problem by adding the following modification factors [15]:

\[
N'_1 = \sum_{i=1}^{N} \left \{ \begin{array}{ll} \frac{N_i}{n_i} & \text{if } \frac{N_i}{n_i} \leq 4, \\ 0 & \text{otherwise} \end{array} \right \}, \quad (3)
\]

\[
N'_2 = \frac{N'_2}{n_i}, \quad (4)
\]

where $N'_i$ is the modified total number of published papers by the author including the multi-author papers;
and \( n_i \) is the total number of authors on the \( i \)th paper. Also, \( N_2 \) is the modified total number of the citations which have been received for all published papers by the author, including the multi-author papers. For authors of a large number of single-author papers the \( A \)-index is to a large extent not affected by the above choice, but authors with a large number of multi-author papers will not gain points from papers where the number of co-authors exceeds 4. It is likely that Eqs. (3) and (4) can also be of use in the implementation of other general conditions or measurements.

Finally, it is interesting to conclude with the statement [16]: “the leaders of the scientific enterprise must accept full responsibility for thoughtfully analyzing the scientific contributions of other researchers. To do so in a meaningful way requires the actual reading of a small selected set of each researcher’s publications, a task that must not be passed by default to journal editors.”

4. Conclusion

We have studied measures of the efficiency of an individual’s scientific research output. A measure, the \( A \)-index, has been proposed which quantifies the research productivity as a function of the single-author’s papers-impact and multi-author’s papers citations and the corresponding \( h \)-index. It also addresses the effect of collective citations and various other factors. When any of the variables related to productivity are found to increase, the suggested index is shown to exhibit a relative rise in value. Using a simple analysis, we have illustrated the effects of a rise in various variables on the index for a number of different cases. In particular, we have noticed an extremely fast convergence of the index for authors who have published most of their work independently. Consequently, the \( A \)-index may provide a useful tool for quantitatively characterizing the difference between authors who are working jointly and authors who are working independently. In any case, we point out that obtaining a deeper understanding of the behavior of an individual’s scientific research output requires more discussion and examination.

Acknowledgments

The author would like to thank Prof. A. S. Obada, Al-Azhar University, Egypt, Prof. A. McGurn, Western Michigan University, USA, Prof. H. Eluech, Université de Montréal, Canada, and Prof. Noor, COMSATS Institute of Information Technology, Pakistan for thoughtful discussion of this problem and helpful comments.

References


