

New exploratory work of evaluating a researcher's output

JUNPING QIU, RUIMIN MA, NI CHENG

Research Center for Chinese Science Evaluation, Wuhan University, Wuhan (P. R. China)

SCI has been popular all over the world since it was published by Garfield in 1963. Researches on evaluating a researcher's output with SCI have always been continuous. In recent years, a great breakthrough has been made since the *h*-index was put forward in 2005. In this paper, we also advance a new method – Paper Quality Index (*PQI*) to evaluate the output of a researcher. The main purpose of our method is to solve two problems that consist in the method of *h*-index: one is that the *h*-index can't compare the outputs of researchers in different fields; the other is that it is unsuitable for evaluating the outputs of young researchers. A simple mathematical expression is constructed to eliminate the difference of citation among different fields and makes the evaluation of short-term outputs of researchers possible.

Introduction

The methods of evaluating the output of a researcher were discussed widely in recent years. Especially, the physicist HIRSCH [2005] advanced a new method named *h*-index that has attracted many interests of authors in different disciplines. Thereinto, the scientometricians have undoubtedly played an important role in promoting the application of *h*-index. BRAUN & AL. [2006] expanded its application to journal

Received November 12, 2007

Address for correspondence:

RUIMIN MA

Research Center for Chinese Science Evaluation, Wuhan University

Wuhan, P. R. China, 430072

E-mail: ruimin.ma@yahoo.com.cn

0138–9130/US \$ 20.00

Copyright © 2008 Akadémiai Kiadó, Budapest

All rights reserved

evaluation. EGGHE & ROUSSEAU [2006] introduced an informetric model for the h -index. EGGHE [2006A] advanced an improvement of the h -index and named g -index. All these studies make the evaluation of outputs of researchers effective and scientific.

In fact, the h -index is an original and simple new measure incorporating both quantity and quality of publications: "A scientist has index h if h of his or her N_p papers have at least h citations each and the other (N_p-h) papers have fewer than $\leq h$ citations each" [HIRSCH, 2005]. However, there are two problems in the evaluation of researchers' outputs via h -index. Firstly, it can't evaluate researchers' outputs in different fields. As known to all, the situations of publication and citation are different among fields. The number of the papers in a given field, the average citations of per paper in the field, and the size of the field can influence the h -index calculation directly [IGLESIAS & PECHARROMÁN, 2007; BORNMANN & DANIEL, 2007]. Papers that are easy to be cited frequently in some field (e.g., biology) will probably lead to a higher h -index of a scientist in corresponding field than one in some other field (e.g., library and information science). In addition, the h -index is unsuitable for evaluating a young scientist's output. Because the h -index values increase over time [HIRSCH, 2005], some older scholars didn't write any paper at all in recent years, but their h -indexes are still high. The h -index always puts the newcomer at a disadvantage and the older scientist at an advantage [BORNMANN & DANIEL, 2007].

Therefore, we manage to advance a method "Paper Quality Index (PQI)" so as to overcome the two problems that h -index has brought.

Method

Basic thoughts about the method

In our opinions, the quality of a paper is determined by two factors: the journal quality and its own performance. Journal is the carrier of papers. A good carrier will promote the paper's level [SUN & XU, 2002]. For example, papers published in *Science* or *Nature* are always considered as the top ones in the corresponding fields. To some extent, the level of a journal can represent the quality of the papers published in it. At present, the journal's level is always measured by the index "impact factor (IF)" [QIU, 2007]. But the average IF should be used to measure the level of a journal since the IF s of a journal are fluctuant in different years. In our method, we use the average IF s of a journal in recent three years, namely 2004, 2005, and 2006 (note that JCR has not provided the 2007 IF data). On the other hand, the performance of a paper mainly depends on its cited times. Generally speaking, a highly cited paper means that it focused a lot of attentions and had wide influence. Only by combining these two factors properly can a paper's quality be measured more accurately. But once a paper has been not cited at all, it is unvalued no matter how great its carrier is.

What's more, the difference among fields should be eliminated before the comparisons of outputs of the researchers in different scientific fields. The difference consists of two parts: the difference of journal impact factor and paper's citation among different field. It is meaningless to compare two journals' absolute *IF*s in two fields with low correlation. For example, the *IF* of any journal in the field of biology may exceed the maximum *IF* in the field of library and information science. But one can't reach the conclusion that the performance of every journal in library and information science is worse than that of any journal in biology. Suppose that there is a journal "A" with 4 *IF* in a certain year belonging to the field of biology with 5 *IF*¹ and a journal "B" with 1.5 *IF* in the same year belonging to the field of library and information science with 1 *IF*. If we want to compare these two journals in terms of *IF*, we need normalize them according to the *IF* of the corresponding field. So selecting which method to measure the performance of a field should be taken into account. We use the ratio of a certain journal's *IF* to the *IF* of the corresponding field as the normalization method. It is a relative index. Therefore, the relative *IF* of journal A is 0.8 (namely 4/5) and that of journal B is 1.5 (namely 1.5/1). Now, one can say that as for this year, the performance of journal B is better than Journal A although its absolute *IF* is lower than Journal A's. The same principle can be applied in the normalization of citations of the papers in different fields.

In addition, we should take into account the citation evolution of a paper. As for the *h*-index, once a paper is selected to the top group, the *h*-index calculated in subsequent years is not at all influenced by this paper's received citations further on: even if the paper doubles or triples its number of citations (or even more) the subsequent *h*-indexes are not influenced by this [EGGHE, 2006B]. At the same time, one should note that the articles of a researcher might scatter in different fields. For example, T. Braun publishes also in chemistry and L. Egghe in mathematics. Furthermore, we should compare the quality of papers published in the same year. It is meaningless to compare the quality of a paper published in 1980 with that of a paper published in 2006.

In a word, at present, a proper method to solve these problems is urgently required.

Mathematical expression of the method and its explanations

On the basis of the thoughts, we advance a method named "paper quality index" (*PQI*). Its mathematical expression is given below:

$$PQI_{ij} = \frac{\overline{IF}}{IF_m} \times \frac{TC}{TC_m} \tag{1}$$

¹ Here involves the concept "field (or subject) impact factor". Its calculation is number of citations in a certain year to the articles published in the field in previous two years / number of papers published in the field in previous two years [DANG, 2002]. JCR database supplies the value of it but name "Aggregate Impact Factor".

In this mathematical expression, “*i*” stands for a paper. “*j*” stands for the publication year of the paper *i*. “*m*” stands for the field to which the paper *i* belongs. *IF* stands for the impact factor of the journal in which paper *i* published. \overline{IF} stands for the journal's average *IF* in recent three years. *IF_m* stands for the impact factor of the field *m*. \overline{IF}_m stands for the average *IF_m* in recent three years. *TC* stands for the total number of citations of the paper after its publication. TC_m stands for the total citations of the articles published in the year *j* in the field *m*. \overline{TC}_m stands for the average citations of per paper published in the year *j* in the field *m*. The calculation is the total number of citations to the articles published in the year *j* in the field *m*/number of papers published in the year *j* in the field *m*. Note that the citations in 2007 should be taken out, because the JCR has not supplied the corresponding data.

The ratio $\overline{IF} / \overline{IF}_m$ is used to eliminate the difference of *IF*s of journals in different fields and measure the relative quality of the journal. And the ratio TC / \overline{TC}_m is employed to eliminate the difference of citations of papers in different fields and measure the relative performance of the paper. In order to evaluate the quality of the paper synthetically, we use multiplication to join them.

The expression (1) measures the quality of a single paper. A researcher's *PQI* of a certain year can be calculated as follows:

$$PQI_j = \sum PQI_{ij}, i = 1, 2, \dots, n \tag{2}$$

The number *n* equals to the number of all articles published by a researcher in the year *j*. *PQI_j* measures the *PQI* of all papers published in the year *j*. Based on this, the whole *PQI* of a researcher can also be calculated as shown in expression (3).

$$PQI = \sum PQI_j, j = a, b, \dots, m \tag{3}$$

The letters “*a, b, ..., m*” denote a series of years, in which the researcher published his/her articles. From the expressions (1), (2), (3), it is found that the method is a hierarchy system starting from the evaluation of a single paper. In view of time factor, one can limit the evaluation year of *PQI*. We can take advantage of (2) to measure a researcher's *PQI* of a separate year and also use (3) to evaluate a researcher's *PQI* of several selected years. For example, if we want to compare *PQI*s of year 2003 and 2004 of some researchers, we can calculate the *PQI* of a researcher like this: $PQI = PQI_{2003} + PQI_{2004}$.

At the same time, the problem of the citation evolution can be resolved well by *PQI*. “*TC*” (see expression (1)) is used to reflect the evolution track. “*TC*” increases as time goes on, which also can influence the *PQI*'s value. However, if “*TC*” of a paper is 0, its *PQI* is also zero, and this paper contributes nothing to the *PQI* of a researcher. So our method also keeps the advantage of the *h*-index not to take into account the “tail” paper [EGGHE, 2006B].

In a word, our method can help us achieve the aim of comparing the output's quality of researchers in different fields and is also suitable for evaluating the quality of a newcomer's output.

Data collection

Issues deserving

Before the data collection, one should pay attention to some special issues that are given as follows:

1. If a journal belongs to several fields in JCR, one should vest it with the first field. A journal should be vested with a unique field.
2. The field "multidisciplinary sciences" should be considered as an independent one. So far, there is no better way to solve this problem.
3. All papers of a researcher in different fields should be calculated, since the two types of difference are eliminated by our method.
4. For a paper published by co-authors, it is possessed by the first author only.
5. There are many types of document in SCI-E. Only the type "Article" is selected.

Instructional steps

In the followings, a series of steps on how to obtain and calculate the *PQI* of a paper will be given in detail by the instance of Egghe's paper "The source-item coverage of the Lotka function", which was published in *Scientometrics* in 2004.

- Click the link "Times Cited" in SCI-E and enter, then count the citations in each year. One can find that this paper was cited 6 times cumulatively between 2004 and 2006 (or subtract the citations in 2007 from total citations).
- *Scientometrics* is uniquely categorized into the scientific field "COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS", which can be found in JCR or SCI-E easily. Select the field in a certain JCR year and enter. Then one will find a button marked with "View category summary list" on the upper-right corner of the page, click it and then enter into the next page. In this page, one will obtain the data as follows: the value of IF_m in the JCR year (shown as the value of "Aggregate Impact Factor" in Figure 1); articles published in the same year (shown as the value of "Articles" in Figure 1); the citations in the same year to the articles published in previous years (shown in the circle in Figure 2). Following the guidance given above we collect the data of each year and show them in Table 1.

Category: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS

Total Cites	Median Impact Factor	Aggregate Impact Factor	Aggregate Immediacy Index	Aggregate Cited Half-life	Aggregate Citing Half-life	# Journals	Articles
121793	0.862	1.609	0.243	6.0	7.9	87	7531

Figure 1. A part of the screen shot of the category summary list

Aggregate Cited Half-Life
 The cited half-life for the category is the median age of its articles cited in the current JCR year. Half of the citations to the category are to articles published within the cited half-life.

Aggregate Cited Half-Life: 6.0 years

Breakdown of the citations to the journals in the category by the cumulative percent of 2006 cites to articles published in the following years:

Year	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996-all
Cites from 2006	1832	9908	12475	13513	11094	11922	8474	7151	7841	4890	32693
Cumulative %	1.50	9.64	19.89	30.98	40.09	49.87	56.83	62.70	69.14	73.16	100

Figure 2. Another part of the screen shot of the category summary list

- According to the data obtained in previous steps, the value of \overline{IF}_m is 1.522 that is calculated like this: $(1.609+1.577+1.379) / 3$; the value of TC_m is 22108 (1414+8291+12475); the value of \overline{TC}_m is 3.440 (22108/6427, the value of denominator can be found in Table 1). So the value of TC / \overline{TC}_m can also be calculated, it is 1.744 (6/3.440).
- Continue to collect the *IFs* of *Scientometrics* in recent three years, and they are 1.120 in 2003, 1.738 in 2004, and 1.363 in 2005, respectively. The average value of them that equals to \overline{IF} is 1.407 $((1.120+1.738+1.363)/3)$. And so the value of $\overline{IF} / \overline{IF}_m$ can also be calculated. It is 0.925 (1.407/1.522).
- Finally, we obtain the *PQI* of this article according to the expression $PQI_{ij} = \overline{IF} / \overline{IF}_m \times TC / \overline{TC}_m$. It is 1.613 (1.744×0.925).

Table 1. The data collected from “View category summary list” in terms of year

Index	2006 IF_m	2005 IF_m	2004 IF_m	Articles published in 2004	04-04	05-04	06-04
Value	1.609	1.577	1.379	6427	1414	8219	12475

Note: “05-04” means that the citations in 2005 to the articles published in 2004 in the field, so does the other similar indexes.

Similarly, one can calculate the *PQIs* of other articles published by Egghe. Table 2 shows the *PQI* of Egghe’s articles published in 2003 and 2004. PY stands for the publication year of the paper. *PQI* denotes the *PQI* of each paper. *PQI1* stands for

Egghe's *PQI* of each year. It is clear to find that the *PQI*₂₀₀₃ of Egghe is 3.197 and *PQI*₂₀₀₄ is 6.051. The *PQI*_{2003, 2004} is 9.248 (3.197+6.051).

Table 2. *PQIs* of papers published by Egghe as first author (data from SCI-E, JCR)

Journal	Field	<i>PY</i>	<i>PQI</i>	<i>PQI1</i>
<i>JASITS</i>	Computer Science, Information Systems	2003	0.760	
<i>JASITS</i>	Computer Science, Information Systems	2003	2.280	
<i>Mathematical and Computer Modelling</i>	Computer Science, Interdisciplinary Applications	2003	0.100	3.197
<i>Mathematical and Computer Modelling</i>	Computer Science, Interdisciplinary Applications	2003	0.050	
<i>Progress in Natural Science</i>	Multidisciplinary Sciences	2003	0.007	
<i>Information Processing & Management</i>	Computer Science, Information Systems	2004	2.355	
<i>Information Processing & Management</i>	Computer Science, Information Systems	2004	0.471	
<i>Scientometrics</i>	Computer Science, Interdisciplinary Applications	2004	1.613	6.051
<i>Scientometrics</i>	Computer Science, Interdisciplinary Applications	2004	1.344	
<i>Scientometrics</i>	Computer Science, Interdisciplinary Applications	2004	0.269	

Limitations under current JCR report

Our method is based on SCI-E and JCR. For personal research, some data, such as *IF_m* and *TC_m*, should be harvested from them directly, or it makes the method complex. But JCR provides the related data after 2003 (in previous JCR edition, the function "Subject Category Summary List" was not given), so only *IF_m*s in recent 4 years can be used. In our method, we use the *IF_m*s in recent three years. Moreover, as figure 2 shows, the index "Aggregate Cited Half-Life" only provides the data of the past ten years, so the data collection of *TC_m* is limited by the "cited years" given by JCR. For example, it is very hard to get the data of the total citations in 2006 to the all articles published in 1994. So although our method is also suitable for evaluating both long-term and short-term output of a researcher, at present we'd better use *PQI* for short-term evaluation. That is, at present we can only evaluate a researcher's *PQIs* of recent 4 years or less for the constraints of JCR.

An example

Table 3 and Table 4 respectively represent *PQI*₂₀₀₃ and *PQI*₂₀₀₄ of the Price Prize medalists (still active) and Fields Prize medallists awarded between 1990 and 2002. *PQI* is the sum of *PQI*₂₀₀₃ and *PQI*₂₀₀₄. The *h*-index data of Price Prize medalists are cited from EGGHE's article [2006b]. In addition, note that we only use the data of SCI-E of Web of Knowledge (WoK) due to the use constraints in our school.

Let us consider the *PQI* first. Obviously, an author's *PQI* varies along with time. For example, White's *PQI*₂₀₀₃ ranks as the highest one with 10.638 while his *PQI*₂₀₀₄ ranks as the 3rd with 5.802. It may accord with the fact that there is little possibility for

an author to publish articles with high quality every year. Because our method eliminates the difference among scientific fields, the *PQIs* of researchers in different fields can be compared with the same criterion, e.g., the *PQIs* of scientometricians versus that of mathematicians. If we set a threshold of *PQI* at 10 subjectively, Price Prize medalists – White, Glänzel, and Fields Prize medalists – Bourgain, McMulle, Kontsevich, Witten, and Lions are the authors who published high quality papers in 2003–2004. So if the total *PQIs* of researchers in different fields have been calculated, we can also pick out the most outstanding scientists in different fields with the same criterion.

However, *h-index* is unsuitable for the comparison of researchers in different fields. It is meaningless to make comparison of *h-index* for scientometricians and mathematicians. At the same time, *h-index* is most suitable for evaluating a researcher's long-term output. In general, we only give an overall *h-index* for a researcher, so the *h-index* of a researcher is relatively static. But a series of *PQIs* of a researcher can be given in terms of year. For example, White's *h-index* ranks at the bottom with 12 but his *PQI*₂₀₀₃, *PQI*₂₀₀₄, and *PQI* all rank in the front. Garfield's *h-index* ranks the 1st but his *PQI*₂₀₀₃, *PQI*₂₀₀₄ as well as *PQI* all rank at the middle level. The *PQI* can monitor the whole process of the research activities of a researcher nicely while the *h-index* mainly emphasizes on the final result.

Table 3. *PQIs* and *h-indexes* of Price Prize medalists (still active, order by *PQI*)

Author	<i>h-index</i>		<i>PQI</i> ₂₀₀₃		<i>PQI</i> ₂₀₀₄		<i>PQI</i>	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
WHITE	12	14	10.638	1	5.802	3	16.440	1
GLÄNZEL	18	5	3.811	2	6.989	1	10.800	2
EGGHE	13	10	3.197	3	6.051	2	9.248	3
LEYDESDORFF	13	10	0.635	7	4.518	4	5.153	4
GARFIELD	27	1	2.864	4	1.751	6	4.615	5
SMALL	18	5	1.140	6	1.613	7	2.753	6
ROUSSEAU	13	10	0.000	8	1.830	5	1.830	7
BRAUN	25	3	1.384	5	0.000	11	1.384	8
VAN RAAN	19	4	0.000	8	1.344	8	1.344	9
MOED	18	5	0.000	8	1.344	9	1.344	10
INGWERSEN	13	10	0.000	8	0.538	10	0.538	11
NARIN	27	1	0.000	8	0.000	11	0.000	12
SCHUBERT	18	5	0.000	8	0.000	11	0.000	12
MARTIN	16	9	0.000	8	0.000	11	0.000	12

Table 4. *PQIs* of Fields Prize medalists (awarded between 1990 and 2002, order by *PQI*)

Author	<i>PQI</i> ₂₀₀₃		<i>PQI</i> ₂₀₀₄		<i>PQI</i>	
	Value	Rank	Value	Rank	Value	Rank
BOURGAIN	39.265	3	36.964	2	76.229	1
MCMULLEN	65.671	1	0.000	4	65.671	2
KONTSEVICH	41.535	2	0.000	4	41.535	3
WITTEN	2.487	5	37.190	1	39.677	4
LIONS	14.850	4	0.000	4	14.850	5
DRINFELD	0.000	7	4.841	3	4.841	6
VOEVODSKY	1.355	6	0.000	4	1.355	7
LAFFORGUE	0.000	7	0.000	4	0.000	8
GOWERS	0.000	7	0.000	4	0.000	8
YOCCOZ	0.000	7	0.000	4	0.000	8
ZELMANOV	0.000	7	0.000	4	0.000	8
BORCHERDS	0.000	7	0.000	4	0.000	8
VAUGHAN	0.000	7	0.000	4	0.000	8
MORI	0.000	7	0.000	4	0.000	8

Conclusions

Our method *PQI* considers both the quality and quantity of a researcher's output simultaneously. It measures the quality of a paper from two aspects synthetically: one is the quality of its carrier "journal"; the other is its own performance "citations". The quality of a journal is measured with impact factor. In order to eliminate the difference of *IF* and paper's citation among different field, the two indexes are also normalized respectively. Finally, we use the expression PQI_{ij} to measure the quality of a paper; PQI_j , which is based on PQI_{ij} , to measure quality of a researcher's outputs of a certain year; *PQI*, which is also based on PQI_j , to measure the quality of a researcher's outputs of all years. The method is hierarchy system. Moreover, the method is dynamic and suitable for the evaluation of a newcomer's output, which is also tested by an example. Furthermore, the method also takes into account the citation evolution of a paper. At the same time, the limitations of our method caused by JCR are also given in details.

In the end, we must emphasize that our method is very sensitive to the number of the citations of a paper. However, it is a serious problem in some countries that self-citation is not very canonical. But in the WoK, there is only one index "Times Cited" which includes self-citation and it is very hard to get rid of self-citation manually one by one. In order to avoid speculative activities in the scientific research, we suggest that the ISI should consider providing a new index "Times Cited without self-citations". It is very significative for our method as well as *h*-index being applied in the different countries with the different evaluation criterions.

*

The authors would like to acknowledge the support from National Natural Science Foundation of China (70673071/G0309) and constructive comments from the reviewers.

References

- BORNMANN, L., & DANIEL, H. D. (2007), What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58 (9) : 1381–1385.
- BRAUN, T., GLÄNZEL, W., SCHUBERT, A. (2006), A Hirsch-type index for journals. *Scientometrics*, 69 (1) : 169–173.
- DANG, Y. R. (2002), Difference analysis of the time factor on subject impact index. *Journal of the China Society for Scientific and Technical Information*, 21 (3) : 328–333.
- EGGHE, L. (2006A), An improvement of the h-index: the g-index. *ISSI Newsletter*, 2 (1) : 8–9.
- EGGHE, L. (2006B), Theory and practise of the g-index. *Scientometrics*, 69 (1) : 131–152.
- EGGHE, L., ROUSSEAU, R. (2006), An informetric model for the Hirsch-index. *Scientometrics*, 69 (1) : 121–129.
- HIRSCH, J. E. (2005), An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102 (46) : 16569–16572.
- IGLESIAS, J. E., PECHARROMÁN, C. (2007), Scaling the h-index for different scientific ISI fields. Retrieved November 3, 2007, from <http://arxiv.org/ftp/physics/papers/0607/0607224.pdf>.
- QIU, J. (2007), *Informetrics*. Wuhan University Press, Wuhan.
- SUN, Y., XU, K. (2002), Evaluating individual scientific research output by using SCI in reason. *Library Tribune*, 22 (5) : 24–26.