Counting Methods, Country Rank Changes, and Counting Inflation in the Assessment of National Research Productivity and Impact

Mu-Hsuan Huang and Chi-Shiou Lin
Department of Library & Information Science, National Taiwan University, Number 1, Section 4, Roosevelt Road, Taipei, Taiwan. E-mail: {mhhuang, chishioulin}@ntu.edu.tw

Dar-Zen Chen
Department of Mechanical Engineering, National Taiwan University, Number 1, Section 4, Roosevelt Road, Taipei, Taiwan. E-mail: dzchen@ntu.edu.tw

The counting of papers and citations is fundamental to the assessment of research productivity and impact. In an age of increasing scientific collaboration across national borders, the counting of papers produced by collaboration between multiple countries, and citations of such papers, raises concerns in country-level research evaluation. In this study, we compared the number counts and country ranks resulting from five different counting methods. We also observed inflation depending on the method used. Using the 1989 to 2008 physics papers indexed in ISI's Web of Science as our sample, we analyzed the counting results in terms of paper count (research productivity) as well as citation count and citation–paper ratio (CP ratio) based evaluation (research impact). The results show that at the country-level assessment, the selection of counting method had only minor influence on the number counts and country rankings in each assessment. However, the influences of counting methods varied between paper count, citation count, and CP ratio based evaluation. The findings also suggest that the popular counting method (whole counting) that gives each collaborating country one full credit may not be the best counting method. Straight counting that accredits only the first or the corresponding author or fractional counting that accredits each collaborator with partial and weighted credit might be the better choices.

Introduction

The counting of publications and citations is fundamental to the assessment of research productivity and impact. In an age of increasing scientific collaboration across national and institutional borders, it becomes conceptually and methodologically challenging to conduct counting for an internationally collaborated paper to show each collaborator's contribution—such as individual authors, institutions, or countries, depending on the unit of analysis. The frequently used counting methods in most scientometric research are those that attribute equal credit to every collaborator (i.e., each collaborator gets a full credit or equal share of one credit) or that accredit one with all or nothing (e.g., only the first author gets the credit) (Gauffriau, Larsen, Maye, Roulin-Perriard, & von Ins, 2007, 2008).

In this study, we examined country rank changes and counting inflation due to the use of different counting methods in country-level research assessment. Certain methods unavoidably produce larger numbers than do other methods for each country being assessed. The changes in paper and citation counts may consequently affect country rankings. The goal of this article is to systematically show the influence of five commonly used counting methods on country rankings and to what extent the numbers may vary depending on the method used. Using the complete 1989 to 2008 papers and citations data of the physics journals from Thomson Reuter's ISI Web of Science (WoS), this article offers a highly accurate picture of how counting methods affect country ranks and the degrees of counting inflation in a subject field characterized by heavy and intensifying international collaboration.

Counting Methods and Their Problems

Types of Counting Method

The existing literature has identified a number of counting methods that are named inconsistently (Gauffriau, et al., 2008; Larsen, 2007), but they can be categorized into three

Received March 23, 2011; revised June 28, 2011; accepted June 30, 2011

 • Published online 8 September 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.21625

major fashions of counting. The first is the all counting approach, which accredits each collaborator with one full credit. This approach is represented by two different counting methods: whole counting and complete counting. The former attributes one full credit to each unique collaborating country or institution; the latter attributes one credit to each author's respective country or institution, depending on the level of analysis (Gauffriau et al., 2007, 2008). The difference between whole counting and complete counting can be shown with an example: In country-level research assessment, a paper to be counted is collaborated on by four institutions in three countries—two in the United States, one in Germany, and one in Japan. In this case, by whole counting, each of the three countries receives one credit. By complete counting, the United States will receive two credits, and Germany and Japan each receive one.

The second approach is straight counting. In this approach, only the most prominent collaborator receives one full credit, and the others receive none. First author counting belongs to this category (Gauffriau et al., 2007). Some scholars, however, have argued for attributing the credit to the corresponding author rather than to the first author in straight counting (Man, Weinkauf, Tsang, & Sin, 2004, as cited in Gauffriau & Larsen, 2005).

The third approach is fractional counting. In this approach, one credit is shared by all collaborators. Two types of counting method are in this category; in whole-normalized counting (WN), all the unique basic units share one credit whereas in complete-normalized counting (CN), all of the basic units in a publication share one credit. Using the previous example mentioned in all counting, when WN is used to attribute credits, each country will equally receive one third of the credit. But if CN is used instead, the United States will then receive one half while Germany and Japan each receive one fourth.

Problems and Consequences of Counting Methods

The various problems associated with counting methods in scientometric research have been discussed in prior studies (Gauffriau & Larsen, 2005; Gauffriau et al., 2008). Gauffriau et al. (2007) indicated that counting methods can be conceptually differentiated by whether they are rank dependent, additive/nonadditive, and normalized/nonnormalized. Each differentiation introduces certain conceptual or technical problems.

Rank dependency refers to whether each basic unit in a collaborated paper receives unequal credit based on the authorship order. Of the methods we tested, all counting and fractional counting are rank-independent methods: They wholly disregard the order of authorship and accredit each collaborator with an equal share of credit. In contrast, straight counting is rank dependent because it accredits only one collaborator who occupies the prominent-authorship position (e.g., the first author or the corresponding author). Other authors have criticized the crudeness of accrediting without carefully differentiating and weighting each collaborator's real contribution (e.g., Egghe, Rousseau, & Hooydonk, 2000; Nudelman & Landers, 1972; Pravdić & Oluić-Vukovic, 1986). This is a valid criticism, but the five methods remain popular in existing scientometric studies for their ease of use or understandability.

The additive/nonadditive and normalized/nonnormalized differences between counting methods are of particular concern for this article. The whole counting method is a nonadditive method. At the aggregate level, it produces a paper count that exceeds the sum of each basic unit's received paper count. Being a nonnormalized method, whole counting also inevitably produces a larger paper count than does straight counting or fractional counting (i.e., WN and CN). This suggests that whole counting will introduce a problem of counting inflation. For instance, Gauffriau and Larsen (2005) found that in country-level assessment using U.S. National Science Foundation statistics, a country's publication number reduction rate could as high as 10 to 32% by going from whole counting to fractional counting. In another study, they found that with the intensifying research collaborations from 1981 to 2002, whole counting had led to chronologically greater counting inflation (Gauffriau et al., 2008).

We particularly focused on comparing whole counting with the other methods because the former is widely used as the de facto method in bibliometric studies (Gauffriau et al., 2008) and in several worldwide research-evaluation programs such as the World University Rankings by Quacquarelli Symonds, the Academic Ranking of World Universities by Shanghai Jiaotong University, the Performance Ranking of Scientific Papers for World Universities by the Higher Education Evaluation and Accreditation Council of Taiwan, and so on (Huang, 2011; also see ARWU, n.d.; Higher Education Evaluation and Accreditation Council of Taiwan, n.d.; Quacquarelli Symonds, n.d.). Given the popularity of the method, to what extent it may inflate paper and citation counts should be studied.

In addition to counting inflation, whether counting methods affect country rankings also has been debated. Scholars have argued that counting methods are of limited impact on ranking (see Gauffriau & Larsen, 2005, p. 90; Gauffriau et al., 2008, pp. 161–163, 172). This article tested the methods on a large set of bibliometric data to empirically examine whether counting method choice did significantly influence country ranks.

Methodology

Counting Methods and Data Processing

We used the complete data of the physics papers published between 1989 and 2008 from the ISI WoS. Our research team wrote a program to automatically parse the WoS data. For each record, we calculated the author number based on the names recorded in the author field (AU). In this article, international-collaboration papers were defined as papers coauthored by two or more institutions in different countries. We determined whether a paper was internationally collaborated based on the author address field (C1), which

DOI: 10.1002/asi

theoretically included all authoring institutions' addresses, including that of the first author, and the corresponding author address field (RP), which listed only the corresponding author's institution address.

Note that 41,390 paper entries (2.86% of the 1,445,273 papers) lacked address information in C1 and RP; they were purged from our analyses. Consequently, the total paper number in the following analyses was 1,403,883. In addition, some individual authors listed more than one institutional affiliation in the publication data. For example, we identified 11,060 single-authored papers (0.77% of the papers) that listed multiple institution addresses. However, we were not able to discern the exact number of authors affiliating with more than one institution in the entire dataset. International collaboration in this article was determined by the presence of multiple nationalities in the address fields. If an author's institutions were located in different countries, the paper was considered an internationally collaborated work.

Based on the dataset, we used five counting methods to count the numbers of each country's international-collaboration papers and citations. The five methods represented the three counting approaches introduced earlier: all counting, straight counting, and fractional counting. We forwent the complete counting method because we did not consider it a reasonable approach. Using the previous example again, when a paper is collaborated on by two U.S. institutions, one German institution, and one Japanese institution, it is reasonable to say that each of the three countries has produced one paper (whole counting) or part of the paper (WN or CN); but it does not make sense to say that the United States has produced two papers (complete counting) when in fact there is just one paper.

- *Whole counting (all counting):* Regardless of the order of authorship, if the collaborating authors' institutions recorded in C1 were located in different countries, then each country was considered to have produced one paper.
- *Straight counting using the first author:* Only the first institution address listed in C1 was counted. In our dataset, 250,956 entries (17.36% of the 1,445,273 papers) lacked the C1 field, so we used RP instead.
- *Straight counting using the corresponding author:* Only the institution address in RP was counted. A total of 40,697 entries (2.82% of the 1,445,273 papers) lacked the RP field, so we used the first address in C1 instead.
- *Complete-normalized counting (fractional counting I):* All institutions listed in C1 were used as the basis for counting. Each institution received an equal share of one credit, and the fractional credits of each institution in the same country were added and formed that country's share. Please note that 17.36% of our dataset lacked the C1 field, so RP was used instead. Because the RP field recorded only one institution address, these entries unavoidably were counted as being produced by one single country, and thus misjudgment may have occurred in the calculation.
- *Whole-normalized counting (fractional counting II):* Regardless of the number of collaborating institutions in C1, only the number of nationalities was considered. Each collaborating country received an equal share of the credit.

The Study Sample: Physics Papers, 1989–2008

In October 2008, 336 journals were listed under the category of physics in Essential Science Indicators (ESI). Within the time frame, a total of 1,445,273 papers with authors from 165 countries were published in those 336 journals. The papers were produced by 6,658,522 authors and were cited 17,005,626 times. On average, each paper had 4.61 authors and received 11.77 citations. A total of 1,189,863 (82.33%) were collaborated papers produced by 6,403,112 authors. These collaborated papers (including intranational- and international-collaboration papers) were cited 14,900,115 times. On average, each collaborated paper had 5.38 authors and received 12.52 citations.

Judging from the author affiliations, 329,447 (22.79%) of the sample papers were international-collaboration papers. These papers were produced by 3,135,587 authors and received 4,786,873 citations. Each of the papers had 9.52 authors and received 14.53 citations, on average. One can see that although international-collaboration papers accounted for only 22.79% of the physics papers, the increase in the average author and citation numbers for all papers, all collaborated papers, and international-collaboration papers reflects noticeable citation-counting inflation.

In the next section, we report the results of the five counting methods on the dataset in terms of paper count (representing productivity assessment), citation count (representing impact assessment), and paper–citation ratio (CP ratio, representing normalized impact assessment). We focus on the country rank changes and the counting inflation rates of the top 30 countries by any of the five counting methods in each assessment.

Country Rank Changes and Counting Inflation

Paper Count

Country rank changes. Table 1 shows slight variation in country rankings by the five methods; the distribution of country ranks was rather similar. The United States ranked first in all five rankings. At the other end, Singapore and Hungary could both drop out of the top 30 when a certain method was applied. Hungary made it into the top 30 only in Method A. In contrast, Singapore was able to reach the top 28 in all the methods but Method A.

Some countries' ranks never changed regardless of the method applied, such as the United States (1), Russia (5), France (6), and so on. For the other countries of varied ranks, a closer examination revealed some interesting patterns. First, one can identify several clusters of countries with adjacent ranks. Within each cluster, country ranks varied by method, but ranks were interchangeable only within the same cluster (e.g., the cluster of Germany, Japan, and China). The two larger clusters are located in the lower half of the table (i.e., the cluster of Switzerland to Taiwan and the cluster of Denmark to Singapore).

Another interesting observation is that when Method A was applied, the Western countries often ranked higher than

DOI: 10.1002/asi

TABLE 1. Paper counts and country rankings by different counting methods.

Country*		1989–2008 Paper numbers by different counting methods					Country rank by paper count					Ratio of counting inflation**			
		A	B	C	D	E	A	B	C	D	E	A/B	A/C	A/D	A/E
Total counts		1,876,809	1,403,883	1,403,883	1,403,880	1,403,881	–	–	–	–	–	1.34	1.34	1.34	1.34
United States		381,902	310,987	310,154	314,185	310,320	1	1	1	1	1	1.23	1.23	1.22	1.23
Germany	–	159,718	109,210	110,667	108,125	109,175	2	4	3	4	3	1.46	1.44	1.48	1.46
Japan	+	157,485	135,041	135,061	135,970	135,150	3	2	2	2	2	1.17	1.17	1.16	1.17
China	+	122,419	110,190	109,656	108,353	107,899	4	3	4	3	4	1.11	1.12	1.13	1.13
Russia		118,943	89,044	87,291	90,045	90,060	5	5	5	5	5	1.34	1.36	1.32	1.32
France		115,471	78,270	78,124	79,640	79,464	6	6	6	6	6	1.48	1.48	1.45	1.45
United Kingdom		102,943	72,637	72,739	72,718	73,314	7	7	7	7	7	1.42	1.42	1.42	1.40
Italy		75,820	54,118	53,560	54,321	51,564	8	8	8	8	8	1.40	1.42	1.40	1.47
India		47,917	41,250	41,109	40,879	40,899	9	9	9	9	9	1.16	1.17	1.17	1.17
Canada	–	43,892	30,170	30,393	30,364	30,963	10	11	11	11	11	1.45	1.44	1.45	1.42
Spain	–	41,946	28,853	29,198	27,958	28,176	11	12	12	12	12	1.45	1.44	1.50	1.49
South Korea	+	40,379	32,984	33,179	32,768	32,775	12	10	10	10	10	1.22	1.22	1.23	1.23
Poland		37,931	24,803	24,884	25,031	25,355	13	13	13	13	13	1.53	1.52	1.52	1.50
Switzerland	–	35,055	19,528	20,141	19,145	20,215	14	16	15	16	15	1.80	1.74	1.83	1.73
Brazil	+	28,183	21,462	21,595	20,972	21,159	15	14	14	14	14	1.31	1.31	1.34	1.33
The Netherlands	–	27,560	17,560	17,661	17,687	18,099	16	17	17	17	17	1.57	1.56	1.56	1.52
Australia	–	23,651	17,390	17,494	17,261	17,445	17	18	18	18	18	1.36	1.35	1.37	1.36
Taiwan	+	23,622	19,955	20,127	19,753	19,866	18	15	16	15	16	1.18	1.17	1.20	1.19
Sweden	–	23,528	14,457	14,663	14,324	14,618	19	20	20	20	20	1.63	1.60	1.64	1.61
Israel	+	21,976	15,182	15,360	14,910	15,292	20	19	19	19	19	1.45	1.43	1.47	1.44
Belgium	–	18,511	11,426	11,615	11,096	11,396	21	22	22	22	22	1.62	1.59	1.67	1.62
Ukraine	+	17,785	12,523	12,381	12,550	12,726	22	21	21	21	21	1.42	1.44	1.42	1.40
Austria	–	13,562	7,869	7,993	7,711	7,992	23	24	24	24	24	1.72	1.70	1.76	1.70
Mexico	+	13,165	9,712	9,784	9,550	9,668	24	23	23	23	23	1.36	1.35	1.38	1.36
Czech Republic		12,570	7,699	7,723	7,661	7,777	25	25	25	25	25	1.63	1.63	1.64	1.62
Denmark	–	12,560	6,986	7,023	7,144	7,375	26	27	26	26	26	1.80	1.79	1.76	1.70
Finland	–	10,565	6,504	6,541	6,313	6,407	27	30	30	30	30	1.62	1.62	1.67	1.65
Greece	–	10,415	6,611	6,713	6,447	6,544	28	29	29	29	29	1.58	1.55	1.62	1.59
Argentina	+	9,744	7,046	6,976	7,069	7,012	29	26	27	27	27	1.38	1.40	1.38	1.39
Hungary	–	8,940	4,873	4,849	4,965	5,071	30	32	32	32	32	1.83	1.84	1.80	1.76
Singapore	+	8,390	6,811	6,873	6,600	6,609	31	28	28	28	28	1.23	1.22	1.27	1.27

Note. *The plus sign (+) indicates rank rise from Method A to other methods; the minus sign (–) indicates rank drop. **The different gray levels of the background indicate the quartered ranges of counting inflation ratio from 1.0 to 2.0 (1.00–1.25; 1.26–1.50; 1.51–1.75; 1.76–2.00).

did the other countries within the same clusters. In contrast, when Methods B to E were applied, the East Asian countries and those that can be described as "newly industrializing economies" (Central Intelligence Agency, n.d.) or emerging (CME Group Index Services, 2010) were ranked higher than were their Western counterparts; for example, Japan and China as opposed to Germany (Cluster 1), and South Korea as opposed to Canada and Spain (Cluster 2), and so on. But even in the larger clusters, for each country the observed rank differences did not exceed 3. This is different from what happened in citation counts and CP ratio rankings (discussed later).

Finally, comparing the results of the five different methods, one can see that Methods B to D resulted in rather consistent rank-change direction as compared to Method A. That is, when any of the methods other than Method A was applied, a country's country rank could either rise (those marked with a plus sign), remain the same (those without signs), or drop (those marked with a minus sign), but the direction of rising or dropping was always the same across countries. Given the similar results of Methods B to E, it seems safe to conclude that of the five methods, Method A may result in a greater difference in productivity assessment than may the others.

Counting inflation in paper counts. We used the paper counts from whole counting (Method A) as the basis to calculate the ratio of counting inflation. The ratio was obtained by dividing each country's paper count from Method A by those from Methods B to E, respectively. For the top-30 countries, the counting inflation ratio ranged between 1.11 (China in Method B) and 1.84 (Hungary in Method C).

One can see that except for the United States, all other countries with an inflation ratio lower than 1.25 were Asian. Only Singapore had an inflation ratio slightly higher than 1.25 in two counting methods. The low inflation suggests that those countries possibly had not been involved as much as the were other countries in international collaboration.

All the Western countries within the top-10 countries, including Russia, had inflation ratios ranging between 1.26 to 1.50. Other Western countries in the top 20 to 30 could have

TABLE 2. Citation counts and country rankings by different counting methods.

Country*		1989–2008 Paper numbers by different counting methods					Country rank by citation count					Ratio of counting inflation**			
		A	B	C	D	E	A	B	C	D	E	A/B	A/C	A/D	A/E
Total counts		24,661,146	16,949,437	16,949,437	16,949,420	16,949,416	–	–	–	–	–	1.45	1.45	1.45	1.45
United States		7,275,777	6,027,548	6,037,877	6,009,977	5,885,082	1	1	1	1	1	1.21	1.21	1.21	1.24
Germany		2,497,703	1,690,025	1,722,411	1,638,338	1,656,176	2	2	2	2	2	1.48	1.45	1.52	1.51
Japan		1,839,626	1,479,329	1,484,280	1,488,909	1,479,183	3	3	3	3	3	1.24	1.24	1.24	1.24
France		1,598,473	1,001,838	1,015,153	1,009,179	1,014,751	4	4	4	4	4	1.60	1.57	1.58	1.58
United Kingdom		1,521,863	998,190	1,007,626	1,000,981	1,012,366	5	5	5	5	5	1.52	1.51	1.52	1.50
Russia	–	1,002,689	529,915	476,481	571,475	568,641	6	8	8	7	6	1.89	2.10	1.75	1.76
Italy	~	976,199	602,440	572,011	606,956	560,290	7	6	6	6	8	1.62	1.71	1.61	1.74
Switzerland	–	722,743	371,418	401,071	352,171	385,701	8	10	9	10	10	1.95	1.80	2.05	1.87
China	+	721,634	565,947	554,833	565,640	563,844	9	7	7	8	7	1.28	1.30	1.28	1.28
Canada	+	644,067	377,796	382,065	381,921	399,212	10	9	10	9	9	1.70	1.69	1.69	1.61
Spain		530,568	310,633	308,610	304,917	306,699	11	11	11	11	11	1.71	1.72	1.74	1.73
The Netherlands		474,071	292,340	298,212	293,691	302,051	12	12	12	12	12	1.62	1.59	1.61	1.57
Poland	–	359,719	156,564	147,913	171,026	177,683	**13**	**18**	**18**	**18**	**18**	2.30	**2.43**	2.10	2.02
India	+	357,443	248,287	243,100	249,175	251,856	14	13	13	13	13	1.44	1.47	1.43	1.42
Israel	–	351,262	192,460	195,338	193,847	204,542	15	16	16	16	15	1.83	1.80	1.81	1.72
Sweden	–	340,635	183,768	185,673	179,449	185,244	16	17	17	17	17	1.85	1.83	1.90	1.84
South Korea	+	320,764	203,504	203,380	208,047	212,256	**17**	**14**	**14**	**14**	**14**	1.58	1.58	1.54	1.51
Australia	+	291,227	199,812	201,678	195,787	199,800	**18**	**15**	**15**	**15**	**16**	1.46	1.44	1.49	1.46
Belgium	–	239,389	128,509	132,993	124,526	130,461	19	20	21	21	20	1.86	1.80	1.92	1.83
Brazil	+	236,580	145,551	142,637	145,131	148,660	20	19	19	19	19	1.63	1.66	1.63	1.59
Denmark	~	233,378	128,177	126,400	125,687	130,334	21	21	22	20	21	1.82	1.85	1.86	1.79
Austria	~	225,859	119,863	134,265	116,119	127,030	22	23	20	23	22	1.88	1.68	1.95	1.78
Taiwan	+	176,364	120,693	122,189	119,702	123,833	23	22	23	22	23	1.46	1.44	1.47	1.42
Finland		157,582	76,559	76,413	75,743	78,036	24	24	24	24	24	2.06	2.06	2.08	2.02
Greece		115,547	58,290	57,634	58,153	59,558	25	25	25	25	25	1.98	2.00	1.99	1.94
Czech Republic	–	110,965	48,630	47,344	52,219	53,826	26	28	28	28	28	2.28	2.34	2.12	2.06
Hungary	–	107,340	43,498	40,707	48,172	49,617	**27**	**32**	**32**	**30**	**30**	**2.47**	**2.64**	2.23	2.16
Mexico	+	103,456	54,700	54,898	56,389	58,868	28	26	26	26	26	1.89	1.88	1.83	1.76
Ukraine	–	91,716	48,508	46,415	51,252	52,683	29	29	30	29	29	1.89	1.98	1.79	1.74
Argentina	+	88,742	52,474	51,320	54,102	53,993	**30**	**27**	**27**	**27**	**27**	1.69	1.73	1.64	1.64
Singapore	+	57,094	45,559	46,511	43,535	44,037	**35**	**30**	**29**	**31**	**31**	1.25	1.23	1.31	1.30

Note. *The plus sign (+) indicates rank rise from Method A to other methods; the minus sign (−) indicates rank drop; the swung dash (~) indicates rank fluctuation. **Country ranks in bold indicate that the rank position difference by the five methods ≥3. **The different graying levels of the background indicate the quartered ranges of counting inflation ratio from 1.0–2.0 and above (1.00–1.25; 1.26–1.50; 1.51–1.75; 1.76–2.00; ≥2.01).

higher inflation. Switzerland, Denmark, Austria, and Hungary had the highest inflation, as compared to other countries. High inflation suggests that those countries, when participating in international collaboration, had served supporting or facilitating roles rather than providing the lead investigators.

The degree of inflation logically echoed the rise and fall of ranks within each cluster of countries. Countries with lower inflation ratios advanced in ranking when Methods B to E were applied. In contrast, countries with relatively higher ratios in each cluster dropped in rank when Methods B to E were used for paper counts. The wider gaps in counting inflation can be observed in Cluster 3 (i.e., Switzerland vs. Taiwan) and Cluster 7 (i.e., Denmark and Hungary vs. Singapore).

Citation Count

Country rank changes. As Table 2 shows, the United States again ranked first by all five methods. Singapore and Hungary could both fall out of the top 30 when a certain method was applied. Hungary made it into the top 30 in whole counting (Method A) and fractional counting (Methods D amd E). In contrast, Singapore entered the list in straight counting using first and corresponding authors (Methods B and C).

Some countries' ranks never changed regardless of counting method applied. Countries with interchangeable ranks formed four clusters. Ten countries rose in ranking by Method A compared with the others (e.g., Canada et al.), nine descended (e.g., Russia et al.), and three fluctuated (Italy, Denmark, and Austria). Rank changes in citation counts were larger than they were in paper counts. For example, Singapore showed drastic change, rising from 35 (Method A) to 29 (Method C); Poland and Hungary both could drop five positions with a change of counting method.

In paper count, whole counting (Method A) tended to favor the Western countries while straight counting (Methods B and C) and fractional counting (Methods D and E) better represented the East Asian and other emerging countries. This tendency, however, was less marked in citation counts. We did

DOI: 10.1002/asi

see that the countries whose ranks dropped from Method A to Method E were all European countries and Israel, and most of the countries with rising ranks were Asian and emerging countries; however, Canada and Australia also both rose. The rank fluctuations of Italy, Denmark, and Austria also were inconsistent with their status in paper count.

In citation count, whole counting (Method A) resulted in ranks that were different from those by the other methods. In contrast, the straight counting and fractional counting (Methods B to E) produced very similar ranks. Methods B to E produced exactly the same ranks in 17 countries. Furthermore, rank differences between Methods B E were usually small.

Note that the United States, Germany, and Japan were invariably the top-three countries by paper and citation count by any counting method; however, a few top-10 countries showed dramatic rank differences between paper count and citation count. Two noticeable examples were China and Switzerland. China ranked rather high in paper count, but its rank dropped in citation count. Switzerland, on the other hand, rose from the middle (14–16th) in paper count to somewhere between the top 8 and top 10 in citation count.

Counting inflation in citation count. Citation count generated from whole counting (Method A) was again used as the basis for calculating the counting inflation ratio. For the top-30 countries, counting inflation ranged between 1.21 (United States in Methods B–E) and 2.64 (Hungary in Method C). Only three countries (U.S., Japan, and Singapore) had inflation ratios lower than 1.25.

In paper count, Asian countries had markedly lower counting inflation; however, the situation varies for each country. For example, although most Asian countries had lower inflation, South Korea had an inflation ratio higher than 1.50. On the other hand, some Western countries such as Australia, Germany, and the United Kingdom could show comparatively lower inflation. The countries with high inflation (e.g., >1.76), however, were predominantly European countries, especially Hungary, Czech, Poland, and Finland (inflation ratios all >2.0).

Russia had a particularly unfavorable citation count in A/C (whole counting vs. straight counting by corresponding author). This indicates that Russia often did not provide the corresponding author in impactful research. Also note that Method D (normalized whole counting) particularly disfavored Switzerland. This suggests that a larger portion of Switzerland's cited papers were collaborated on by several countries. But compared to the countries which also had high inflation in A/D (e.g., Poland, Czech, Hungary), Switzerland had provided more research leadership so that its inflation in both A/B and A/C was markedly lower.

In addition, in paper count, the upper half of the countries generally showed lower inflation than did the lower half. But the distribution of the lower and higher inflation ratio values among the top-30 countries was irregular in citation count. This shows that counting methods somewhat differently affected paper and citation counts.

Citation count inflation did echo the rank rise or rank drop from Method A to the others. All countries with rising ranks except Mexico had relatively lower inflation than did countries with dropping ranks. Countries with unchanged ranks could demonstrate low inflation (e.g., Japan, Germany), medium inflation (e.g., France, Netherlands), or very high inflation (e.g., Finland). This suggests that country rank and counting inflation may not be as closely related in citation count as they are in paper count.

Ratio of CP Count

Country rank changes. Citation count is one way to observe the research impact of a country. Another way is to divide the citation count by the paper count to observe research impact per paper. Note that while the CP ratio is capable of producing a normalized impact assessment, it also can be problematic. A country with small paper and citation counts may turn out to have a much higher CP ratio than does another country that has large quantities of both papers and citations. For instance, when Method A was used on our sample, the top country among the 165 countries by CP ratio was a particular country that had produced only four papers and had received 432 citations over 20 years. Although its CP ratio is as high as 108.00, it would be controversial to say that this particular country was more impactful than, say, Switzerland (CP ratio = 20.62), which produced 35,055 papers and received 722,743 citations over the same period of time. The comparison of CP ratio must be reasonably confined to countries that have a relatively large production of papers and citations. Therefore, we report the CP ratio observation within the top-30 countries previously identified by paper and citation counts—in fact, 31 countries in total due to the ranking differences resulting from the use of different counting methods. Table 3 shows the CP ratio values and rankings of the 31 countries by the five methods.

Within the top-30 countries, a few rough clusters were observed within which ranks were interchangeable. But the rank interchangeability within clusters as well as the number of countries within the larger clusters were both greater than those seen in paper and citation counts (see Tables 1 and 2). This suggests that counting methods affected the CP ratio based rankings more than it did those based on paper and citation counts.

One can see that the top-12 countries in the upper half of the table (from Switzerland to France) were undoubtedly the best performing countries. Each of them could be within the top-10 countries by one of the counting methods. Moreover, most of their CP ratio values were higher than those of the rest of the world's by the five methods, except Finland's CP ratio in Methods B, C, and D. The second cohort of countries was those ranked between 13 and 20 by any of the methods (from Belgium to Argentina). The rest of the countries formed the third cohort. Ukraine invariably fell out of the top 30 by CP ratio.

CP ratio revealed some major differences from citation count. First, the United States no longer dominated

DOI: 10.1002/asi

TABLE 3. CP ratio and country rankings by different counting methods.

Country*		CP ratio by different counting methods					Country rank by CP ratio**					Ratio of counting inflation**			
		A	B	C	D	E	A	B	C	D	E	A/B	A/C	A/D	A/E
World		13.14	12.07	12.07	12.07	12.07	–	–	–	–	–	**1.09**	**1.09**	**1.09**	**1.09**
Switzerland	−	20.62	19.02	19.91	18.39	19.08	1	2	1	2	1	1.08	1.04	1.12	1.08
United States	+	19.05	19.38	19.47	19.13	18.96	2	1	2	1	2	0.98	0.98	1.00	1.00
Denmark		18.58	18.35	18.00	17.59	17.67	3	3	3	3	3	1.01	1.03	1.06	1.05
The Netherlands		17.20	16.65	16.89	16.60	16.69	4	4	4	4	4	1.03	1.02	1.04	1.03
Austria	−	16.65	15.23	16.80	15.06	15.90	5	6	5	6	5	1.09	0.99	1.11	1.05
Israel	−	15.98	12.68	12.72	13.00	13.38	**6**	**10**	**9**	8	8	1.26	1.26	1.23	1.20
Germany	+	15.64	15.48	15.56	15.15	15.17	7	5	6	5	6	1.01	1.00	1.03	1.03
Finland	−	14.92	11.77	11.68	12.00	12.18	**8**	**12**	**12**	**12**	**12**	1.27	1.28	1.24	1.22
United Kingdom	+	14.78	13.74	13.85	13.77	13.81	9	7	7	7	7	1.08	1.07	1.07	1.07
Canada	~	14.67	12.52	12.57	12.58	12.89	10	11	11	10	9	1.17	1.17	1.17	1.14
Sweden	+	14.48	12.71	12.66	12.53	12.67	11	9	10	11	11	1.14	1.14	1.16	1.14
France	+	13.84	12.80	12.99	12.67	12.77	**12**	**8**	**8**	**9**	10	1.08	1.07	1.09	1.08
Belgium	−	12.93	11.25	11.45	11.22	11.45	13	14	14	14	14	1.15	1.13	1.15	1.13
Italy	−	12.88	11.13	10.68	11.17	10.87	14	15	16	15	17	1.16	1.21	1.15	1.18
Spain	−	12.65	10.77	10.57	10.91	10.88	15	17	17	17	16	1.17	1.20	1.16	1.16
Australia	+	12.31	11.49	11.53	11.34	11.45	**16**	**13**	**13**	**13**	**13**	1.07	1.07	1.09	1.08
Hungary	−	12.01	8.93	8.39	9.70	9.78	17	18	19	18	18	1.35	1.43	1.24	1.23
Japan	+	11.68	10.95	10.99	10.95	10.94	**18**	16	**15**	16	**15**	1.07	1.06	1.07	1.07
Greece	+	11.09	8.82	8.59	9.02	9.10	19	19	18	19	19	1.26	1.29	1.23	1.22
Poland	−	9.48	6.31	5.94	6.83	7.01	**20**	**24**	**26**	22	22	1.50	1.60	1.39	1.35
Argentina	+	9.11	7.45	7.36	7.65	7.70	21	20	20	20	20	1.22	1.24	1.19	1.18
Czech	−	8.83	6.32	6.13	6.82	6.92	22	23	23	23	23	1.40	1.44	1.30	1.28
Russia	−	8.43	5.95	5.46	6.35	6.31	**23**	**28**	**29**	**26**	**26**	1.42	1.54	1.33	1.34
Brazil	+	8.39	6.78	6.61	6.92	7.03	**24**	**21**	22	**21**	**21**	1.24	1.27	1.21	1.19
South Korea	+	7.94	6.17	6.13	6.35	6.48	25	25	24	25	25	1.29	1.30	1.25	1.23
Mexico	−	7.86	5.63	5.61	5.90	6.09	**26**	**29**	28	**29**	**29**	1.40	1.40	1.33	1.29
Taiwan	~	7.47	6.05	6.07	6.06	6.23	27	26	25	28	27	1.23	1.23	1.23	1.20
India	+	7.46	6.02	5.91	6.10	6.16	28	27	27	27	28	1.24	1.26	1.22	1.21
Singapore	+	6.81	6.69	6.77	6.60	6.66	**29**	**22**	**21**	**24**	**24**	1.02	1.01	1.03	1.02
China		5.89	5.14	5.06	5.22	5.23	30	30	30	30	30	1.15	1.17	1.13	1.13
Ukraine		5.16	3.87	3.75	4.08	4.14	31	31	31	31	31	1.33	1.38	1.26	1.25

Note. *The plus sign (+) indicates rank rise from Method A to other methods; the minus sign (−) indicates rank drop; the swung dash (~) indicates rank fluctuation. **Country ranks in bold indicate that the rank position difference by the five methods ≥ 3. ***The different graying levels in the background indicate values between 1.00–1.25, 1.26–1.50, and >1.50.

the rankings when CP ratio was used to assess research impact. Switzerland and the United States took turns sharing the top spot depending on the counting method. Second, Switzerland, Denmark, and The Netherlands were all ranked very highly by CP ratio, but this was not the case in rankings by citation count. Other examples of dramatic rank rise from citation count to CP ratio were Israel and Finland, both located in the lower half of Table 2. These countries proportionally produced more highly cited papers, so their research impact per paper was significantly higher than that of the other countries.

On the other hand, some countries suffered a huge rank drop in CP ratio rankings. Japan was one example; it occupied third place in citation count, but dropped greatly in CP ratio rankings (between 15th and 18th). China suffered the most dramatic rank drop from citation count to CP ratio. When ranked by citation count, it was able to enter the top-10 most impactful countries. But when CP ratio was used, it fell to 30th no matter which counting method was used. Other countries with a similar rank drop include Russia and India; these countries proportionally produced less highly cited papers.

Comparing the CP ratio rankings by the five methods, 13 countries were ranked higher or at least the same when using straight counting (Methods B and C) and fractional counting (Methods D and E); 12 countries ranked lower or the same. Two countries saw fluctuating ranks, and four countries had exactly the same ranks by all methods. Compared to Tables 1 and 2, Table 3 shows markedly fewer countries with stable rank positions. In addition, in the previous two tables, the countries formed roughly two equally sized groups representing rank rise and rank drop due to method choices. However, this pattern was less clear in CP ratio rankings, which suggests that counting methods had a stronger divergent effect on CP ratio based rankings.

For some countries, rank changes due to counting method were relatively large. For instance, Singapore rose from 29th by whole counting (Method A) to 21st by straight counting

TABLE 4. Summary of statistical tests on number counts and country ranks.

Paper count	Paper numbers	Pearson	• All highly correlated; lowest coefficient value: 0.995 (A–B) ($p < 0.01$) • Method A was slightly less correlated with the others; B, C, D, & E were all completely correlated ($r = 1$).
		ANOVA	• Significant: $F = 25.595$ ($p = 0.000$)
	Country ranks	Spearman	• All highly correlated; lowest coefficient value: 0.987 (A–B) ($p < 0.01$) • Method A was slightly less correlated with the others; B–C & C–E were completely correlated ($\rho = 1$).
		Friedman	• Insignificant: $F = 1.147$ ($p = 0.887$)
Citation count	Citation numbers	Pearson	• All highly correlated; lowest coefficient value: 0.996 (A–B; A–C; A–D; A–E) ($p < 0.01$) • B, C, D, & E were all completely correlated ($r = 1$).
		ANOVA	• Significant: $F = 22.761$ ($p = 0.000$)
	Country ranks	Spearman	• All highly correlated; lowest coefficient value: 0.979 (A–C) ($p < 0.01$)
		Friedman	• Insignificant: $F = 0.928$ ($p = 0.920$)
CP ratio	CP ratio values	Pearson	• All highly correlated; lowest coefficient value: 0.973 (A–C) ($p < 0.01$)
		ANOVA	• Significant: $F = 68.277$ ($p = 0.000$)
	Country ranks	Spearman	• All highly correlated; lowest coefficient value: 0.953 (A–C) ($p < 0.01$)
		Friedman	• Insignificant: $F = 0.383$ ($p = 0.984$)

using corresponding author (Method C). Similarly, France and Brazil could both be ranked higher when straight counting and fractional counting were used. In contrast, Poland, Russia, Israel, and Finland were much higher ranked by the whole counting method.

Counting inflation in CP ratio. Similar to that in Tables 1 and 2 higher inflation was more or less concentrated in the lower half of the table. Countries with lower ranks experienced higher inflation, with a few exceptions (e.g., Japan, Singapore, and China). But different from the previous two tables, counting inflation in CP ratio based rankings no longer echoed the rank rise and rank fall from Method A to other methods.

Straight counting using first author and corresponding author particularly disfavored Hungary, Poland, the Czech Republic, and Russia; these countries had the highest inflation in A/B and A/C. The highest inflation was found in Poland in A/C (1.60).

For the first time, we found inflation lower than 1.00. The United States and Austria both experienced "counting deflation" in A/B and A/C. This means that the whole counting method underrated their research impact per paper. The two countries had proportionally provided the first author or corresponding author of more impactful papers than had the other countries.

Discussion

Some scholars have suggested that counting methods have only minor effects on research evaluation results and conclusions (Bourke & Butler, 1994 and May, 1997, both as cited in Gauffriau & Larsen, 2005). Our findings partially support this claim. However, the influence of counting methods was not the same for all evaluation measures (i.e., paper or citation counts or CP ratio). In paper count, the country rankings were not greatly affected by counting methods; however, in citation count and CP ratio, the country rank changes became larger and the change patterns were more complex than those in paper count. The clustering of countries of interchangeable ranks also was more complex in citation count and CP ratio. This suggests that counting methods differently affected the rankings.

Statistical tests further assisted in our observation of the influences from counting methods. Pearson's correlation analyses showed that all number counts by the five methods in paper, citation, or CP ratio based rankings were highly correlated at the .01 significance level; the correlation coefficient values were all greater than 0.995. In addition, Spearman's test on the country rankings showed that the five methods' ranking results also were highly correlated in each of the paper, citation, and CP ratio based rankings ($\rho s = >0.95$ at a significance level of 0.01). Given the high correlation in number counts and country rankings, however, analysis of variance did show that the selection of methods resulted in significant variation in number counting. In contrast, the Friedman test showed that variation in country rankings was insignificant. This suggested that the selection of counting method did influence number counting, which served as the basis for evaluation. Given the limited number of countries being sorted, however, the country rank variations did not achieve statistical significance.

We also conducted Spearman's test and the Friedman test on the ranking results to see whether the countries performed differently using the three evaluation measures. The results corroborated the aforementioned observation. The Spearman's correlation analysis showed that by each method, the all *paper count* and *citation count* based rankings were highly correlated. The *citation count* and *CP ratio* based rankings were somewhat correlated. However, the paper count and CP ratio based rankings were not statistically correlated. The Friedman test also showed that the ranking differences in the three tables were not statistically significant. We found the statistical results reasonable. Given the huge disparity of

DOI: 10.1002/asi

TABLE 5. Spearman's and Friedman tests of ranking results by each method.

Spearman's test	Paper vs. citation	• Always highly correlated; the lowest coefficient value: 0.909 (Method D) ($p < 0.01$)
	Citation vs. CP ratio	• Not correlated at $p < 0.01$
		• Correlated at $p < 0.05$; coefficient values between 0.394 (Method C) and 0.427 (Method A)
	Paper vs. CP ratio	• Not correlated at $p < 0.05$
Friedman test	Paper, citation, and CP ratio	• Method A: $F = 0.052$, $p = 0.974$
		• Method B: $F = 0.325$, $p = 0.850$
		• Method C: $F = 0.622$, $p = 0.733$
		• Method D: $F = 0.729$, $p = 0.695$
		• Method E: $F = 0.471$, $p = 0.790$

scientific development of the 30 countries, it was not surprising to see the high correlation between the paper count and citation count rankings because both were measures based on direct number counting. It also made sense to see the citation count and CP ratio rankings slightly correlated because both of them assessed research impact. The paper count and CP ratio rankings were not statistically correlated. The results showed that the evaluation of scientific research should not rely solely on a single measure or on highly homogeneous measures (e.g., all measures based on direct number counting).

Although the selection of counting methods did not greatly affect the entire country-level assessment results, our findings still support Gauffiau et al. (2008) in that straight counting and fractional counting were the better methods for research evaluation. First, having observed the various degrees of counting inflation in whole counting (Method A), which expanded number counts for the countries by varying degrees (some rather large), the validity of the method raised concerns. Second, in paper count, the highly consistent cluster-based, country-rank changes showed that the whole counting method obviously favored the more advanced Western countries. This suggests that the method may result in systematic bias in evaluation. In contrast, statistical testing showed that the number counts and country rankings from the other four methods were relatively consistent. As such, straight counting or fractional counting might be better choices for large-scale country-level assessment.

Conclusion

This study echoes the series of studies by Gauffriau and collaborators (Gauffriau & Larsen, 2005; Gauffriau et al., 2007, 2008) and used large-scale data to examine the consequences of counting-method uses. Specifically, it reports country rank changes and counting inflation in productivity and impact assessment using paper counts, citation counts, and CP ratio. Compared to studies using sample data (e.g., Glänzel, 2002; Golnabi & Mahdieh, 2006; Kao, 2009), the large and relatively complete publications data of the physics field can yield more robust understandings of what really happens in the international scientific research arena. Our analyses of citation count and CP ratio based results further bridge the gap in the existing literature where the effects of counting methods on impact evaluation have been rarely studied (Gauffriau & Larsen, 2005).

The pros and cons of whole counting, straight counting, or fractional counting in scientometric studies and research evaluation have been argued in the existing literature. While whole counting is considered more intuitive and easier for research evaluation, straight counting and fractional counting are more mathematically logical and have been advocated as the better choices (Gauffriau et al., 2008). Our analyses further support the use of the latter in larger scale, country-level assessment.

The differences between whole counting and the other methods in this study may have occurred for two reasons. First, institutions in some newly developed countries may have assumed leadership roles in international collaboration papers, but this was less discernible in whole counting in which every collaborating country received equal credit. Alternatively, it is likely that the rank differences were due to less participation of the newly developed countries in international collaboration. Future research needs to examine the extent of international and intranational collaboration of each country to understand the cause of the differences.

Further, although counting methods were shown to be of minor influence on country-level evaluation, whether this is still the case in institutional- or individual-level evaluations awaits investigation. In country-level assessment, the effect of counting method might have been greatly mitigated by other factors also influencing the quantity of papers or citations, such as each country's size, population, and the level of national and scientific development. However, the conclusion might not hold when more research institutions and researchers on a similar basis become the target of evaluation. The relationship between counting methods and subject disciplines also requires empirical investigation. Future studies should continue to examine the use and effects of counting methods at different levels and across subject disciplines.

References

Academic Ranking of World Universities. (n.d.). Ranking methodology. Retrieved from http://www.arwu.org/ARWUSubjectMethodology2010.jsp

Central Intelligence Agency. (n.d.). The world factbook: Appendix B: International organizations and groups. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/appendix/appendix-b.html#N

CME Group Index Services. (2010). Dow Jones total stock market indexes. Retrieved from http://www.djindexes.com/mdsidx/downloads/brochure_info/Dow_Jones_Total_Stock_Market_Indexes_Brochure.pdf

Egghe, L., Rousseau, R., & van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. Journal of the American Society for Information Science, 51(2), 145–157.

Gauffriau, M., & Larsen, P.O. (2005). Counting methods are decisive for rankings based on publications and citation studies. Scientometrics, 64(1), 85–93.

Gauffriau, M., Larsen, P.O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publications, cooperation and productivity measures in scientific research. Scientometrics, 73(2), 175–214.

Gauffriau, M., Larsen, P.O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2008). Comparisons of results of publication counting using different methods. Scientometrics, 77(1), 147–176.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. Library Trends, 50(3), 461–473.

Golnabi, H., & Mahdieh, M.H. (2006). Trend of laser research developments in global level. Optics and Laser Technology, 38(2), 122–131.

Higher Education Evaluation and Accreditation Council of Taiwan. (n.d.). Indicators. Retrieved from http://ranking.heeact.edu.tw/en-us/2010/Page/Indicators

Huang, M.-H. (2011). A comparison of three major academic rankings for world universities: From a research evaluation perspective. Journal of Library and Information Studies, 9(1), 1–25.

Kao, C. (2009). The authorship and country spread of Operation Research journals. Scientometrics, 78(3), 397–407.

Larsen, P.O. (2007, September 18). The state of the art in publication and citation counting [Abstract]. In Program of the 12th Nordic Workshop on Bibliometrics and Research Policy. Retrieved from http://www.iva.dk/nbw2007/files/1b_1_Peder_Olesen.pdf

Nudelman, A.E., & Landers, C.E. (1972). The failure of 100 divided by 3 to equal $33\frac{1}{3}$. American Sociologist, 7, 9.

Pravdić, N., & Oluić-Vukovic, V. (1986). Dual approach to multiple authorship in the study of collaboration/scientific output relationship. Scientometrics, 10(5), 259–280.

Quacquarelli Symonds. (n.d.). Data indicator. Retrieved from http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/data-indicators

DOI: 10.1002/asi