

Equality of Google Scholar with Web of Science Citations: Case of Malaysian Engineering Highly Cited Papers

Nader Ale Ebrahim¹, Hadi Salehi², Mohamed Amin Embi³, Mahmoud Danaee⁴, Marjan Mohammadjafari⁵, Azam Zavvari⁶, Masoud Shakiba⁶ & Masoomeh Shahbazi-Moghadam⁷

¹ Research Support Unit, Centre of Research Services, Institute of Research Management and Monitoring (IPPP), University of Malaya, Malaysia

² Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran

³ Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Bangi, 43600, Malaysia

⁴ Faculty of Agriculture, Roudehen Branch, Islamic Azad University, Roudehen, Iran

⁵ Department of Industrial Engineering, Faculty of Engineering, Science and Research Branch, Islamic Azad University, Kerman, Iran

⁶ Center for Software Technology and Management, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

⁷ Perdana School of Science, Technology and Innovation Policy, Universiti Teknologi Malaysia

Correspondence: Hadi Salehi, Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran. E-mail: hadisalehi1358@yahoo.com

Received: June 1, 2014 Accepted: June 23, 2014 Online Published: August 6, 2014

doi:10.5539/mas.v8n5p63

URL: <http://dx.doi.org/10.5539/mas.v8n5p63>

Abstract

This study uses citation analysis from two citation tracking databases, Google Scholar (GS) and ISI Web of Science, in order to test the correlation between them and examine the effect of the number of paper versions on citations. The data were retrieved from the Essential Science Indicators and Google Scholar for 101 highly cited papers from Malaysia in the field of engineering. An equation for estimating the citation in ISI based on Google scholar is offered. The results show a significant and positive relationship between both citation in Google Scholar and ISI Web of Science with the number of versions. This relationship is higher between versions and ISI citations ($r = 0.395$, $p < 0.01$) than between versions and Google Scholar citations ($r = 0.315$, $p < 0.01$). Free access to data provided by Google Scholar and the correlation to get ISI citation which is costly, allow more transparency in tenure reviews, funding agency and other science policy, to count citations and analyze scholars' performance more precisely.

Keywords: bibliometrics, citation analysis, evaluations, equivalence, Google Scholar, High cited, ISI Web of Science, research tools, H-index

1. Introduction

Citation index as a type of Bibliometrics method traces the references in a published article. It shows that how many times an article has been cited by other articles (Fooladi et al., 2013). Citations are applied to evaluate the academic performance and the importance of information contained in an article (Zhang, 2009). This feature helps researchers get a preliminary idea of the articles and research that make an impact in a field of interest. The avenues to evaluate citation tracking have greatly increased in the past years (Kear & Colbert-Lewis, 2011). Citation analysis was monopolized for decades by the system developed by Eugene Garfield at the Institute for Scientific Information (ISI) now owned by Thomson Reuter Scientific (Bensman, 2011). ISI Web of Science is a publication and citation database which covers all domains of science and social science for many years (Aghaei Chadegani et al., 2013). In 2004, two competitors emerged, Scopus and Google Scholar (Bakkalbasi, Bauer, Glover, & Wang, 2006). Google Inc. released the beta version of 'Google Scholar' (GS) (<http://scholar.google.com>) in November 2004 (Pauly & Stergiou, 2005). These three tools, ISI from Thomson Reuters, Google Scholar (GS) from Google Inc. and Scopus from Elsevier are used by academics to track their citation rates. Access to ISI Web of Science is subscription-based service while GS provides a free alternative to retrieve the citation counts. Therefore, the researchers need to estimate their citation in ISI by knowing the GS

citation counts. On the other hand, publishing a research paper in a scholarly journal is necessary but not sufficient for receiving citations in the future (Nader Ale Ebrahim, 2013). The paper should be visible to the relevant users and authors in order to get citations. The visibility of the paper is defined by the number of paper versions which are available in the Google Scholar database. The number of citations will be limited to the versions of the published article on the web. The literature has shown increased visibility by making research outputs available through open access repositories, wider access results and higher citation impact (Nader Ale Ebrahim et al., 2014; Amancio, Oliveira Jr, & da Fontoura Costa, 2012; Antelman, 2004; Ertürk & Şengül, 2012; Hardy, Oppenheim, Brody, & Hitchcock, 2005). A paper has greater chance of becoming highly cited whenever it has more visibility (Nader Ale Ebrahim et al., 2013; Egghe, Guns, & Rousseau, 2013).

The objectives of this paper are two-fold. The first objective is to find the correlation between Google Scholar and ISI citation in the highly cited papers. The second objective is to find a relationship between the paper availability and the number of citations.

2. Google Scholar & Web of Science Citations

The citation facility of Google Scholar is a potential new tool for Bibliometrics (Kousha & Thelwall, 2007). Google Scholar, is a free-of-charge by the giant Google search engine, has been suggested as an alternative or complementary resource to the commercial citation databases like Web of Knowledge (ISI/Thomson) or Scopus (Elsevier) (Aguillo, 2011). Google Scholar provides Bibliometrics information on a wide range of scholarly journals, and other published material, such as peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations (Orduña-Malea & Delgado López-Cózar, 2014). GS also introduced two new services in recent years: Google Scholar Author Citation Tracker in 2011 and Google Scholar Metrics for Publications in April 2012 (Jasco, 2012). Perhaps some of these documents would not otherwise be indexed by search engines such as Google, so they would be "invisible" to web searchers, and clearly some would be similarly invisible to Web of Science users, since it is dominated by academic journals (Kousha & Thelwall, 2007). On the other hand, the Thomson Reuters/Institute for Scientific Information databases (ISI) or Web of Science database (actually there is ambiguity between different names of former ISI), include three databases: Science Citation Index/Science Citation Index Expanded (SCI/SCIE) (SCIE is the online version of SCI), Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI) (Larsen & von Ins, 2010). Since 1964 the Science Citation Index (SCI) has been a leading tool in indexing (Garfield, 1972).

Few studies have been done to find a correlation between GS with WoS citations. Cabezas-Clavijo and Delgado-Lopez-Cozar (2013) found that the average h-index values in Google Scholar are almost 30% higher than those obtained in ISI Web of Science, and about 15% higher than those collected by Scopus. GS citation data differed greatly from the findings using citations from the fee-based databases such as ISI Web of Science (Bornmann et al., 2009). Google Scholar overestimates the number of citable articles (in comparison with formal citation services such as Scopus and Thomson Reuters) because of the automated way it collects data, including 'grey' literature such as theses (Hooper, 2012). The first objective of this study is to find the correlation between Google Scholar and ISI citation in the highly cited papers.

3. Visibility and Citation Impact

Nader Ale Ebrahim et al. (2014) based on a case study confirmed that the article visibility will greatly improve the citation impact. The journal visibility has an important influence on the journal citation impact (Yue & Wilson, 2004). Therefore, greater visibility caused higher citation impact (Zheng et al., 2012). In contrast, lack of visibility has condensed a significant citation impact (Rotich & Musakali, 2013). Nader Ale Ebrahim et al. (2013) by reviewing the relevant papers extracts 33 different ways for increasing the citations possibilities. The results show that the article visibility has tended to receive more download and citations. In order to improve the visibility of scholars' works and make them relevant on the academic scene, electronic publishing will be advisable. This provides the potential to readers to search and locate the articles at minimum time within one journal or across multiple journals. This includes publishing articles in journals that are reputable and listed in various databases and peer reviewed (Rotich & Musakali, 2013). Free online availability substantially increases a paper's impact (Lawrence, 2001a). Lawrence (2001a, 2001b) demonstrated a correlation between the likelihood of online availability of the full-text article and the total number of citations. He further showed that the relative citation counts for articles available online are on average 336% higher than those for articles not found online (Craig, Plume, McVeigh, Pringle, & Amin, 2007).

However, there are limited resources to explain the relationship between the paper availability and the number of citations (Craig et al., 2007; Lawrence, 2001b; McCabe & Snyder, 2013; Solomon, Laakso, & Björk, 2013).

None of them discussed about the relationship between the number of versions, and citation. The number of “versions” will be shown in any Google Scholar search result. Figure 1 shows 34 different versions of an article entitled “Virtual Teams: a Literature Review (Nader Ale Ebrahim, Ahmed, & Taha, 2009)” and number of citations. The second objective of this research is to find a relationship between the paper availability and the number of citations.

Virtual Teams: a Literature Review.

[NA Ebrahim, S Ahmed, Z Taha - Australian Journal of Basic ... , 2009 - search.ebscohost.com](#)

Abstract In the competitive market, virtual teams represent a growing response to the need for fast time-to-market, low-cost and rapid solutions to complex organizational problems.

Virtual teams enable organizations to pool the talents and expertise of employees and ...

Cited by 95 Related articles All 34 versions Cite Saved

Figure 1. The number of “versions” in the Google Scholar search result

4. Methodology

Highly cited papers from Malaysia in the field of engineering were retrieved from the Essential Science Indicators (ESI) which is one the Web of Knowledge (WoK) databases. ESI provides access to a comprehensive compilation of scientists’ performance statistics and science trend data derived from WoK Thomson Reuters databases. Total citation counts and cites per paper are indicators of influence and impact of each paper. There is a threshold to select highly cited papers according to the baseline data in ESI. This threshold is different from one discipline to another one. ESI rankings are determined for the most cited authors, institutions, countries, and journals (The Thomson Corporation, 2013). The paper must be published within the last 10-year plus four-month period (January 1, 2003-April 30, 2013) and must be cited above threshold level, in order to be selected. Essential Science Indicators data used in this research have been updated as of July 1, 2013.

Google Scholar which is a free online database was used for deriving the number of citations and versions of the ESI highly cited papers. The data from ESI was collected on 29 July 2013 and Google Scholar data was collected on 31 July 2013. The total numbers of 101 papers were listed in ESI as highly cited papers from Malaysia in the field of engineering. The lists of 101 papers were retrieved from ESI database and then were exported to an Excel sheet. A search engine was developed to get the number of citations and versions from Google Scholar. This gadget assisted the present researchers to collect the data more precisely and faster than searching for the papers one by one. The Statistical Package for the Social Sciences (SPSS) was used for analyzing the data. The results are illustrated in the following section.

5. Results and Discussion

The number of citations which were derived from Web of Knowledge platform hereafter are called ISI citation. To study the relationship among the number of citations in Google scholar and ISI and the number of versions, correlation coefficients were computed.

Table 1 shows descriptive statistics of the variables.

Table 1. Descriptive statistic of variables

	N	Minimum	Maximum	Mean	Std. Deviation
Versions	101	2	28	5.62	3.078
Cited in Google Scholar	101	4	348	80.76	71.718
ISI citation	101	5	189	43.15	36.076

As both numbers of citations in Google scholar and ISI were distributed normally, Pearson correlation coefficient (r) was used and the results showed a very high positive and significant association ($r = 0.932$, $P < 0.01$) between the number of citations in Google scholar and ISI for the articles that were published during 2006 to 2012 from Malaysia in the field of engineering. To study the relationship between both citation and the number of versions,

Spearman Rho was used due to the non-normal distribution of the versions. The results showed a significant and positive relationship between both citations in Google Scholar and ISI with the number of versions. This relationship was higher between versions and ISI citations ($r = 0.395, p < 0.01$) than between versions and Google Scholar citations ($r = 0.315, p < 0.01$). Linear regression was also applied to predict the number of citations in ISI based on Google Scholar citations. The results showed a very high predictability ($R^2 = 0.836$) for the linear model (see

Figure 2) which was significant ($F = 511.63, p < 0.01$). Therefore, the final equation for estimating the citation in ISI based on Google Scholar is:

$$\text{ISI Citation} = 5.961 + 0.460 (\text{Google Scholar citation})$$

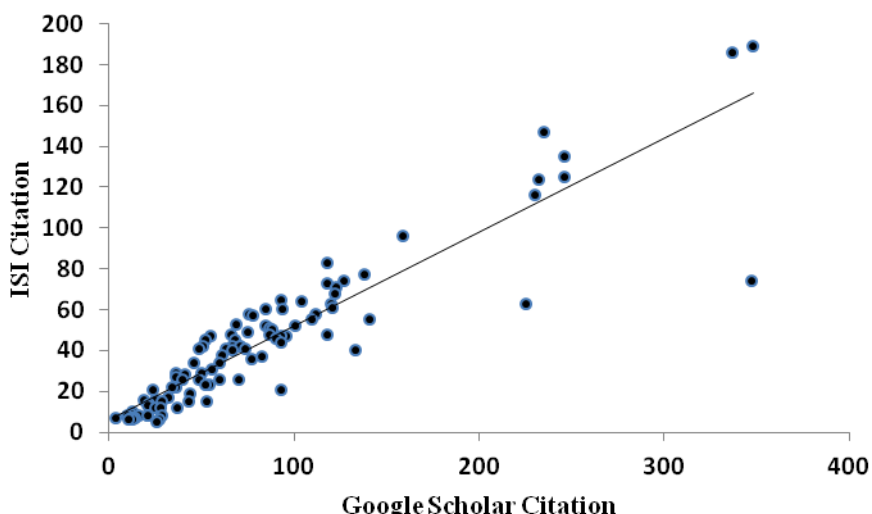


Figure 2. Scatter diagram between ISI citation and Google Scholar citation

To study the effect of the number of versions on both citations in Google Scholar and ISI, simple linear regression was applied. The results indicated that the number of versions had a significant positive effect on citations in both databases (see Table 2 and Table 3).

Table 2. Summary of regression analysis results

	R Square	F	β	t	p
Model^a	0.276	39.12**	0.532	6.255	<0.01
Model^b	0.272	38.316**	0.528	6.19	<0.01

Predictors: Versions

a: Dependent Variable: Cited in Google Scholar, b: Dependent Variable: ISI citation

Table 3. Descriptive statistics of variables - Year

Year	N	Versions		Cited in Google Scholar		ISI citation	
		Mean	SD	Mean	SD	Mean	SD
Before 2009	20	7.75	5.25	152.85	103.741	79.8	46.6
2009	26	6.08	1.695	101.19	38.948	56.96	20.577
2010	18	5.11	2.193	70.17	50.097	41.44	26.86

2011	16	4.31	1.352	49.25	33.66	21.31	12.015
2012	21	4.48	2.089	19.9	9.518	9.24	3.315

A comparison between Google Scholar and ISI citation for highly cited papers from Malaysia in the field of engineering (see Figure 3) shows that the citation counts in Google Scholar are always higher than the number of citations in ISI.

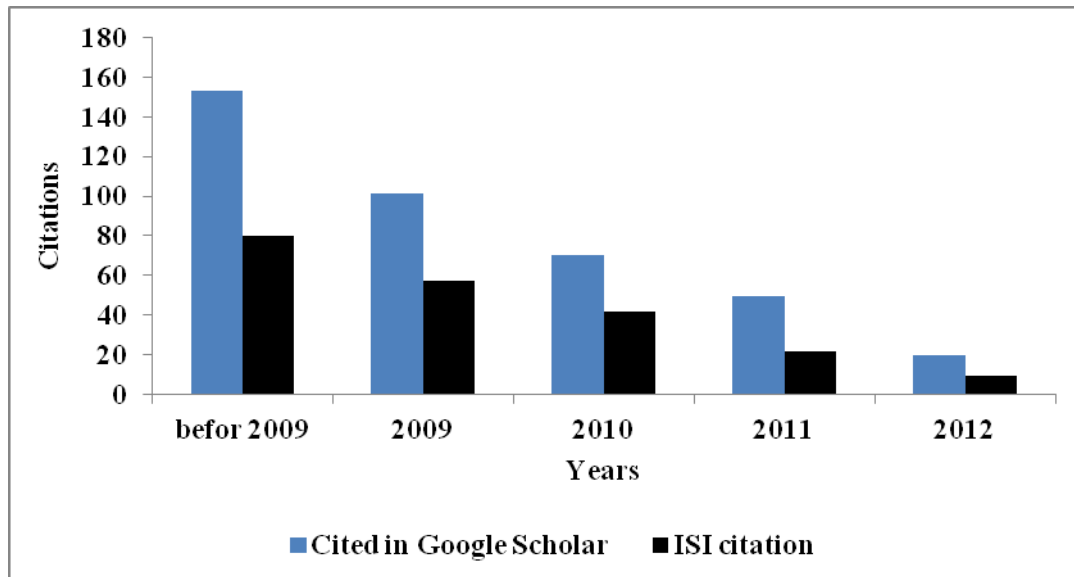


Figure 3. Comparison between Google Scholar and ISI citations

6. Conclusion

The number of publications and the number of citations in ISI Web of Science are used to measure the researchers' scientific performance and their research impact. However, these numbers are not freely available. Therefore, the offered equation can be used as a reference to convert the number of Google Scholar citations to ISI citations. On the other hand, the number of versions of both systems has a significant positive effect on the number of citations. This finding supports other researchers' (Amancio et al., 2012; Antelman, 2004; Egghe et al., 2013; Ertürk & Şengül, 2012; Hardy et al., 2005) findings related to the paper visibility. The results of this study indicate that there is a strong correlation between the number of citations in Google Scholar and ISI Web of Science. Therefore, the researchers can increase the impact of their research by increasing the visibility of their research papers (or paper versions). Future study is needed to determine the relationship between citation counts on the other databases such as Microsoft Academic Research, Scopus, SiteSeer index and ISI by considering journal article and conference papers.

Reference

- Aghaei Chadegani, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. *Asian Social Science*, 9(5), 18-26. <http://dx.doi.org/10.5539/ass.v9n5p18>
- Aguillo, I. F. (2011). Is Google Scholar useful for Bibliometrics? A Webometric Analysis. In E. Noyons, P. Ngulube, & J. Leta (Eds.), *Proceedings of Issi 2011: The 13th Conference of the International Society for Scientometrics and Informetrics, Vols 1 and 2* (pp. 19-25). Leuven: Int Soc Scientometrics & Informetrics-Issi.
- Ale Ebrahim, N. (2013). Introduction to the Research Tools Mind Map. *Research World*, 10(4), 1-3. <http://dx.doi.org/10.5281/zenodo.7712>
- Ale Ebrahim, N., Ahmed, S., & Taha, Z. (2009). Virtual Teams: a Literature Review. *Australian Journal of Basic and Applied Sciences*, 3(3), 2653-2669. <http://dx.doi.org/10.6084/m9.figshare.1067906>
- Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi Tanha, F., Gholizadeh, H., & Motahar, S. M. (2014). Visibility

- and Citation Impact. *International Education Studies*, 7(4), 120-125. <http://dx.doi.org/10.5539/ies.v7n4p120>
- Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi Tanha, F., Gholizadeh, H., Motahar, S. M., & Ordi, A. (2013). Effective Strategies for Increasing Citation Frequency. *International Education Studies*, 6(11), 93-99. <http://dx.doi.org/10.5539/ies.v6n11p93>
- Amancio, D. R., Oliveira Jr, O. N., & da Fontoura Costa, L. (2012). Three-feature model to reproduce the topology of citation networks and the effects from authors' visibility on their h-index. *Journal of Informetrics*, 6(3), 427-434. <http://dx.doi.org/10.1016/j.joi.2012.02.005>
- Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, 65(5), 372-382. <http://dx.doi.org/10.5860/crl.65.5.372>
- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3(1), 7. <http://dx.doi.org/10.1186/1742-5581-3-7>
- Bensman, S. (2011). Anne-Wil Harzing: The publish or perish book: Your guide to effective and responsible citation analysis. *Scientometrics*, 88(1), 339-342. <http://dx.doi.org/10.1007/s11192-011-0388-8>
- Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & Daniel, H. D. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry-Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics*, 3(1), 27-35. <http://dx.doi.org/10.1016/j.joi.2008.11.r001>
- Cabezas-Clavijo, A., & Delgado-Lopez-Cozar, E. (2013). Google Scholar and the h-index in biomedicine: The popularization of bibliometric assessment. *Medicina Intensiva*, 37(5), 343-354. <http://dx.doi.org/10.1016/j.medin.2013.01.008>
- Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?: A critical review of the literature. *Journal of Informetrics*, 1(3), 239-248. <http://dx.doi.org/10.1016/j.joi.2007.04.001>
- Egghe, L., Guns, R., & Rousseau, R. (2013). Measuring co-authors' contribution to an article's visibility. *Scientometrics*, 95(1), 55-67. <http://dx.doi.org/10.1007/s11192-012-0832-4>
- Ertürk, K., & Şengül, G. (2012). Self Archiving in Atılım University. In S. Kurbanoglu, U. Al, P. Erdoğan, Y. Tonta, & N. Uçak (Eds.), *E-Science and Information Management* (Vol. 317, pp. 79-86): Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-33299-9_11
- Fooladi, M., Salehi, H., Yunus, M. M., Farhadi, M., Aghaei Chadegani, A., Farhadi, H., & Ale Ebrahim, N. (2013). Do Criticisms Overcome the Praises of Journal Impact Factor? *Asian Social Science*, 9(5), 176-182. <http://dx.doi.org/10.5539/ass.v9n5p176>
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178, 471-479. <http://dx.doi.org/10.1126/science.178.4060.471>
- Hardy, R., Oppenheim, C., Brody, T., & Hitchcock, S. (2005). Open Access Citation Information.
- Hooper, S. L. (2012). Citations: not all measures are equal. *Nature*, 483(7387), 36-36. <http://dx.doi.org/10.1038/483036c>
- Jacso, P. (2012). Google Scholar Metrics for Publications The software and content features of a new open access bibliometric service. *Online Information Review*, 36(4), 604-619. <http://dx.doi.org/10.1108/14684521211254121>
- Kear, R., & Colbert-Lewis, D. (2011). Citation searching and bibliometric measures: Resources for ranking and tracking. *College & Research Libraries News*, 72(8), 470-474.
- Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web-URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065. <http://dx.doi.org/10.1002/asi.v58:7>
- Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575-603. <http://dx.doi.org/10.1007/s11192-010-0202-z>
- Lawrence, S. (2001a). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521-521. <http://dx.doi.org/10.1038/35079151>

- Lawrence, S. (2001b). Online or invisible. *Nature*, 411(6837), 521. <http://dx.doi.org/10.1038/35079151>
- McCabe, M. J., & Snyder, C. M. (2013). Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals: SSRN working paper.
- Orduña-Malea, E., & Delgado López-Cózar, E. (2014). Google Scholar Metrics evolution: an analysis according to languages. *Scientometrics*, 98(3), 2353-2367. <http://dx.doi.org/10.1007/s11192-013-1164-8>
- Pauly, D., & Stergiou, K. I. (2005). Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar service. *Ethics in Science and Environmental Politics*, 5, 33-35.
- Rotich, D. C., & Musakali, J. J. (2013). *Publish or Perish: Remaining Academically Relevant and Visible In the Global Academic Scene through Scholarly Publishing*. Paper presented at the Conference and Programme Chairs, South Africa.
- Solomon, D. J., Laakso, M., & Björk, B.-C. (2013). A longitudinal comparison of citation rates and growth among open access journals. *Journal of Informetrics*, 7(3), 642-650. <http://dx.doi.org/10.1007/s11192-013-1164-8>
- The Thomson Corporation. (2013). Essential Science Indicators, Product Overview. Retrieved from http://esi.webofknowledge.com/help/h_whatism.htm
- Yue, W. P., & Wilson, C. S. (2004). Measuring the citation impact of research journals in clinical neurology: A structural equation modelling analysis. *Scientometrics*, 60(3), 317-332. <http://dx.doi.org/10.1023/B:SCIE.0000034377.93437.18>
- Zhang, C.-T. (2009). The e-Index, Complementing the h-Index for Excess Citations. *PLoS ONE*, 4(5), e5429. <http://dx.doi.org/10.1371/journal.pone.0005429>
- Zheng, J., Zhao, Z. Y., Zhang, X., Chen, D. Z., Huang, M. H., Lei, X. P., ... Zhao, Y. H. (2012). International scientific and technological collaboration of China from 2004 to 2008: a perspective from paper and patent analysis. *Scientometrics* 91(1), 65-80. <http://dx.doi.org/10.1007/s11192-011-0529-0>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).