

## Improving Fairness of H-index: RA-index

Adian Fatchur Rochim\*, Abdul Muis, and Riri Fitri Sari

Universitas Indonesia, Depok - 16424, Indonesia

\*E-mail: adian.fatchur@ui.ac.id

### ABSTRACT

H-index has been widely used as one of the bibliometric measurement methods for researchers' performance. On the other hand, H-index has been unfair for figuring authors that have high number of citations but fewer number of papers (perfectionist researcher) and researchers that have many papers but fewer citations (productive researcher). The main objective of this article is to improve H-index for accommodating and calculating perfectionist and productive researchers' impact based on Jain's Fairness Index algorithm and Lotka's Law. For improving H-index by RA-index is proposed. To prove the proposed method, 1,710 citation data sets of top cited researchers from Scopus based on author names list from Webometrics site are used. Fairness index of the RA-index has the average of 91 per cent, which is higher than the fairness of H-Index 80 per cent has been found.

**Keywords:** RA-index; H-index; Fairness-index; Jain fairness-index.

### 1. INTRODUCTION

Researcher's impact is a quantitative representation of citation and number of paper of a researcher. The most well-known indices for researchers' impact is called the H-Index. In 2005, Hirsch proposed the H-Index, an indicator to measure the productivity of researchers, combining two parameter, i.e. the number of papers and the number of citation<sup>1</sup>. Most indexing services such as Scopus, Clarivate Analysis, and Google Scholar use the H-index to represent a researcher profile.

On the other hand, from the literatures we found weaknesses in the implementation of the H-index is found. Bornmann<sup>1</sup>, stated that one of the weaknesses of the H-Index is the difficulty for new researchers to get a citation. Secondly, for senior researchers who have produced many publications, although they no longer produce papers, their citations continue to grow<sup>2</sup>. Mesiar<sup>3</sup>, stated that the H-Index was less sensitive to researchers who have a large number of publications but with fewer citations. Likewise, this is also the case with a researcher that has a few publications with a high number of citation was not accommodated by H-index calculation<sup>1,4,5</sup>. Two type of researchers who are not well accommodated in the H-Index method are the productive and perfectionist researcher types<sup>3,6,7</sup>. The excess number of citations and/or number of papers of those researchers do not contribute to increase the H-index value.

H-index has widely been used i.e. to measure performance of researchers, to compare scientific production, also to be implemented in organisational level<sup>2</sup>, although, which still have the weaknesses on the calculation. Figure 1 shows the

weakness of H-index such as, H-index is not sensitive to calculate the impact of productive researcher and perfectionist researchers. Moreover, H-index does not count those papers that have never been cited<sup>1,8</sup>.

To improve H-index, a scientrometrician named Egghe<sup>9</sup> proposed the G-index. G-index accommodated some papers with high citations' number. Bihui<sup>6</sup>, *et al.* stated that it was impossible to compare researchers that belong to different period of time and field of study and proposed the R-index to complement the H-index. Kosmulski, in 2007 proposed Maxprod for assessing the scientific output<sup>10</sup>. Gangan<sup>11</sup>, proposed impact-Citation-Energy (iCX) as an alternative bibliometric based on analogy in electricity energy. Abramo<sup>12</sup>, stated that the definition of the average impact cannot be claimed as a researcher's performance. He proposed some methods called Mean Normalised Citation Score (MNCS) and Fractional Scientific Strength (FSS). Rodriguez<sup>13</sup>, proposed a double rank for research assessment. For improving H-index, some proposals have been submitted, but the H-index is still used as the most popular indexing services. This condition occurred because the correlation of the H-index and its variants, etc. high, which is around 0.8-0.9<sup>14</sup>. In literature, H-index has some advantages, i.e. mathematically simple, it can be applied to any level of aggregation, it combines publications and citations, and its robust characteristic<sup>6</sup>.

Wang<sup>15</sup> described the area of the H-index as an h-core, and the papers that have few citations as an h-tail area, while Alonso<sup>13</sup>, described the papers that have high citations as upper-tail. Those descriptions are used for identifying areas excluded by H-index calculation. H-index value excludes the upper-h-tail and lower-h-tail in calculating the *h* value of a researcher. Egghe<sup>9</sup> improved the calculation method of

G-index by including the upper-h-tail to calculate the  $g$  value of a researcher.

In this paper, four parts of areas in citation function of researcher data set are as follows a) upper-h-tail, is an area that the papers have higher citations than h-index value, b) h-core, is an area of the number of  $h$  papers that have  $h$  citation, c) lower-h-tail, is an area that the papers have lower citations than h-index value, d) uncited-h-tail, is an area that the papers have no citations.

Proposed a new method called RA-index that combines all the tail-areas of the citation function, including uncited papers for improving the sensitivity of H-index. To prove RA-index calculation, 1,710 data set of citation from Scopus were used. Our discrimination-index tool used discrimination index, validation test and reliability test based on Pearson and Cronbach Alpha method.

## 2. RESEARCH QUESTION

The research question is how to improve the H-index for accommodating the group of researchers who have a number of highly cited papers, and researchers those have many papers but with a few citations?

Gagolewski<sup>7</sup>, studied the use of the geometric analysis to the data set of the citation distribution. They performed analysis by mapping the characteristic of the distribution data set of researcher's citations. Three types of researchers are defined as follows i.e. perfectionist researcher, moderate researcher, big producer. Among these, the productive and perfectionist researchers are not well accommodated in H-Index calculation method<sup>1,3,10</sup>.

Radko in 2016, classified four groups of researchers have the same H-index are as follows:

- researchers who have  $n$  number of paper and  $n$  citation
- researchers who have "infinitely many" number of papers and  $n$  citation
- researchers who have  $n$  number of papers and "infinitely many" citations
- researchers who have  $n$  number of papers with "infinitely many" number of citations, plus they have "infinitely many" papers with  $n$  number of citation<sup>3</sup>.

Three groups were excluded on H-index, i.e. group B, C and D. Figure 1 shows four groups researcher (A, B, C and D).

We would like to introduce the notion of fairness in our RA-index. The definition of fairness in Cambridge Dictionary is "the quality of treating people equally or in a way that is right or reasonable"<sup>16</sup>. From Fig. 1, it is unfair if the researchers' performance of researcher groups i.e. B, C and D have the same as h-index value of group A, because they have more citation and or more papers. Group B, C and D should have the higher h-index than group A.

Likewise, it will be fair if group D has the highest h-index. In order to make the H-index calculation more sensitive and fairer? A method to measure the fairness index is provided. Discrimination index used to differentiate the calculation of index value of the group of A, B, C and D. Discrimination is opposite of fairness. Discrimination is treatment of a person or group of persons differently, because

of skin colour, sex, sexuality<sup>17</sup>. Discrimination in this paper is performance calculation treatment of a person or group of researchers differently, because several of their number of papers and citations in upper-h-tail, lower-h-tail<sup>18</sup>, and uncited papers was excluded by the h-index calculation.

Two main parameters for the h-index calculation are the number of citations and the number of papers. Discrimination is due to:

- The number of citations in upper- $h$ -tail and lower- $h$ -tail are not counted
- The number of papers in upper- $h$ -tail and lower- $h$ -tail are not counted.

Uncited papers do not have the contribution to increase the h-index. It seems the paper does not exist, or the paper was not indexed. Likewise, citations' addition for papers in the outside of the h-index core does not contribute to improve the h-index value. If papers have no citations, it means the papers do not have much influence on science, and therefore, they have no great results<sup>19</sup>.

Nevertheless, in our country, which has many new authors need to have the h-index, this reason makes a problem. They have been hard work but may need more time to get a citation for their papers. On the other hand, our government policy requires the h-index value for the researcher career. So, we need calculation method that has been more sensitive than h-index.

Therefore, new measuring tool to calculate the discrimination index and can be used to rank scientists required. Discrimination index is important, because it can contribute to differentiate researchers who have the same h-index by calculating the discrimination it receives from the results from the h-index calculations.

Figure 2 shows the comparison of H-index from the citation data set. In this work, data set is gained from the Scopus

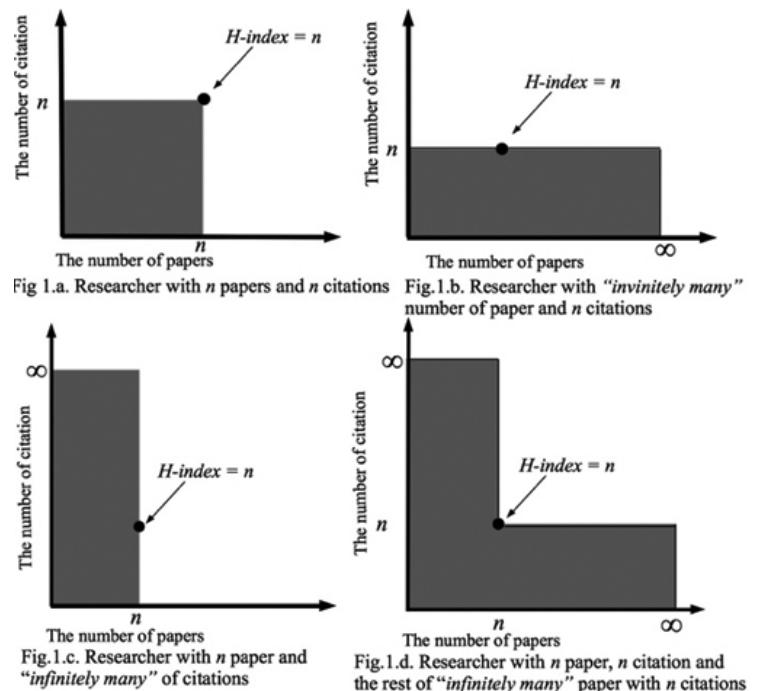


Figure 1. Four group of researchers (A,B,C, and D).

database in January 3<sup>rd</sup>, 2018. This figure shows that Egghe has the H-index of 24, which is lower than that of Sangwal's H-index (26), although the top cited papers and the number of papers of Egghe are more than Sangwal. This shows that there is a discrimination on that case. Therefore, it is important to measure the level of discrimination of the indices' method. The index value that has a smaller discrimination index, it is fairer<sup>20</sup>.

Figure 2 shows an example of where the bias. The high cited papers of Egghe are more than those of Sangwal, but the H-index of Sangwal, which is 26, is higher than the H-index

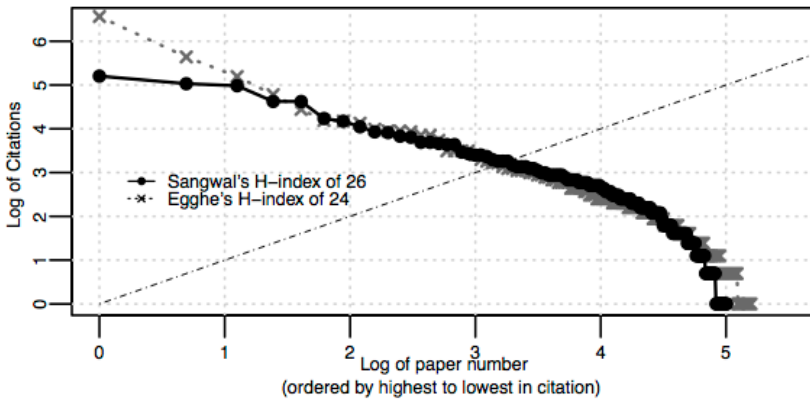


Figure 2. The discrimination of H-index.

of Egghe (24). The first hypothesis is that the high citation of paper of an author does not correlate directly to the H-index. The second hypothesis is that the number of indexed papers on the lower-h-tail does not contribute to the increase of the H-index.

Figure 2 shows the weaknesses of the H-index. This figure shows the log diagram of the number of paper and the citation for two top researchers (Egghe and Sangwal). The production of paper by Egghe is more than that of Sangwal, but the H-index of Sangwal is 26, which is higher than Egghe (24).

In order to provide the figure of the discrimination for this work, a calculation comparison sample of the H-index calculation value is used. The parameters to be compared are the number of cited papers, the total number of citations, and the number of papers. Figure 2 shows the comparison of H-index from the citation data set.

To compare of the researchers based on their h-index, three parameters has been considered are as follows.

- First parameter, comparing the number of cited papers of researcher ( $m$ ) and number of cited papers of h-index ( $n$ ), as shown in Fig. 3a.
- Second parameter, comparing the total number of citations of researcher ( $aobpa$ ) and area of h-index calculation ( $onpn$ ), as shown in Fig. 3b.
- Third parameter, comparing the number of indexed papers of a researcher ( $a$ ) and number

of indexed papers based on h-index calculation ( $n$ ), as shown in Fig. 3c.

The discrimination index value is obtained from some vectors such as the absolute and real vector obtained<sup>21</sup>. It has a value between 0 and 1. The discrimination index is the opposite of the fairness index<sup>21</sup>.

$$\text{Discrimination index} = 1 - \text{Fairness index} \quad (1)$$

The definition of fairness index here is perceived as a factor of fairness received and represented into a number of properties<sup>21</sup>. Fairness index can be calculated as follows:

$$f(x) = \frac{\left[ \sum_{i=1}^n x_i \right]^2}{N * \sum_{i=1}^n x_i^2}, \text{ where } x_i \geq 0 \quad (2)$$

Discrimination index can be defined as:

$$\text{Discrimination} = (1 - \text{Fairness index}) * 100\% \quad (3)$$

Parameters to calculate the discrimination index in this work consists of: 1) the index value, 2) the number of the cited papers, and 3) the sum of the number of citations ( $NoC$ ). The sum of the number of citations is as follows

$$NoC = NoC_1 + NoC_2 + \dots + NoC_n \quad (4)$$

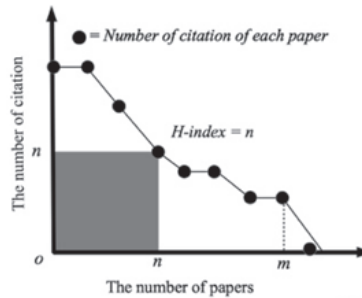


Fig 3.a. Illustration of ratio of  $n$  and  $m$

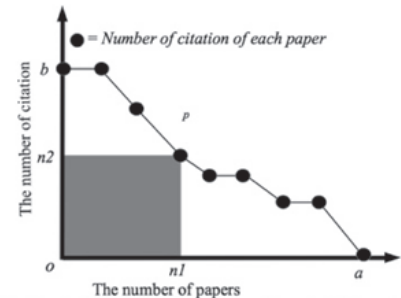


Fig 3.b. Illustration ratio of the total number of citation of researcher in  $obpa$ -area and area of H-index calculation ( $onpn$ )

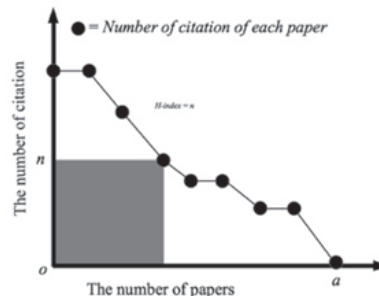


Fig 3.c. Illustration ratio of  $n$  vs  $a$

Figure 3. Parameters to calculate the discrimination level of H-index.

The sum of the number of citations =

$$\sum_{i=1}^n NoC_i \tag{5}$$

We proposed some parameters for calculating the discrimination index as follows.

- $\alpha 0$  is the square of the H-index, and  $\alpha t$  is a total number of citations of a researcher
- $\beta 0$  is the number of papers of the H-index calculation, and is  $\beta t$  the number of papers of researcher
- The total number of the cited papers of researcher is  $\gamma t$ , and the number of cited papers in H-index calculation is  $\gamma 0$ . It depends on the H-index value, for example  $\gamma 0 = 3$ , the number of cited papers based on H-index calculation is 3. Likewise,  $\gamma 0 = 3$ , so the number of the cited papers based on H-index calculation is 10.

Discrimination index can be calculated as:

$$\alpha = \alpha t / \alpha 0 \tag{6}$$

$$\beta = \beta t / \beta 0 \tag{7}$$

$$\gamma = \gamma t / \gamma 0 \tag{8}$$

$$Discrimination\ index = 1 - \left( \frac{|a+b+g|^2}{3 * (a^2+b^2+g^2)} \right) \tag{9}$$

Equations of (1), (2) and (9) are given by Jain (1985)<sup>21</sup>.

### 3. DATA COLLECTION

In order to prove the proposed method, this paper used the citation data set from Scopus. Here, the author names are from the list of the top cited researcher from the webometrics site<sup>22</sup>. The webometrics site provides the ranking of highly cited researchers' data. On the other hand, Scopus is a database owned by Elsevier. Scopus indexed journals and books along with bibliographic data of authors and publishers. Scopus has indexed more than 22,748 peer-reviewed journals and over 3,643 journals with full open access types<sup>23</sup>.

Scopus provides data access services through two ways as follows 1) Accessing the web with an authenticated user password at <http://scopus.com>, 2) Accessing the data via an Application Programming Interface (API) by token authentication and Internet Protocol (IP) listed by Scopus, via URL address <http://api.elsevier.com>.

Both modes of access can be done if the user has been registered to Scopus services. This paper used the API access, since the website access takes a lot of time to retrieve citation data set. For example, data set of a researcher can take 3-5 minutes to be retrieved by the web access, whereas the data taken in this study was about 1,000 researchers' data set.

Indeed, accessing data through API service is preferable. Here, Scopus has data set restriction through API as follows 1) Use a given API key, with a maximum limit of 10 API keys per user, 2) Limits the data queries on Scopus to 50,000 queries per seven days, 3) Each query is limited to 200 results, and 4) Each citation data takes 2-3 seconds per citation per paper.

Based on Scopus restrictions, it is necessary to define

query method that satisfied those restrictions. The step is as follows:

- Getting the author name list of top cited researchers from several countries, which represents regions, i.e. Argentina, Philippine, Israel, Mexico, Portugal and Spain. Author name list of researchers is taken from Webometrics site. The data was accessed from the web, processed and saved as a spreadsheet file
- Retrieving the Scopus ID number data of the researcher using the API access provided by Scopus via an URL address of <http://api.elsevier.com>.

Figure 4 shows our method to get data from some provision that was used the Scopus API service as follows, limitations of data queries on the Scopus 50,000 query database every seven days, and each query is limited to 200 results. It takes 2-3 seconds per citation per paper to be processed.

Data parsing has been conducted to file data of the researcher's name and the researcher number in Scopus (Scopus ID). Table 1 shows the data of the first stage of parsing a result.

Table 2 shows a total sample of 1,710 data set citation by Scopus which is obtained before being filtered. Figure 4 shows a sample of timing diagram of the query process of 539 paper of a researcher. In this research, we created an application and run 10 script of code simultaneously.

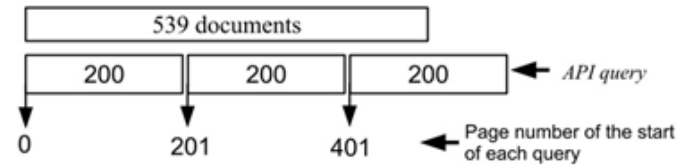


Figure 4. A sample of split query method for accommodating limitation of Scopus API Query of 200 query per session (sample of 539 paper).

Table 1. A sample of the data representation after being parsed.

Name	Scopus ID	Number of documents
researcher 1	7202074046	171
researcher 2	55911103200	10
researcher 3	35406091300	106
researcher 4	55978052600	1
researcher 5	36077704000	148
...	...	...
researcher 999	16151582900	195
researcher 1000	35463345800	195

Table 2. Distribution of researcher with the H-index of 1 to 216

H-index	1 to 54	55 to 108	109 to 162	163 to 216
Data Distribution	761	793	141	15

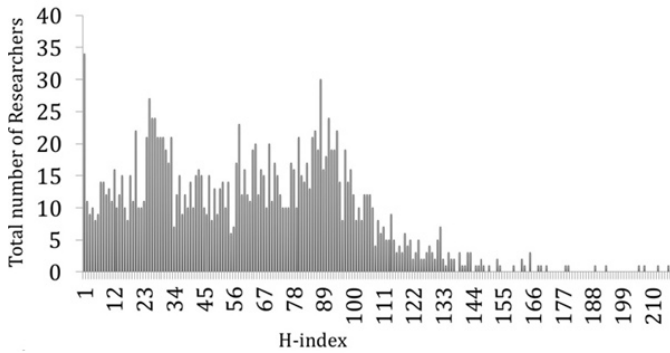


Figure 5. H-index distribution data from 1,710 from Scopus, list of top scientist names from webometrics. (Webometrics, March 2017)

4. DATA PREPARATION

After data collection process finished, we obtained the data of 1,710 researchers, and then continued to prepare the data set.

Figure 5 shows the Scopus data of the search results from the list of names of world’s top webometrics researchers. The researchers’ data were clustered on the range of the H-Index. Table 2. shows the data recapitulation obtained from the range of H-index from 1 to 54, 55 to 108, 109 to 162 and 163 to 216.

Table 3. Data distribution of 1,710 top world scientists (webometrics version, March 2017)

Mean	Median	Minimum	Maximum	H-index range
7,9	7	0	34	1 to 216

5. METHODOLOGY

The weaknesses of the H-Index have currently become some objects to debate, and several proposals for the calculation method suggest alternative schemes for improving the H-Index method. Therefore, we propose our RA-Index method with the advantage of accommodating lower-h-tail and uncited-h-tail.

Figure 6 (a), shows two shading areas A (upper-h-tail) and B (lower-h-tail) that was excluded in H-index calculation. Figure 6 (b). shows the calculation of G-index, which accommodate the area of A in Figure 6 (a). Figure 6 (c). It shows the RA-index, which consider half of area C in Figure 6 (b).

Methodology to construct RA-index is as follows:

- Parameters Identification  
This step identifies parameters used in RA-index
- Identification of the citation data set
- Calculation RA-index (1<sup>st</sup> stage)  
On this step the calculation of RA-index based on G-index used Lotka’s Law approach
- Calculation RA-index (2<sup>nd</sup> stage)  
This step calculates lower-h-tail and uncited-h-tail areas
- Integration RA-index.

5.1 Parameters Identification

To identify RA-index, we used the parameters as follows:

- Geometric area concept of the H-index method
- Accumulation of the number of citations based on G-index
- Giving weight for uncited papers and indexed paper that has the number of citations that is less than the H-index value
- Measurement of the discrimination index of the RA-index

5.2 Identification of the Citation Data Set

Identification of the characteristic of citation. The function set of a citation data set can be stated as follows<sup>7</sup>:

- Figure 7 shows data set citation function, the data were sorted from the highest to the lowest number of citations.
- Function  $F_c(x)$  describes the characteristic of the data set function as a step function. Each step is represented by  $x$ , and  $x$  is an element of natural numbers.
- All of  $x$  are members of the positive real numbers, so that the function  $F_c(x)$  is a member of the natural number.

5.3  $RA_i$  Calculation

The third step is the calculate the area of upper-h-tail and h-core of citation function.

The equation can be written as follows:

$$RA_{i-1} = \sqrt{\sum_{i=1}^n c_i} \tag{10}$$

$$RA_i = RA_{i-1}, \text{ which is } RA_{i-1} \geq i \tag{11}$$

$RA_i$  is similar with  $i$ , where  $i$  is the ranking number of the paper that has a smaller value equal to  $RA_i$ . Eq. (13) is the

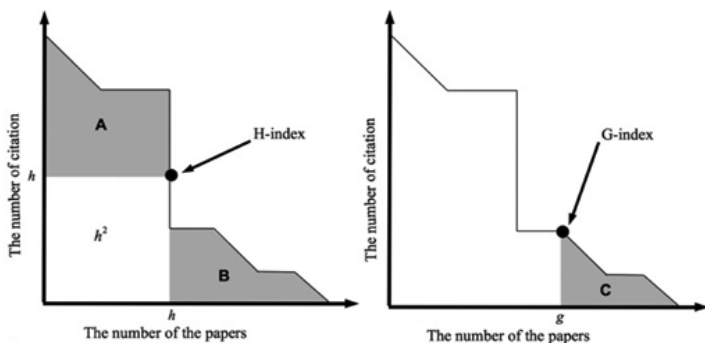


Fig. 6.a. A and B areas those were excluded in H-index calculation

Fig. 6.b. G-index accommodate area of A

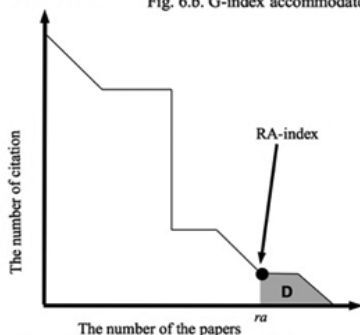


Fig. 6.c. RA-index accommodate area of C

Figure 6. Comparison of H-index (a), G-index (b), an RA-index (c).

square root of the sum of the number of citations of the papers. From our observation, the  $RA_1$  is similar to the  $g$  value in the G-index.

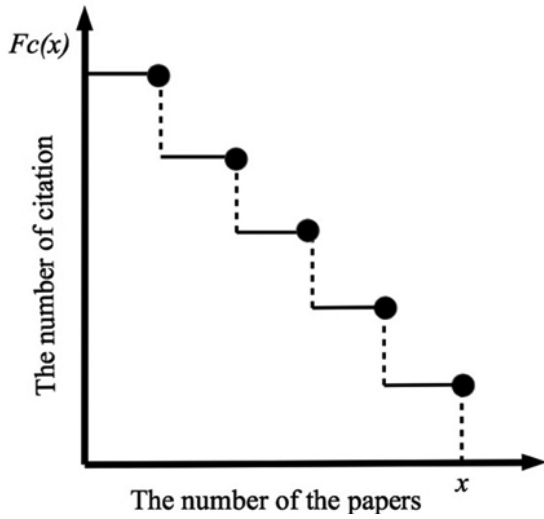


Figure 7. Citation function of the number of papers and citation.

5.4  $RA_2$  Calculation

The fourth step considers the third parameter. The parameters are 1) the number of uncited papers and 2) the number of papers with the citations, which is lower than  $RA_1$  value. Based on Lucio method, the number of papers that can be used as a reference for the performance of the researchers <sup>24</sup>. The equation can be written as follows:

$$RA_2 = \sqrt{(a * tts) + (b * (td - tts - RA_1))} \tag{12}$$

- $td$  = number of indexed papers
- $ts$  = number of cited papers
- $tts$  = number of uncited papers.

$RA_2$  is a parameter to increase the fairness of H-index. The parameter consists of values of lower- $h$ -tail ( $b * (td - tts - RA_1)$ ) and uncited papers value ( $a * tts$ ). We use the approach the square root of the sum (RA) of the lower- $h$ -tail and the uncited papers value to get the value of  $RA_2$ .

The importance of the weighting of uncited papers cannot be ignored, because they may be temporarily unnoticed<sup>25</sup>. If uncited paper and lower- $h$ -tail are not used to increase the researcher’s impact value. In general, researchers need a lot of effort, time and resources to make a paper. An uncited paper is not considered to exist or published, it will have no impact value.

$a$  and  $b$  are the RA-index coefficient of 0.5.  $a$  and  $b$  are the coefficient number that gives the weighting on the uncited papers and the indexed paper that has citation less than the H-index value. On the future work, we will try to optimise the  $a$  and  $b$  coefficients.

5.5 The final step is to integrate the RA-index

The final equation of the RA-index can be written as follows:

$$RA = RA_1 + RA_2 \tag{13}$$

$$RA - index = RA, \text{ in which } i \leq RA \tag{14}$$

Finally RA-index is  $RA$  where is  $RA$  less than or equal to the paper number of  $i$ .

Figure 8 shows the concept of RA-index area of the calculation method. RA-index is relatively higher than the G-index of  $\Delta CD$ , because RA-index considers uncited- $h$ -tail and lower- $h$ -tail areas to increase the  $ra$  value. Therefore, from the propose method the *productive researchers* index values will be accommodated.

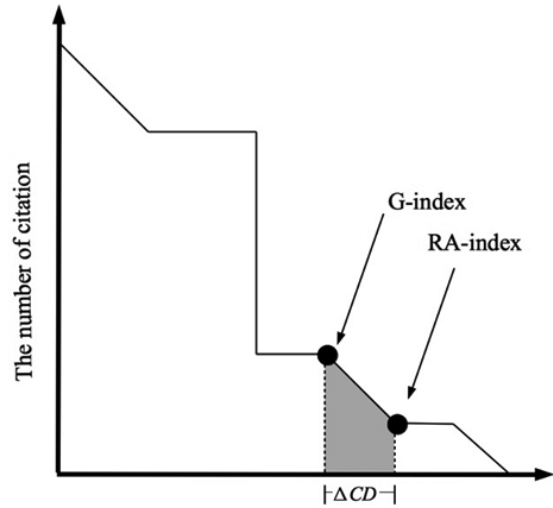


Figure 8. RA-index area.

The following example helps to understand RA-index calculation. Case: The citation data set of an author with nine papers with the data set citation of (6,5,4,3,2,1,0,0,0).

From the data set citation, we found that:

- number of indexed papers ( $td$ ) = 9
- number of cited papers ( $ts$ ) = 6
- number of uncited papers ( $tts$ ) = 3

Let consider, that to find RA-index, the calculation of RA-index for the 1<sup>st</sup> stage:

Table 4 shows the number of citations ( $TC$ ), paper numbers ( $i$ ), sum of citations  $\sum TC$  and calculation

Table 4. Calculation of RA-index

$TC$ (citation)	$i$	$\sum TC$	$\sqrt{\sum TC} = RA_{1-i}$
6	1	6	2.45
5	2	11	3.31
4	3	15	3.87
3	4	18	4.24
2	5	20	4.47
1	6	21	4.58
0	7	21	4.58
0	8	21	4.58
0	9	21	4.58

$$RA_{i-1} = \sqrt{\sum_{i=1}^n c_i}$$

From Eq. (11):  $RA_1 = RA_{i-1}$ , Which is  $RA_{i-1} \geq i$

Table 4 shows that the highest  $\sum TC$  4.58, but lower than  $i$ .

The requirement  $RA_i$  is  $RA_{i-1} \geq i$ , so

$$\begin{aligned} RA_{1,4} &= 4.24 \geq 4, i = 4, \\ RA_{1,5} &= 4.47 \leq 5, i = 5, \\ RA_1 &= i, \text{ where } RA_{1,i} \geq I, \\ RA_1 &= 4 \end{aligned}$$

After get  $RA_i$  continue to calculate RA-index in the 2<sup>nd</sup> stage. From the Equation (14):

$$RA_2 = \sqrt{(a * tts) + (b * (td - tts - RA_1))}$$

Filling the value of  $td$ ,  $ts$ ,  $tts$  and  $RA_1$ :  
We find,

$$RA_2 = \sqrt{(0.5 * (09 - 6)) + (0.5 * (6 - 3 - 3))}$$

And then,

$$RA_2 = \sqrt{(1.5 + 0)}$$

$$RA_2 = 1.2$$

Continue to add  $RA_2$  and  $RA_1$ , that we find the RA-index:

$$\begin{aligned} RA_{3-i} &= RA_1 + RA_2 \\ RA_{3-i} &= 4 + 1.2 \\ RA_{3-i} &= 5.2 \\ RA - index &= i, \text{ which is } i \leq RA_{3-i} \\ RA - index &= 5 \end{aligned}$$

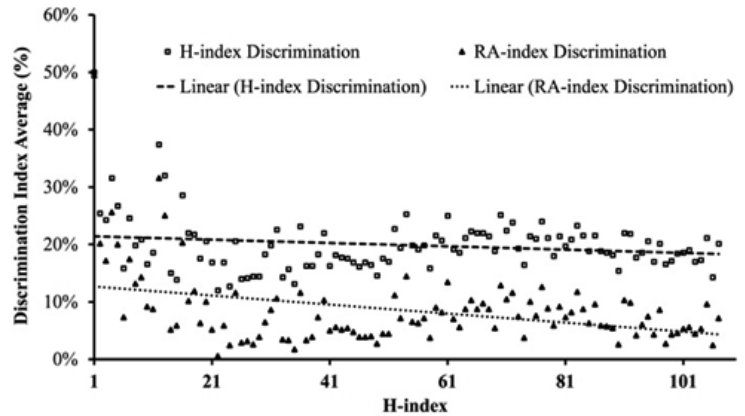
## 6. RESULTS AND DISCUSSIONS

This study finds that the Correlation test between H-index and RA-index by Pearson's Product Moment method has result of 0.973. To measure the fairness of the H-index and to compare with the RA-index we used a discrimination-index based on Jain's fairness index. Table 5. shows the comparison of Discrimination the H-index and the RA-index.

**Table 5. The discrimination comparison of H-index and RA-index**

Statistic results	Discrimination Index	
	H-index (%)	RA-Index (%)
Mean	20	9
Median	18	5
Minimum	2	0
Maximum	73	73
Deviation Standard	9	11

Our test result shows that the H-index is ranged from 1 to 107, and the average of discrimination index of H-index is 20%. This is higher than the RA-index's discrimination index which is 9%. This mean that RA-index is fairer than H-index. However, the deviation standard of RA-index is 11% higher than that of H-index's, which is 9%, because of the existence of the  $a$  and  $b$  parameters for giving weight to lower- $h$ -tail and uncite- $h$ -tail areas of RA-index to optimize the calculation. Figure 9 shows the comparison of the discrimination index between H-index and RA-index. The  $x$  axis is based on H-index. RA-index discrimination in red dotted line is lower than the H-index discrimination in blue dotted line.



**Figure 9. Distribution of the discrimination index for H-index and RA-index.**

For measuring the validity of RA-index, we used the Pearson Product-Moment Correlation Coefficient method. We found that RA-index has a high correlation of the number of indexed paper (P.i), number of cited paper (P.c) and uncited paper (P.no.c). Table 6 shows the correlation of RA-index. The correlation of RA-index is higher than H-index. It could be stated that the RA-index was closer to the expectation than the H-index.

For measuring the reliability of RA-index, we used Cronbach Alpha method. We found that RA-index has a Cronbach Alpha score of 0.842.

**Table 6. Comparison of the validity test of H-index and RA-index**

	H	RA
P.i	0.502	0.634
P.c	0.519	0.637
P.no.c	0.224	0.399

## 7. CONCLUSIONS

This paper proposed a new method called RA-index for improving the sensitivity of H-index to accommodate productive researchers and perfectionist researcher types. It can be reported that the average of the discrimination index of the RA-index is 9%. This is lower than the H-index's discrimination index (20%). The RA-index accommodates improving the sensitivity of the H-index on the number of

highly cited papers and has many cited paper and uncited paper under the H-core. This improvement can enhance the measurement sensitivity of the H-index. The correlation test between H-index and RA-index is 0.973 by the Pearson's Product Moment. The reliability test using Cronbach Alpha of RA-index is 0,842. The Pearson correlation test and Cronbach Alpha test of RA-index is high, which means that the RA-index is valid and reliable as an alternative method for quantifying researcher's impact.

Furthermore, we found that the standard deviation of the RA-index method is 11%, which is higher than that of the H-index method (9%). High standard deviation indicates that the discriminations' points are spread out over a wider range of the discrimination index. The RA-index has the lowest average discrimination index. It can be concluded that the RA-index is fairer than the H-index.

## REFERENCES

- Bormmann, L.; Society, M.P.; Munich, D. & Marx, W. The h index as a research performance indicator. *European Science Editing*, 2011, **37**(3), 77–80.
- Waltman, L. A review of the literature on citation impact indicators. *J. Informetr.* **10**(2), 365–391 (2016). doi: 10.106/j.joi.2016.02.007.
- Mesiar, R. & Gagolewski, M. H-index and other sugeno integrals: Some defects and their compensation. *IEEE Trans. Fuzzy Syst.*, 2016, **24**(6), 1668–1672. doi: 10.1109/TFUZZ.2016.2516579.
- Zitt, M. Paving the way or pushing at open doors? A comment on Abramo and D'Angelo 'Farewell to size-independent indicators'. *J. Informetr.*, 2016, **10**(2), 675–678. doi: 10.1016/j.joi.2016.04.003
- Abramo, G.; D'Angelo, C.A. & Viel, F. The suitability of h and g indexes for measuring the research performance of institutions. *Scientometrics* **97**(3), 2013, 555–570. doi: 10.1007/s11192-013-1026-4.
- BiHui, J. The R- and AR-indices: Complementing the h-index. *Chinese Sci. Bull.*, 2007, **52**, 855–863. doi: 10.1007/s11434-007-0145-9.
- Gagolewski, M. & Grzegorzewski, P.A. Geometric approach to the construction of scientific impact indices. *Scientometrics*, 2009, **81**(3), 617–634. doi: 10.1007/s11192-008-2253
- Alonso, S.; Cabrerizob, F.J.; Herrera-Viedmac, E. & Herrera, F. H-Index: A review focused in its variants, computation and standardisation for different scientific fields. *J. Informetr.*, 2009, **3**(4), 273–289. doi: 10.1016/j.joi.2009.04.001.
- Egghe, L. Theory and practise of the g-index. *Scientometrics*, 2006, **69**(1), 131–152. doi: 10.1007/s11192-006-0144-7.
- Kosmulski, M. MAXPROD - A new index for assessment of the scientific output of an individual, and a comparison with the h-index. *Cybermetrics*, 2007, **11**(1), 14–18.
- Prathap, G. Qualifying Scholarly Impact using an iCX (impact-Citations-Exergy) Analysis. *DESIDOC J. Libr. Inf. Technol.*, 2011, **31**(5), 382–386. doi: 10.14429/djlit.31.5.1196.
- Abramo, G. & D'Angelo, C.A. A Farewell to the MNCS and like size-independent indicators. *J. Informetr.*, 2016, **10**(2), 646–651. doi: 10.1016/j.joi.2016.04.006.
- Rodríguez-Navarro, A. & Brito, R. Double rank analysis for research assessment. *J. Informetr.*, 2018, **12**(1), 31–41. doi: 10.1016/j.joi.2017.11.004.
- Bornmann, L.; Mutz, R.; Hug, S.E. & Daniel, H.D. A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *J. Informetr.*, 2011, **5**(3), 346–359. doi: 10.1016/j.joi.2011.01.006.
- Wang, C.; Li, Y.; Chen, D. & Huang, M. Uses Unexpired patent to re-evaluate innovation performance by h-index and h-core Indicators. *PICMET*, 2013, 965–971.
- Cambridge. Fairness meaning in the Cambridge English Dictionary. (2017). Available at: <http://dictionary.cambridge.org/dictionary/english/fairness>. (Accessed: 17th January 2017)
- Cambridge, Meaning of 'Discrimination' in the English Disctionary. (2017). Available at: <http://dictionary.cambridge.org/dictionary/english/fairness>. (Accessed: 21st March 2018)
- Zhang, C.T.; The h'-Index, Effectively improving the h-Index based on the citation distribution. *PLoS One*, 2013, **8**(4), doi: 10.1371/journal.pone.0059912
- Hirsch, J.E. & Buela-Casal, G. The meaning of the h-index. *Int. J. Clin. Heal. Psychol.*, 2014, **14**(2), 161–164. doi: 10.1016/S1697-2600(14)70050-X.
- Rochim, A.F.; Muis, A. & Sari, R.F. Discrimination measurement method on H-index and G-index using Jain's fairness index. *Proc. ISSI 2017 - 16th International Conference Scientometrics and Informetrics* 446–447 (2017).
- Jain, R.; Chiu, D.M. & Hawe, W.R., A quantitative measure of fairness and discrimination for resource allocation in shared computer system. *DEC Technical Report TR301 cs.NI/9809*, 1984, 1–38.
- F. Aguillo, I. Highly cited researchers (h>100) according to their Google Scholar Citations public profiles. *CSIC, Madrid, Spain* (2017). Available at: <http://www.webometrics.info/en/node/58>. (Accessed: 1st September 2017)
- Scopus. Content - Scopus - Solutions | Elsevier. *Scopus* (2017). Available at: <https://www.elsevier.com/solutions/scopus/content>. (Accessed: 14th May 2017)
- Bertoli-Barsotti, L. & Lando, T. On a formula for the h-index. *J. Informetr.* 2015, **9**(4), 762–776. doi: 10.1016/j.joi.2015.07.004.
- Lou, W. & He, J. Does author affiliation reputation affect uncitedness? *Proc. Assoc. Inf. Sci. Technol.*, 2015, **52**, 1–4.

## ACKNOWLEDGEMENT

This work was supported by Research Grant Funding of



Tugas Akhir Doktoral (TADOK) from Universitas Indonesia No.1361 /UN2.3.1/HKP.05.00/2018.

## CONTRIBUTORS

**Mr Adian Fatchur Rochim** is PhD Candidate in Department of Electrical Engineering, Computer Engineering, Faculty of Engineering, Universitas Indonesia. He received Bachelor degree at electrical engineering Diponegoro University and Master degree at Computer Engineering, Bandung Institute of Technology. He research interest are Scientometrics, and H-index variants. He is working on 'impact of researcher using quantity on the number of publication and citation'. He has proposed Fairness Index by Raj Jain to measure the fairness of H-index and variants.

**Dr Abdul Muis** received his bachelor degree from Electrical Engineering Universitas Indonesia in 1998. He received his Master and PhD degrees in mechatronics from KEIO University, Japan in 2001 and 2007. He joined Electrical Engineering Department as permanent lecturer in 1999.

**Prof. Riri Fitri Sari**, is a Professor of Computer Engineering at the Department of Electrical Engineering, Faculty of Engineering, University of Indonesia. She was the CIO / Head of Information System Development and Services of the University of Indonesia UI (2006-2014). She received her PhD from Leeds University, UK (2004). She is currently actively teaching and researching in the field of Internet of Things, Computer Network, Protocol Engineering, and the implementation of Information and Communication Technology. Since 2010, she has been actively involved with University Ranking and has been the Chairperson of UI GreenMetric Ranking of World University since then.