Qual Quant (2015) 49:1513–1530
DOI 10.1007/s11135-014-0110-z

Factors affecting citation networks in science and technology: focused on non-quality factors

Minho So · Jiyoung Kim · Sangki Choi · Han Woo Park

Published online: 16 October 2014

Abstract This paper analyses citations of journal articles to find out key bibliographic factors that have an effect on citation. It helps to understand the degree of measurable factors which affects to impact factor by scale and correlation analysis between each factors. It ensures moderating variables work to increase impact factor as a moderating effect. Five factors can be measured—the number of authors, the number of affiliations, the number of references, the number of pages, and the number of keywords—were selected. We found that five bibliographic factors have positive effects on citation. All of common bibliographic items, such as subject, nation, year showed general effects in the analysis to examine if the moderating variables increase or decrease the effects on bibliographic factors.

Keywords Bibliometrics · Citation analysis · Factors affecting article citations · Moderating effect · Academic journal policy

1 Introduction

Objective evaluations of science and technology (S&T) research activities and achievements have played a crucial role in policymaking as well as decision-making processes in science

M. So · S. Choi
Department of Library and Information Science, Chonbuk National University, Jeonju-si, South Korea
e-mail: minhoso@kaist.ac.kr

S. Choi
e-mail: choisk@jbnu.ac.kr

J. Kim · H. W. Park (✉)
Department of Media and Communication, Yeungnam University, Gyeongsan-si, South Korea
e-mail: hanpark@ynu.ac.kr

J. Kim
e-mail: Jiyoungkim999@yu.ac.kr

policy. In such evaluations, the number of papers published and that of citations received represent two of the most important indices. As important achievements of researchers and as valuable sources of data, existing research papers are cited as data in new papers. In general, the larger the number of citations a paper receives, the higher its assumed quality is. Therefore, the number of papers and that of citations for each country in S&T may be useful indicators of the level of S&T; crucial data for reviewing policymaking; and effective tools for evaluating S&T investment and research activities (Marks 2001, p. 91).

Paper citations are verified by their frequency in a given journal by other authors (Garfield and Merton 1979, p. 24). A simple but powerful ranking tool has been designed based on this frequency, and the impact factor for each specific topic category is calculated using the tool to rank journals. For example, in determining citation frequency, the larger the impact of a study, higher its assumed quality (Van Leeuwen et al. 2003, pp. 262–263). In general, citation frequency in academic research is strongly correlated with academic success. Simple analyses by Ashton and Oppenheim (1978), Garfield (1992), and Opthof (1997) concerning the citation frequency of papers by Nobel Prize winners provided support for the correlation between these papers and the number of citations received.

In this regard, previous studies have identified various factors that influence the number of citations received by a paper from other papers but have been limited in terms of the consideration of paper topics or the development of indices for measuring leverage. In addition, few studies have analyzed the factors that can increase the number of incoming citations based on large data by country. For citation frequency diverse factors may be relevant in addition to paper quality, which has generally been assumed to have considerable influences. Considering other variables that are quantitatively measurable may be an important and necessary task for strengthening the influence of journals and researchers. This study examined the number of citation of journal articles to find out key bibliographic factors that have an effect on citation, to understand the degree of measurable variables that affect impact factor through the scale analysis and correlation analysis between the factors, and to ensure that moderating variables work on increasing impact factor as a moderating effect.

2 Literature review

2.1 Bibliometrics as knowledge diffusion network

Bibliometrics is a set of methods to quantitatively analyze academic literature. Citation analysis is considered as the knowledge diffusion network. Many research fields use bibliometric methods to explore the impact of their field, a set of researchers, or a particular paper. Kim (2012) examined the factors influencing the speed and frequency of citations in scientific papers through a bibliometric analysis using 21,443 data points from the NCR-Korea database and showed that papers written by foreign first authors or reprint authors were cited more quickly based on the speed of first citations or the average speed of citations; papers written in collaboration with foreign researchers were cited much more frequently than those written only by domestic researchers; and three bibliographic factors (the country of origin of the first author, the country of origin of the corresponding author, and the number of coauthors) had significant effects on the relationship between the speed and number of first citations. Norris (2008) examined advantages of open-access (OA) papers in citation and selected four topics, namely ecology, applied mathematics, sociology, and economics, to assess whether there would be an advantage for journal papers with an OA version on the Internet relative to those that were exclusively toll access (TA). He considered a total of 4,633 papers. Here

2,280 (49 %) were OA and had a mean citation count of 9.04, whereas the mean for TA papers was 5.76.

There are studies about collaboration aspect on citation. Levitt and Thelwall (2009) investigated and studied whether carrying out research in collaboration should end up with being of a high quality of results while an analysis on research in collaboration was considered as an important end for building a research policy. They extracted data on Library and Information Science from 1976 to 2004. Papers differentiated into five levels according to collaboration portions, and were compared. In addition research works done by 35 influential information specialists have been analyzed. It was verified that there was considerable relationship between research in collaboration and citation frequency. Rousseau (2000) studied on advantages that papers of co-authors had in citation over single-authors and revealed that it is not always true that papers of co-authors would be cited more than those of single-authors, and insisted that the number of citations was not proportionate to the number of co-authors either. Papers of co-authors in some fields were cited more than of single-authors, however, that did not take place in all the science areas.

Turpin et al. (2011) traced changes in institutional and social dynamics underlying cross-sector R&D collaboration in Australia and found that public policy provided the initial push toward cross-sector collaboration. According to their research, the Cooperative Research Centres (CRCs) program is Australia's oldest national arrangement for university–industry–government research collaboration. In the last two decades, the program has grown to become a dominant model for cross-sector R&D cooperation in the country. Because of the size of the program in the Australian innovation system, it has also become a major debate topic about science policies. Universities have institutionalised this imperative in diverse ways to steer research funding and career opportunities for their academic staff, and expectations and aspirations of CRC staff members, doctoral students, and potential staff members and students have become deeply embedded in the CRC's evolutionary processes.

2.2 Assessment of scientific knowledge with big data

Bibliometrics approaches have been increasingly used to assess the evolution and structure of scientific knowledge and R&D output. Wuchty et al. (2007) examined the increasing dominance of teams in knowledge production by considering 19.9 million papers over a five-decade period from 1955 to 2000 and 2.1 million patents and found that teams increasingly dominated solo authors in knowledge production. In addition, research projects were increasingly done in teams across nearly all fields. Teams typically produced more frequently cited research than individuals did, and this advantage increased over time. In addition, teams also produced exceptionally high-impact research. They detailed these results for sciences and engineering, social sciences, arts and humanities, and patents, suggesting a fundamental change in the process of knowledge creation. Lee (2013) identified the key factors influencing the success of multidisciplinary and trans-disciplinary R&D projects in transition economies using big data. Integrating knowledge management is an integrative approach to R&D, followed by providing the major sub-constructs of team science and policy implications to better facilitate multidisciplinary, interdisciplinary, and transdisciplinary R&D projects in knowledge-based transition economies.

The interdisciplinary science of team science promotes team-based research through an empirical examination of processes by which scientific teams organize, communicate, and conduct research. This field focuses on understanding and managing circumstances that facilitate or hinder the effectiveness of large-scale collaborative research, training, and transnational initiatives. This includes understanding how teams connect and collaborate to achieve scientific breakthroughs that would otherwise be unattainable by either individual or simply additive efforts. Yoo and Chung (2010) examined the factors influencing patent citations by conducting a correlation analysis of the number of patent citations and 17 explanatory variables (morphological, technological, and conceptual factors) based on a test data set of US patents in five research fields and found that 7 of these 17 variables had an impact of 5 % or greater: the number of patent citations, the number of pages, the number of claims, the ratio of references to average citations, the ratio of increases in patents to decreases, the strength of bibliographic couplings, the number of co-citations, and document similarity.

2.3 Research questions and hypotheses

This study is guided by the following two research questions: first, which bibliographic factors, aside from paper quality, can be assumed to influence incoming citations of S&T papers? Second, is there any causal relationship between incoming citations and the researcher's language (native language)? In this regard, the following hypotheses are proposed:

- *H1-1* the number of coauthors has a positive correlation with the likelihood of being cited.
- *H1-2* the number of collaborating institutions has a positive correlation with the likelihood of being cited.
- *H1-3* the number of references has a positive correlation with the likelihood of being cited.
- *H1-4* the number of pages has a positive correlation with the likelihood of being cited.
- *H1-5* the number of keywords has a positive correlation with the likelihood of being cited.
- *H2-1* the paper topic has moderating effects on relationships between factors influencing citations and the number of citations.
- *H2-2* the paper's country of origin has moderating effects on relationships between factors influencing citations and the number of citations.
- *H2-3* the year of a paper has moderating effects on relationships between factors influencing citations and the number of citations.

2.4 Research method

The sample of three countries for the citation analysis was selected by considering both native languages of researchers and regions in terms of the similar research activity in S&T field: Australia as an English-speaking-country in Oceania, Korea as a non-English-speaking country in Asia, and Switzerland as a multilingual country in Europe. SCI papers from three countries (Australia, South Korea, and Switzerland) published within the last 10 years were examined to investigate the effects of various factors (other than paper quality) on citations in the S&T field.

For this, bibliographic data were collected from the web version of SCI publications, the Web of Science, for papers published in the last 10 years in the three countries, and then the data were verified for three specific years 2006, 2008, and 2010 and three large S&T fields (natural science, life sciences, engineering). Usually the peak time of citation is 3 years after the publication, but it seems according to different disciplines (Moed 2005, 95, Fig. 5.2). There are over 50 % ratio of citation of more than once cited researches at least 2 years after the paper published according to the Korea Advanced Institute of Science and Technology (So 2012). Thus, we selected 2006, 2008 and 2010 for the empirical data from the data extraction period of 10 years, from 2001 to 2010 at the 2012 data collection point.

Factors influencing citations were classified into five types that were measurable according to quantitative fluctuations based on bibliographic information from papers. For bibliographic information on authors, the number of coauthors and that of co-affiliations were selected, and numbers of references, pages, and keywords were adopted from the bibliographic information as five factors for the citation analysis.

To test the hypotheses, a frequency analysis was conducted through SPSS 15.0. Descriptive statistics were used to examine the variables, and the Pearson correlation coefficient was employed for a correlation analysis. Finally, an ANOVA was conducted to determine any significant differences between variables, and a regression analysis was conducted to examine the impact factors.

2.5 Variables

Table 1 shows the factors investigated in this study. Various factors considered in previous studies for analyzing citations were classified into five categories, including author attributes, institutional attributes, research attributes, journal characteristics, and methods. A total of 36 factors included 10 factors for author attributes, including the country of origin, 4 factors for institutional attribute, including reputation, 9 factors for research attributes, including language, 7 factors for journal characteristics, including reputation, and 6 factors for methods, including test methods.

All factors were evaluated to select quantitative ones extractable from the Web of Science, the target database for extraction, and measurable in terms of quantitative fluctuations. As a result, numbers of coauthors and co-affiliations for author attributes were determined, and numbers of references, pages, and keywords were adopted as research attributes. These five factors were determined as the target of analysis for increasing the number of incoming citations and as an independent variable. In addition, the number of citations was selected as a dependent variable.

With a moderating variable put into the causal relationship between independent and dependent variables, it comes to have a moderating effect once the independent variable is affected. To test the moderating effects on relationships between various factors and the number of citations, the country of origin of the author was selected, and the topic and publication year of the paper were selected as journal characteristics.

In this study, a total of 45,822 papers in the S&T field (including natural sciences, life sciences, and engineering) with a cumulative rate of 80 % or higher in terms of the number of citations in three specific years (2006, 2008, and 2010) in Australia, South Korea, and Switzerland were used. These 45,822 papers served as target data for the impact factor analysis, and 126,590 papers were selected initially in the S&T field out of a total of 218,316 papers published in these three countries. Then those papers with a cumulative rate of 80 % or higher in terms of the number of incoming citations and published in 2006, 2008, and 2010 in Australia, South Korea, and Switzerland were selected to help discriminate data required for hypothesis testing and statistical analysis and ensure fairness in the number of papers for each topic (Table 2).

3 Results

Table 3 shows a frequency analysis by subject, country, and year. In terms of topics, natural sciences accounted for 45.71 % of the 45,822 papers, followed by life sciences (31.44 %) and engineering (22.85 %). The most common country was South Korea (45.67 %), followed

Table 1 Factors in the bibliometric analysis

Factors	Web of Science (available factors)	Quantitative measurement	Factor selection
Author attribute			
Country	O	X	⋆
Department	O	X	
Author	O	X	
Number of coauthors	O	O	•
Number of co-affiliations	O	O	•
Gender	X	X	
Age	X	O	
Career	X	O	
Reputation	X	X	
Number of previous studies	X	O	
Affiliation attribute			
Reputation	X	X	
Collaboration	O	X	
International cooperation	O	X	
Funding	△	X	
Research attribute			
Language	O	X	
Format	O	X	
Novelty of the title	X	X	
Number of references	O	O	•
Novelty of references	O	X	
Number of figures and tables	X	O	
Number of pages	O	O	•
Number of keywords	O	O	•
Number of citations	O	O	◎
Journal characteristics			
Reputation (impact factor)	△	O	
Subject areas	O	X	⋆
Year published	O	X	⋆
Cited half-life	△	O	
Citation peak	△	O	
Immediacy index	△	O	
Self-citation rate	△	O	
Methods			
Experimental method	X	X	
Goodness	X	X	
Bibliographic coupling	X	O	
Co-citation	X	O	
Similarity between documents	X	O	
Application level	X	X	

Possible (O), impossible (X), and partially possible or a possible coupling with another database (△)

In factor selection, ◎ is selected as a dependent variable, • as variable (impact factor), and ⋆ to test the moderating effect

Table 2 Overview of target data for data design

Years	Countries	Total number of papers ⓐ	Number of papers on engineering, life sciences, and physical sciences ⓑ	Number of papers for top 80 % of all citations ⓒ	ⓒ/ⓑ (%)	ⓑ/ⓐ (%)
2006	Australia	21,808	10,592	4,053	38.26	48.57
	Korea	22,492	16,159	5,725	35.43	71.84
	Switzerland	14,163	8,343	3,244	38.88	58.91
2008	Australia	26,673	12,596	4,922	39.08	47.22
	Korea	30,289	20,777	7,394	35.59	68.60
	Switzerland	15,465	8,610	3,401	39.50	55.67
2010	Australia	30,884	14,369	5,386	37.48	46.53
	Korea	37,982	24,910	7,810	31.35	65.58
	Switzerland	18,560	10,234	3,887	37.98	55.14
Sum		218,316	126,590	45,822	36.20	57.98

Table 3 Frequency analysis

$N = 45,822$

Classification	Frequency	%
Subject		
Natural sciences	20,943	45.71
Life sciences	14,407	31.44
Engineering	10,472	22.85
Countries		
Australia	14,361	31.34
Korea	20,929	45.67
Switzerland	10,532	22.99
Years		
2006	13,022	28.42
2008	15,717	34.30
2010	17,083	37.28

by Australia (31.34 %) and Switzerland (22.99 %). The year 2010 accounted for 37.28 %, followed by 2008 (34.30 %) and 2006 (28.42 %).

Table 4 shows the descriptive statistics (minimums, maximums, averages, and standard deviations) for numbers of coauthors, co-affiliations, references, pages, keywords, and citations. The average number of references was 37.22; that of citations, 22.53; that of coauthors, 11.72; that of pages, 9.46; that of keywords, 7.23; and that of co-affiliations, 3.75. Standard deviations were used to measure changes in variables. The highest standard deviation was found for coauthors (50.81), followed by citations (34.38), references (21.76), co-affiliations (8.11), pages (6.26), and keywords (3.38).

Table 5 shows the Pearson correlation coefficient for the analysis of correlations between numbers of coauthors, co-affiliations, references, pages, keywords, and citations. The results of the correlation analysis indicate significant correlations between all these variables

Table 4 Descriptive analysis

Variables	Minimum	Maximum	Mean	SD
#Coauthors	1.00	967.00	11.72	50.81
#Co-affiliations	1.00	155.00	3.75	8.11
#References	1.00	782.00	37.22	21.76
#Pages	1.00	165.00	9.46	6.26
#Keywords	0.00	10.00	7.23	3.38
#Citations	2.00	2328.00	22.53	34.38

Table 5 Correlation analysis

Variables	Number of citations	Number of co-affiliations	Number of references	Number of pages	Number of keywords	Number of citations
#Coauthors	1					
	(.)					
#Co-affiliations	0.899	1				
	(0***)	(.)				
#References	0.003	0.048	1			
	(0.579)	(0***)	(.)			
#Pages	0.066	0.099	0.451	1		
	(0***)	(0***)	(0***)	(.)		
#Keywords	−0.062	−0.032	0.563	0.158	1	
	(0***)	(0***)	(0***)	(0***)	(.)	
#Citations	0.081	0.120	0.085	0.031	0.084	1
	(0***)	(0***)	(0***)	(0***)	(0***)	(.)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

except for the number of authors and that of references. There were significant negative (−) correlations between the number of keywords and the number of coauthors ($r = -0.062$, $p < 0.001$) and between the number of keywords and the number of co-affiliations ($r = -0.032$, $p < 0.001$) and significant positive (+) correlations between other factors. There were relatively strong correlations between the number of co-affiliations and the number of coauthors ($r = 0.899$), between the number of pages and the number of references ($r = 0.451$), and between the number of keywords and the number of references ($r = 0.563$). The number of citations had the highest correlation with the number of co-affiliations ($r = 0.120$), followed by the number of references ($r = 0.085$), the number of keywords (0.084), and the number of pages ($r = 0.031$).

Table 6 shows the ANOVA results by topic, county, and year. The number of coauthors, the number of affiliations, the number of references, the number of pages, and the number citations showed significant differences. The largest number of coauthors was found in natural sciences (17.80), followed by life sciences (7.86) and engineering (4.89). The largest number of co-affiliations was found in natural sciences (4.78), followed by life sciences (3.37) and engineering (2.21). The average number of references in life sciences was 42.54; that in natural sciences, 38.86; and that in engineering, 26.61. The average number of pages in engineering was 10.16; that in life sciences, 9.43; and that in natural sciences, 9.12. The largest number of keywords was found in natural sciences (9.04), followed by life sciences (7.31) and engineering (4.58). Life sciences and natural sciences showed similar average

Table 6 ANOVA test: topics

Variables	Countries	Mean	SD	F-value	p-value
#Coauthors	Natural sciences	17.80	73.07	289.617	0.000***
	Life sciences	7.86	10.28		
	Engineering	4.89	18.14		
#Co-affiliations	Natural sciences	4.78	11.04	380.747	0.000***
	Life sciences	3.37	4.77		
	Engineering	2.21	2.89		
#References	Natural sciences	38.86	23.46	1,877.54	0.000***
	Life sciences	42.54	20.08		
	Engineering	26.61	16.11		
#Pages	Natural sciences	9.12	7.33	98.348	0.000***
	Life sciences	9.43	3.58		
	Engineering	10.16	6.74		
#Keywords	Natural sciences	7.32	3.17	6,903.475	0.000***
	Life sciences	9.04	2.19		
	Engineering	4.58	3.42		
#Citations	Natural sciences	24.49	28.75	559.547	0.000***
	Life sciences	26.66	45.69		
	Engineering	12.93	22.93		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

numbers of citations (26.66 and 24.49, respectively), and engineering, a relatively low average (12.93).

Table 7 shows the results of the F-test for a variance analysis to check for any differences in research factors between the countries. There were significant differences in all factors, including numbers of coauthors, co-affiliations, references, pages, keywords, and citations.

Switzerland showed the largest number of coauthors (15.81), followed by Korea (12.21) and Australia (8.02). Switzerland also showed the largest number of co-affiliations (4.87), followed by Korea (3.50) and Australia (3.36). Switzerland and Australia showed similar results in the number of references (43.57 and 41.65, respectively), followed by Korea (30.97). Switzerland and Australia showed similar results for the number of references (10.78 and 10.49, respectively), followed by Korea (30.97). The number of keywords showed a similar pattern. Switzerland and Australia showed similar results (7.84 and 7.72, respectively), followed by Korea (6.59). Switzerland showed the largest number of citations (45.86), followed by Australia (24.08) and Korea (16.43).

In terms of the publication year, there were no significant differences in numbers of co-affiliations, references, pages, keywords, and citations, but there was a significant difference in the number of coauthors (Table 8). In general, the more recent the publication year, the larger the number was. The average number of co-affiliations in 2006 was 3.42, followed by 2008 (3.71) and 2010 (4.04). The average number of references in 2006 was 34.90, followed by 2008 (35.93) and 2010 (40.16). The average number of keywords in 2006 was 7.10, followed by 2008 (7.14) and 2010 (7.43). The year 2006 showed the largest number of pages (9.79), followed by 2010 (9.67) and 2008 (9.30). The average number of citations in 2006 was 35.30, followed by 2008 (23.80) and 2010 (11.64). This may be explained by the cumulative

Table 7 ANOVA test: countries

Variables	Countries	Mean	SD	F-value	p-value
#Coauthors	Australia	8.02	26.53	73.471	0.000***
	Korea	12.21	56.60		
	Switzerland	15.81	62.20		
#Co-affiliations	Australia	3.50	6.15	132.557	0.000***
	Korea	3.36	8.40		
	Switzerland	4.87	9.64		
#References	Australia	41.65	23.55	1,729.409	0.000***
	Korea	30.97	17.44		
	Switzerland	43.57	23.56		
#Pages	Australia	10.49	6.32	980.452	0.000***
	Korea	8.08	4.76		
	Switzerland	10.78	7.99		
#Keywords	Australia	7.72	3.20	721.745	0.000***
	Korea	6.59	3.51		
	Switzerland	7.84	3.13		
#Citations	Australia	24.08	37.68	818.207	0.000***
	Korea	16.43	21.50		
	Switzerland	32.54	45.96		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

nature of citations starting from the publication year to the point of data collection such that the more recent the publication year, the smaller the number of citations.

Table 9 shows the results of a regression analysis of factors influencing the number of incoming citations. The number of coauthors ($B = -0.083$, $p < 0.001$) and that of pages ($B = -0.079$, $p < 0.001$) had significant negative (−) effects on the number of citations, whereas the numbers of co-affiliations ($B = 0.979$, $p < 0.001$), that of references ($B = 0.074$, $p < 0.001$), and that keywords ($B = 0.612$, $p < 0.001$) had significant positive (+) effects. That is, with other independent variables taken into consideration, the number of citations increased significantly with an increase in the number of affiliations, the number of references, or the number of keywords, whereas it decreased significantly with an increase in the number of authors or the number of pages.

The number of co-affiliations had the greatest effect on the number of citations ($\beta = 0.231$), followed by the number of coauthors ($\beta = -0.122$), the number of keywords ($\beta = 0.060$), the number of references ($\beta = 0.047$), and the number of pages ($\beta = -0.014$).

To classify the paper topic, a moderating variable, it was converted into a dummy variable. More specifically, life sciences were set to $(0, 0)$; natural sciences, to $(1, 0)$; and engineering, to $(0, 1)$. If the factors affecting the number of citations, including numbers of coauthors, co-affiliations, references, pages, and keywords, are designated as X, then the models suggested in Baron and Kenny (1986) can be expressed as

Model 1: $Y = \beta_0 + \beta_1 X + \beta_{21} Z_1 + \beta_{22} Z_2,$
Model 2: $Y = \beta_0 + \beta_1 X + \beta_{21} Z_1 + \beta_{22} Z_2 + \beta_{31} X Z_1 + \beta_{32} X Z_2.$

Here, Z_1 is a dummy variable for natural sciences and Z_2, for engineering. Table 10 shows the results of a hierarchical regression analysis for these two models. If an interaction term

Table 8 ANOVA test: years

Variables	Years	Mean	SD	*F*-value	*p*-value
#Coauthors	2006	11.26	51.58	0.776	0.460
	2008	11.96	52.27		
	2010	11.86	48.81		
#Co-affiliations	2006	3.42	7.24	21.861	0.000***
	2008	3.71	8.10		
	2010	4.04	8.71		
#References	2006	34.90	18.50	259.817	0.000***
	2008	35.93	18.83		
	2010	40.16	25.86		
#Pages	2006	9.51	6.25	7.324	0.001***
	2008	9.30	6.07		
	2010	9.55	6.43		
#Keywords	2006	7.10	3.42	44.171	0.000***
	2008	7.14	3.44		
	2010	7.43	3.28		
#Citations	2006	35.30	48.64	1,912.824	0.000***
	2008	23.80	31.40		
	2010	11.64	14.73		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9 Factors influencing citations

IV	Nonstandard		Standard regression coefficients	*t*-value	*p*-value
	Regression coefficients	Standard error			
Constant	13.406	0.426		31.496	0.000
#Coauthors	−0.083	0.007	−0.122	−11.535	0.000***
#Co-affiliations	0.979	0.045	0.231	21.804	0.000***
#References	0.074	0.01	0.047	7.467	0.000***
#Page	−0.079	0.029	−0.014	−2.737	0.006***
#Keywords	0.612	0.057	0.060	10.673	0.000***
Model statistics	*R*	R^2	R^2	*F* value	*p* value
	0.163	0.026	0.026	248.853	0.000***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

is constructed for a hierarchical regression analysis, then there is a potential problem of multicollinearity, and therefore the average of independent and dependent variables was set to 0, and the variance was set as 1 in a standardizing manner.

As shown in Table 10, the results for Model 1 indicate $B = -0.070,\ t = -6.411\ (p < 0.001)$, for natural sciences and $B = -0.351,\ t = -24.073\ (p < 0.001)$, for engineering. In terms of the moderating effects of independent variables, a significant negative (−) moderating effect was found for #coauthors * Natural [$B = -0.976,\ t = -10.377\ (p < 0.001)$]; a

Table 10 Mediating effects: topics

IV	B-value	t-value	p-value	B-value	t-value	p-value
Constant	0.112	12.993	0.000	0.200	16.334	0.000
#Coauthors	−0.116	−11.061	0.000***	0.913	9.787	0.000
#Co-affiliations	0.211	19.996	0.000***	0.048	1.507	0.132
#References	0.033	5.268	0.000***	0.028	2.426	0.015
#Page	0.010	1.911	0.056	0.059	3.378	0.001
#Keywords	0.001	0.234	0.815	−0.028	−2.030	0.042
Natural	−0.070	−6.411	0.000***	−0.156	−11.097	0.000***
Engineering	−0.351	−24.073	0.000***	−0.410	−22.056	0.000***
#Coauthors * Natural				−0.976	−10.377	0.000***
#Co-affiliations * Natural				0.087	2.519	0.012*
#References * Natural				0.007	0.491	0.623
#Page * Natural				−0.063	−3.346	0.001***
#Keywords * Natural				0.026	1.609	0.108
#Coauthors * Engineering				−0.752	−6.786	0.000***
#Co-affiliations * Engineering				0.034	0.499	0.618
#References * Engineering				0.031	1.456	0.145
#Page * Engineering				−0.039	−1.936	0.053
#Keywords * Engineering				0.035	1.912	0.056
Model statistics	R^2	F-value	p-value	ΔR^2	F-value	p-value
	0.040	270.795	0.000***	0.008	37.278	0.000***

* $p < .05$, ** $p < .01$, *** $p < .001$

significant positive (+) moderating effect, for #co-affiliations * Natural [$B = 0.087$, $t = 2.519\,(p < 0.001)$]; a significant negative (−) moderating effect, for #pages * Natural [$B = -0.063$, $t = -3.346\,(p < .0001)$]; and a significant negative (−) moderating effect, for #coauthors * Engineering [$B = -0.752$, $t = -6.786\,(p < 0.001)$]. That is, for the number of coauthors, the effect of natural sciences was weaker than that of the other fields, and the effect of engineering was also weaker than that of the non-engineering fields. For the number of co-affiliations, natural sciences had a greater effect than the other fields, and the effect of natural sciences on the number of pages was weaker than that of the other fields. No significant differences were verified for #references * Natural, #keywords * Natural, #co-affiliations * Engineering, #pages * Engineering, and #keywords * Engineering.

In Model 1, the coefficient of determination (R^2), used to measure the fit of the model, was 0.040, indicating that explanatory power increased to 0.8 % in Model 2 because of the inclusion of interaction items. The moderating effect of the paper topic was significant ($F = 37.278$, $p < 0.001$).

To test the moderating effect by country, the variable was converted into a dummy variable, as in the case of topics. Switzerland was set as the base category (0, 0); Australia, as (1, 0); and Korea, as (0, 1). If the factors influencing citations (numbers of coauthors, co-affiliations, references, pages, and keywords) are designated as X, then the models suggested in Baron and Kenny (1986) can be expressed as

Model 1:$Y = \beta_0 + \beta_1 X + \beta_{21} Z_1 + \beta_{22} Z_2$,

Model 2:$Y = \beta_0 + \beta_1 X + \beta_{21} Z_1 + \beta_{22} Z_2 + \beta_{31} X Z_1 + \beta_{32} X Z_2$.

Table 11 Mediating effects: countries

IV	Model 1			Model 2		
	B-value	*t*-value	*p*-value	*B*-value	*t*-value	*p*-value
Constant	0.263	27.317	0.000	0.269	27.029	0.000
#Coauthors	−0.095	−9.027	0.000***	−0.095	−5.750	0.000
#Co-affiliations	0.197	18.809	0.000***	0.194	11.395	0.000
#References	0.018	2.961	0.003**	−0.020	−1.760	0.078
#Pages	−0.030	−5.711	0.000***	−0.008	−1.016	0.309
#Keywords	0.053	9.469	0.000***	0.061	5.045	0.000
Australia	−0.225	−17.978	0.000***	−0.231	−17.667	0.000***
Korea	−0.421	−34.963	0.000***	−0.426	−34.584	0.000***
#Coauthors * Australia				−0.029	−0.878	0.380
#Co-affiliations * Australia				0.071	2.720	0.007**
#References * Australia				0.035	2.341	0.019*
#Pages * Australia				−0.027	−2.202	0.028*
#Keywords * Australia				0.001	0.036	0.971
#Coauthors * Korea				0.076	3.074	0.002***
#Co-affiliations * Korea				−0.091	−3.492	0.000***
#References * Korea				0.087	5.345	0.000***
#Pages * Korea				−0.039	−3.002	0.003**
#Keywords * Korea				−0.030	−2.032	0.042*
Model statistics	R^2	*F*-value	*p*-value	ΔR^2	*F*-value	*p*-value
	0.052	359.622	0.000	0.002	7.726	0.000***

* $p < .05$, ** $p < .01$, *** $p < .001$

Here Z_1 is a dummy variable for Australia and Z_2, for Korea. Independent and dependent variables were standardized. Table 11 shows the results of a hierarchical regression analysis for these models.

As shown in Table 11, the results for Model 1 indicate $B = -0.070$, $t = -6.411$ ($p < 0.001$), for Australia and $B = -0.351$, $t = -24.073$ ($p < 0.001$), for Korea, implying that Natural science had significant negative effects on Engineering. In terms of interaction effects of independent and moderating variables in Model 2, a significant positive (+) moderating effect was found for #co-affiliations * Australia [$B = 0.071$, $t = 2.720$ ($p < 0.01$)] at the 0.05 level; a significant positive (+) moderating effect, for #references * Australia [$B = 0.035$, $t = 2.341$ ($p < 0.05$)]; a significant negative (−) moderating effect, for #pages * Australia ($B = -0.027$, $t = -2.202$ ($p < 0.05$)]; a significant positive (+) moderating effect, for #coauthors * Korea [$B = 0.076$, $t = 3.074$ ($p < 0.01$)]′; a significant negative (−) moderating effect, for #co-affiliations * Korea [$B = -0.091$, $t = -3.492$ ($p < 0.001$)]; a significant positive (+) moderating effect, for #references * Korea [$B = 0.087$, $t = 5.345$ ($p < 0.001$)]; a significant negative (−) moderating effect, for #pages * Korea [$B = -0.039$, $t = -3.002$ ($p < 0.01$)]; and a significant negative (−) moderating effect, for #keywords * Korea [$B = -0.030$, $t = -2.032$ ($p < 0.01$)]. In other words, Korea had a greater effect on the number of coauthors than the other two countries, and Australia had a greater effect on the number of co-affiliations than the other two countries. In terms of the number of references, Australia had a greater effect than the other countries,

Table 12 Mediating effects: years

IV	Model 1			Model 2		
	B-value	*t*-value	*p*-value	*B*-value	*t*-value	*p*-value
Constant	0.045	5.922	0.000	0.048	6.363	0.000
#Coauthors	−0.153	−15.150	0.000***	−0.273	−14.864	0.000***
#Co-affiliations	0.267	26.418	0.000***	0.407	21.482	0.000***
#References	0.088	14.655	0.000***	0.128	10.601	0.000***
#Pages	−0.032	−6.338	0.000***	−0.038	−4.397	0.000***
#Keywords	0.051	9.493	0.000***	0.043	4.605	0.000***
2006	0.348	31.148	0.000***	0.357	31.755	0.000***
2010	−0.385	−36.786	0.000***	−0.378	−36.256	0.000***
#Coauthors * 2006				0.116	4.286	0.000***
#Co-affiliations * 2006				−0.095	−3.222	0.001***
#References * 2006				0.051	2.728	0.006**
#Pages * 2006				−0.015	−1.170	0.242
#Keywords * 2006				0.019	1.291	0.197
#Coauthors * 2010				0.162	6.610	0.000***
#Co-affiliations * 2010				−0.213	−8.931	0.000***
#References * 2010				−0.079	−5.429	0.000***
#Pages * 2010				0.016	1.332	0.183
#Keywords * 2010				−0.029	−2.312	0.021*
Model statistics	R^2	*F*-value	*p*-value	ΔR^2	*F*-value	*p*-value
	0.113	830.901	0.000***	0.006	32.442	0.000***

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

and Korea. In terms of the number of pages, Australia had a weaker effect than the other countries, and Korea also had a relatively weaker effect. Korea had a weaker effect on the number of keywords than the other countries. In addition, no significant differences were found for #coauthors * Australia and #keywords * Australia.

In Model 1, the coefficient of determination, used to measure the fit of the model, was 0.052, indicating 5.2 % of citation variable change, and explanatory power increased to 0.2 % in Model 2 because of the inclusion of interaction items. The moderating effect of was significant [$F = 7.726\,(p < 0.001)$].

To test the moderating effect by year, the variable was converted into a dummy variable, as in the case of topics and countries. The year 2008 was set as the base category $(0, 0)$; 2006, as $(1, 0)$; and 2010, as $(0, 1)$. If the factors influencing citations (numbers of coauthors, co-affiliations, references, pages, and keywords) are designated as X, then the models suggested in Baron and Kenny (1986) can be expressed as

$$\text{Model 1:} Y = \beta_0 + \beta_1 X + \beta_{21} Z_1 + \beta_{22} Z_2,$$
$$\text{Model 2:} Y = \beta_0 + \beta_1 X + \beta_{21} Z_1 + \beta_{22} Z_2 + \beta_{31} X Z_1 + \beta_{32} X Z_2.$$

Here Z_1 is a dummy variable for 2006 and Z_2, for 2010. Each variable was standardized. Table 12 shows the results of a hierarchical regression analysis for the models.

As shown in Table 12, the results for Model 1 indicate $B = 0.348$, $t = 31.148\,(p < 0.001)$, for 2006 and $B = -0.385$, $t = -36.786\,(p < 0.001)$, for 2010. In terms of inter-

Table 13 Results of hypothesis testing

Hypothesis	*t*-value	*F*-value	(*p*)	Support
H1-1: the number of coauthors has a positive correlation with the likelihood of being cited	−11.535		0.000***	No
H1-2: the number of collaborating institutions has a positive correlation with the likelihood of being cited	21.804		0.000***	Yes
H1-3: the number of references has a positive correlation with the likelihood of being cited	7.467		0.000***	Yes
H1-4: the number of pages has a positive correlation with the likelihood of being cited	−2.737		0.006**	No
H1-5: the number of keywords has a positive correlation with the likelihood of being cited	10.673		0.000***	Yes
H2-1: the paper topic has moderating effects on relationships between factors influencing citations and the number of citations		37.278	0.000***	Yes
H2-2: the paper's country of origin has moderating effects on relationships between factors influencing citations and the number of citations		7.726	0.000***	Yes
H2-3: the year of the paper has moderating effects on relationships between factors influencing citations and the number of citations		32.442	0.000***	Yes

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

action effects of the year and numbers of coauthors, co-affiliations, pages, and keywords in Model 2, a significant positive (+) moderating effect was found for #co-authors * 2006 [$B = 0.116$, $t = 4.286\,(p < 0.001)$]; a significant negative (−) effect, for #affiliations [$B = -0.095$, $t = -3.222\,(p < 0.001)$]; a significant positive (+) moderating effect, for #references * 2006 [$B = 0.051$, $t = 2.728\,(p < 0.01)$]; a significant positive (+) moderating effect, for #coauthors * 2010 [$B = 0.162$, $t = 6.610\,(p < 0.001)$], for #affiliations * 2010 [$B = -0.213$, $t = -8.91\,(p < 0.001)$]; a significant negative (−) effect, for #references * 2010 [$B = -0.079$, $t = -5.429\,(p < 0.001)$]; and a significant negative (−) effect, for #keywords * 2010 [$B = -0.029$, $t = -2.312\,(p < 0.05)$]. In other words, the year of 2006 had a greater effect on the number of coauthors than the other years, and 2010 also had a greater effect than the other years. For the number of co-affiliations, however, 2006 had a weaker effect than the other years, and 2010 also had a weaker effect than the other years. The year 2010 had a greater effect on the number of references than the other years, and 2010 had a weaker effect on the number of keywords than the other years. In addition, there were no significant differences for #pages * 2006, #keywords * 2006, and #pages * 2010.

In Model 1, the coefficient of determination, used to measure the fit of the model, was 0.113, and explanatory power increased to 0.6 % in Model 2 because of the inclusion of interaction items. The moderating effect was significant [$F = 32.442\,(p < 0.001)$].

Table 13 shows the results of hypothesis testing for each factor based on the analysis results for impact factors. As shown in Table 13, the hypotheses were generally supported at the 0.05 level. Here the casual coefficient was assumed to be significant if the *t*-value of the impact factor exceeded ±1.96. The moderating effect was assumed to be significant at the 0.01 level. However, the number of coauthors and that of pages were rejected (i.e., negative coefficients and thus negative effects on citations).

To select factors influencing incoming citations in S&T papers (except for paper quality), those factors appearing in the paper's bibliographic information were classified into five factors based on the quantitatively measurable strength of the effect. Numbers of coauthors and co-affiliations were selected from this bibliographic information on authors, and those of references, pages, and keywords were adopted from the bibliographic information. In addition, topics, countries, and years were selected to test the moderating effect. Based on the selected factors, eight hypotheses were developed, including relationships between bibliographic factors and the number of citations and moderating effects on the number of citations, and incorporated into research models.

Papers were reviewed according to their topics, countries, and publication years through a frequency analysis of target research papers. Life sciences, Korea, and the year 2010 accounted for the highest proportion of papers (45.7, 45.67, and 37.28 %, respectively).

Based on the descriptive statistics for target variables, the highest standard deviation was for the number of coauthors (50.81 %). High standard deviations were found for numbers of citations and references, whereas low ones, for numbers of co-affiliations, pages, and keywords.

The Pearson correlation coefficient was used to analyze correlations between target variables. Nonsignificant correlations were verified only between the number of coauthors and that of references, whereas others showed significant correlations. There were significant negative (−) correlations between the number of keywords and that of coauthors as well as between the number of keywords and that of co-affiliations and significant positive (+) correlations between other variables. In terms of correlations between the number of citations and other variables, the number of co-affiliations showed the highest correlation. High correlations were found between the number of co-affiliations and that of coauthors ($r = 0.899$) as well as between the number of keywords and that of references ($r = 0.563$).

In terms of each field, natural sciences showed the largest number of coauthors (17.80), and it also showed the highest average number of co-affiliations. Life sciences showed the highest average number of references (42.54). Engineering showed the highest average number of papers (10.16), and life sciences had the highest average number of keywords (9.04). Both life sciences and natural sciences showed high average numbers of citations, whereas engineering, a low average number.

In terms of countries, Switzerland and Australia had higher country impact factors than Korea and higher averages than Korea.

For the three publication years, 2010 showed the highest average numbers of co-affiliations, references, and keywords, followed by 2008 and 2006, and 2010 showed the highest average number of pages, followed by 2006 and 2008. There was no significant difference in the number of coauthors, and the results for the number of citations are not discussed because of its cumulative characteristics in each year.

According to the results of a regression analysis of impact factors for the number of citations, the number of co-affiliations had the most effect on the absolute number of citations, followed by the number of coauthors, the number of keywords, the number of references, and the number of pages, in that order. The number of coauthors [$B = -0.083$ ($p < 0.001$)] and that of pages [$B = -0.079$ ($p < 0.001$)] had significant negative (−) effects, and the number of co-affiliations [$B = 0.979$ ($p < 0.001$)], that of references [$B = 0.074$ ($p < 0.001$)], and that of keywords [$B = 0.612$ ($p < 0.001$)] had significant positive (+) effects. Topics, countries, and years had significant moderating effects. Table 14 shows the effect of each moderating variable.

According to the comprehensive analysis, relevant variables should be applied to policy-making elements for journals.

Table 14 Moderating effects on the relationship between moderating and independent variables

	Topics			Countries			Years		
	Natural sciences	Life Sciences	Engineering	Australia	Korea	Switzerland	2006	2008	2010
#Coauthors	⊖	⊕	⊕	⊖	⊖	⊖	⊖	⊖	⊖
#Co-affiliations	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
#References	⊕	⊕	⊕	⊕	⊕	⊖	⊕	⊕	⊕
#Pages	⊖	⊕	⊕	⊖	⊖	⊖	⊖	⊖	⊖
#Keywords	⊖	⊖	⊕	⊕	⊕	⊕	⊕	⊕	⊕

4 Summary

This study defines some major bibliographic factors influencing the number of citations and analyzes their dimensions and relationships to determine their moderating effects. The study compares these dimensions between three countries, namely Australia (an English-speaking-country in Oceania), Korea (a non-English-speaking country in Asia), and Switzerland (a multilingual country in Europe), and investigates Switzerland as a country producing papers with the largest number of incoming citations to verify whether any factor with the identical phenomenon would be identifiable. The results are summarized as follows.

First, the results for the Pearson correlation coefficient, used to analyze correlations between target variables (numbers of coauthors, co-affiliations, references, keywords, and citations), verify significant correlations between them except for the number of coauthors and that of references. Relatively high correlations (above 0.5) were found between the number of co-affiliations and that of references as well as between the number of keywords and that of references.

Second, in terms of differences in variables across topics and countries, significant differences were found for all target variables, including numbers of coauthors, co-affiliations, references, pages, keywords, and citations. For the publication year, a significant difference was not found only for the number of coauthors.

Third, the results of a regression analysis for the number of citations verify significant negative ($-$) moderating effects of numbers of coauthors and pages on the number of citations and significant positive ($+$) moderating effects of numbers of co-affiliations, references, and keywords on the number of citations. In other words, the larger the number of co-affiliations, references, or keywords, the larger the number of citations was. For the other independent variables, however, the larger of the number of coauthors or pages, the smaller the number of citation was.

In terms of moderating effects on bibliographic factors as independent variables by topic, country, and year, all factors had significant moderating effects. The publication year had the greatest moderating effect ($R^2 = 0.113$), followed by country ($R^2 = 0.052$) and topic ($R^2 = 0.040$), in terms of explanatory power. In terms of topics, natural sciences had positive effects only on numbers of co-affiliations and references, and life sciences had negative effects only on the number of keywords. For engineering, all factors had positive effects. Both Australia and Korea had positive effects on numbers of co-affiliations, references, and keywords, whereas Switzerland had negative effects on numbers of coauthors, references, and pages. For the publication year, all years had positive effects on numbers of co-affiliations, references, and keywords. Based on a comprehensive review of effects of various variables on the number of incoming citations, the results for the research models, which assumed positive (+) effects of the five bibliographic factors provide support for all hypotheses except for those two. In order to increase citation of articles, based on the result of this research, a collaborative research from several affiliations or teams should be facilitated, a broad range of reference retrieval is needed as the higher number of references means vigorous research, and the more number of keywords needs to be allowed when submitting articles through modification of editing policies of academic journals. The novelty of our study is big size of data. Research is being conducted on a number of factors affecting the citation. There are not many cases have been carried out so far have analyzed the data of a large-scale regional factors influence the increase in citations of research papers.

Acknowledgments This paper is part of dissertation entitled "Factors Affecting Citations in Science and Technology" under the supervision of Sangki Choi and the previous version of this paper was presented at Daegu Gyeongbuk International Social Network Conference in South Korea (December, 2013).

References

Ashton, S.V., Oppenheim, C.: A method of predicting Nobel prizewinners in chemistry. Soc. Stud. Sci. **8**(3), 341–348 (1978)

Baron, R.M., Kenny, D.A.: The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical consideration. J. Personal. Soc. Psychol. **51**(6), 1173–1182 (1986)

Garfield, E.: Of Nobel class: part 2 forecasting Nobel prizes using citation data and the odds against it. Curr. Comments **35**, 127–136 (1992)

Garfield, E., Merton, R.K.: Citation Indexing: Its Theory and Application in Science, Technology, and Humanities, vol. 8. Wiley, New York (1979)

Kim, W.J.: A study on the factors influencing citation speed and citation frequency of scientific articles using bibliometric analysis in South Korea. J. Korean Soc. Libr. Inf. Sci. **46**(4), 285–309 (2012)

Lee, Y.G.: Multidisciplinary team research as an innovation engine in knowledge-based transition economies and implication for Asian countries. J. Contemp. East. Asia **12**(1), 49–63 (2013)

Levitt, J.M., Thelwall, M.: Citation levels and collaboration within library and information science. J. Am. Soc. Inf. Sci. Technol. **60**(3), 434–442 (2009)

Marks, R.: Learned societies adapt to new publishing realities. In: Fredrickson, E. (ed.) A Century of Science Publishing: A Collection of Essay, pp. 91–96. IOS Press, Amsterdam (2001)

Moed, H.: Citation Analysis in Research Evaluation. Springer, Dordrecht (2005)

Norris, M.: The Citation advantage of open-access articles. J. Am. Soc. Inf. Sci. Technol. **59**(12), 1963–1972 (2008)

Opthof, T.: Sense and nonsense about the impact factor. Cardiovasc. Res. **33**(1), 1–7 (1997)

Rousseau, R. (2000). Are multi-authored articles cited more than single-authored ones? Are collaborations with authors from other countries more cited than collaborations within the country? A case study. In: Proceedings of the Second Berlin Workshop on Scientometrics and Informetrics, pp. 173–176

So, M.H.: SCI Analysis Research. Korea Advanced Institute of Science and Technology (2012)

Turpin, T., Woolley, R., Garrett-Jones, S.: Cross-sector research collaboration in Australia: the Cooperative Research Centres Program at the crossroads. Sci. Public Policy **38**(2), 87–97 (2011)

Van Leeuwen, T., et al.: The holy grail of science policy: exploring and combining bibliometrics tools in search of scientific excellence. Scientometrics **57**(2), 257–280 (2003)

Wuchty, S., Jones, B.F., Uzzi, B.: The increasing dominance of teams in production of knowledge. Science **316**, 1036–1039 (2007)

Yoo, J.B., Chung, Y.M.: Analysis of factors influencing patent citations. J. Korean Soc. Inf. Manag. **27**(1), 103–118 (2010)