The *h*-Index in the Russian Scientific Citation Index

Viktor Bolotov, Natela Kvelidze-Kuznetsova, Vladimir Laptev, Svetlana Morozova

Viktor Bolotov

full member of the Russian Academy of Sciences, Professor, Director of Research and Development at the Russian Training Center, Institute for Educational Management, Russian Academy of Sciences. Address: 8, Pogodinskaya str., Moscow, 119121, Russian Federation. Email: vikbolotov@yandex.ru

Natela Kvelidze-Kuznetsova

CEO of the Fundamental Library, Herzen State Pedagogical University of Russia. Address: 48, Reki Moyki nab., Saint Petersburg, 191186, Russian Federation. Email: natela@herzen.spb.ru

Vladimir Laptev

full member of the Russian Academy of Sciences, Professor, Pro-rector for research at the Herzen State Pedagogical University of Russia. Address: 48, Reki Moyki nab., Saint Petersburg, 191186, Russian Federation. Email: laptev@herzen.spb.ru

Svetlana Morozova

Deputy CEO of the Fundamental Library, Herzen State Pedagogical University of Russia. Address: 48 Reki Moyki nab., Saint Petersburg, 191186, Russian Federation. Email: morozova@herzen.spb.ru

Abstract. The recent years have witnessed using numerical measurements of research work, apart from indicators of financial and administrative support and those of educational activity, as a ranking and monitoring criterion in assessing output of universities and scientific institutions. We analyze the h-index, one of publication activity indexes, considering it to be the most appropriate scientometric indicator that allows to smooth over many drawbacks of scientific output assessment by mere calculation of the number of publications or citations. We also discuss using the Web of Science and Scopus scientometric platforms to assess scientific productivity of Russian researchers. There are two main reasons why use of these platforms sometimes provides information that is inaccurate or incomplete: a) only publications in English are taken into account, and b) mostly natural science journals are selected. The paper demonstrates how the h-index is currently calculated in the Russian Science Citation Index and how the indexes can be optimized if the existing RSCI data is further processed and if new data is added. Based on the experiment of calculating the *h*-index for three authors on the RSCI platform, we have come to the conclusion that quantitative methods of assessing scientific output will most often be inaccurate and may only be applied together with peer reviews.

Keywords: scientometrics, Russian Science Citation Index, cumulative citation index, *h*-index, science efficiency, university ranking.

Decree of the President of the Russian Federation No. 599 "On measures to implement the state policy in education and science" dated 7 May 2012, a major role played by scientometric indicators in producing university rankings and in university monitoring, toughening of requirements to composition of thesis and expert councils under

the Higher Attestation Commission—all of these have provided for the State's official attention to the problem that has been extensively discussed in Russia over the past decade.

The objective of increasing the proportion of Russian research papers in the total number of publications in global science journals indexed in Web of Science database to 2,44% by 2015, as well as the objective of including five Russian universities to the top 100 global rankings by 2020 [On measures to implement the state policy...], have put an end to the discussion about the key criteria of research paper assessment.

Academic performance indicators calculated in citation indexes will prevail over peer reviews. The whole spectrum of academic achievements will be reflected in columns of numbers entitled as "number of publications", "cumulative citation index", "h-index", "impact factor".

> In order to perform comparative assessment of efficiency of academic performance in different fields of science, the major scientometric databases develop special analytical tools. Lately, such tools have been used more and more often all over the world in strategic science planning, assessment of performance of individual organizations, and making decisions on allocation of funds to specific projects and institutions [Moskalyova 2012].

Revolutionary changes in methods of academic performance assessment in Russia are a step towards integration of Russian research projects into the global research process where every researcher is familiar with scientometric tools.

This paper aims to demonstrate to researchers in humanities and social sciences how the h-index is calculated in the Russian Science Citation Index (hereinafter RSCI) at the present time and how the indicators may be improved if the existing RSCI bulk of information is processed or if new data is added. Clearly, this study and its results do not claim to be a global scientific research; they are rather part of the ongoing debate over scientometric indicators and methods of their calculation. It is important that discussion is not restricted to words, models, experiments, research for the sake of research. Instead, it should become a step towards resolving the pressing issues and problems encountered by Russian authors. Indeed, discussion "represents a series of statements made alternately by the participants. <...> [This] is one of the essential forms of communication, a powerful method of resolving controversial issues, and a style of learning, in its own way" [Ivin, Nikiforov, 1997. P. 90–91].

From integral indexes to the *h*-index

Without dwelling on the origins of scientometrics and the history of its development, which are described in hundreds of works (and this is in Russian only), we will point out the most important thing. Analytical

component of scientometrics is aimed at the triad of objects investigated: author—publication—journal (publishing). An array of data is processed using scientometric tools, and the major indicators obtained in this study are divided into two large groups: indicators reflecting the number of publications and those reflecting the number of citations per publication.

Scientometric data has always been the focus of attention in countries with developed market economies, especially in the United States, where universities have traditionally been centers for scientific research, while recognized importance of academic performance of universities provides a guaranteed inflow of promising students and a possibility of receiving state support.

> As international and national scientific citation indexes were put on stream, results of citation analysis became widely used by science administrators and chief executives of science foundations as one of the decision-making tools in assessing the role of the indexes in a specific field of research [Bredikhin, Kuznetsov, Shcherbakova, 2013. P. 5].

Cumulative (total) citation of publications of an author (a group of authors, a department, or an institution as a whole), a major indicator scientometrics is based on, was extremely unpopular with those who used it to assess scientific productivity of researchers or scientific organizations. An integral index calculated by merely counting the number of times this or that published work is mentioned does not always reflect the real weight of the publication, as it doesn't take into account:

- possible fraud, or 'paid' citations;
- self-citations (an author citing himself/herself, co-authors citing each other, a journal citing its own publications, Ph.D. students citing their research supervisors, etc.);
- "counter-citations" (citing a publication to disagree with the author).

"When we say that one paper cites another, it only means that the second paper is referred to in bibliography of the first one" [Pislyakov, 2011]. The cumulative citation index shows relevance of an author among other researchers but neither assesses the quality of her/his work nor reflects novelty of the results provided. Besides, specific nature of each field of science imposes specific traditions of citing one's own publications or those of other researchers.

Admitting that using the integral citation index to compare productivity of researchers and research teams working in different domains of science is inappropriate, scientists of various fields searched for an alternative calculation tool based on the third group of results obtained by combining the number of publications and the number

of citations. Their search provided for the following existing methods of calculation:

- *g*-index [Egghe, 2006]. Given a set of articles ranked in decreasing order of the number of citations that they received, the *g*-index is the (unique) largest number such that the top *g* articles received (together) at least *g*2 citations;
- *hg*-index [Alonso et al., 2010]. The *hg*-index is the *h*-index and the *g*-index combined;
- *e*-index [Zhang, 2009]. The *e*-index is an attempt to include publications ignored by the *h*-index;
- *AR*-index [Jin, 2007]. The *AR*-index includes the year of articles omitted by the *h*-index.

This series may be continued with a method proposed very recently by Ukrainian researchers:

> The new index is a mixed fraction modification of the h-index. Its integer component equals the usual *h*-index, while the fractional component shows progress of the author in reaching the next unit of the *h*-index. In accordance with the scientometric tradition, the new indicator is named the *Sh*-Index [Shtovba, Shtovba, 2011].

There are many more similar indicators incoherently used in scientometrics, most often only to prove wrong the data obtained with other tools.

The assessment method developed by Jorge Hirsch, physicist from the University of California (San Diego), was recognized as the most precise one by the academic community. Very soon, it came into use (together with the cumulative index) in official global citation indexes. The method was introduced in the laconic article *An Index to Quantify an Individual's Scientific Research Output* published in 2005 in *Proceedings of the National Academy of Sciences of the United States of America.*

The calculation formula is simple:

> Assume that *N* (*N* > 1) is the total number of publications of an author. Let us presume that this author has an *h*-index equal to *h* if *h* of his *N* research papers are cited at least *h* times each, while the rest (*N—h*) of the papers are cited no more than *h* times each. Otherwise speaking, an author has an *h*-index equal to *h* if (s) he has *h* papers published, each of them cited at least *h* times [Bredikhin, Kuznetsov, 2012. P. 151].

Moreover, the *h*-index also takes into account total number of citations, distribution of papers in time, and duration of research relevance reflected in citations in other publications. It is important

to note that the *h*-index is an integer, so its dynamics is low and its growth is determined by a significant set of factors.

A single brilliant publication (perhaps, created by a number of authors) with hundreds of citations will not allow the researcher to have a high h-index if there are no citations to other works of the same author, even though (s) he might have a pretty high cumulative citation index. Most likely, the *h*-index will be 1 or 2 in this case. In return, the *h*-index will provide a fair assessment of academic contributions made by authors who have dozens of citations to dozens of their papers created throughout many years. The *h*-index can also be applied to assess performance of an institution as a whole. Individual papers of individual authors recognized by the academic community (through multiple citations) will provide a high cumulative citation index for the employer institution. However, notably high values of the *h*-index will only be available to those organizations where most authors perform research projects recognized by their international counterparts every year, have their results published on a regular basis, and have their publications consistently referred to in research papers of other authors.

Thus, the index proposed by Hirsch replied in part to the challenges of the scientific world but couldn't resolve all of the problems. Specifically, it didn't close the gap between scientometric indicators in natural sciences and humanities (which is a topic for another research, as Hirsch never formulated that problem).

> Hirsch believed that using this measuring tool alone can only provide a rough approximation of a specific researcher's activity. So, the *h*-index should be mainly used to decide on grant allocation or to confirm the status of a scientist [Bredikhin, Kuznetsov, Shcherbakova, 2013. P. 267].

Published in 2005, the *h*-index became an integral tool of the two most recognized multidisciplinary citation indexes, Web of Science (Thomson Reuters)[1] and Scopus (Elsevier)[2], by the end of the decade. On these scientometric platforms, the *h*-index may be calculated for any group of documents: publications of an individual author or a group of authors (for any period), a selected bulk of articles, publications of an institution, a country, or a research team.

> The fact that bibliographic databases use the *h*-index as their indicator (less than two years after it was suggested!) demonstrates

---

[1] Web of Knowledge (citation index and scientometric tools). New York: Thomson Reuters http://isiknowledge.com; Web of Knowledge (information portal in Russian) http://wokinfo.com/russian/

[2] SciVerse. Scopus (citation index). Amsterdam: Elsevier B. V. http://www.scopus.com; Elsevier (official Russian website) http://elsevierscience.ru/products/scopus/

that it has become a generally accepted tool to measure academic performance. The *h*-index has provided the basis for a large amount of other indexes intended to rectify its drawbacks or to be used together with it. Moreover, a number of authors suggest that one-dimensional metrics is useless in the multidimensional space of bibliometrics [Bredikhin, Kuznetsov, Shcherbakova, 2013. P. 269].

Both Web of Science and Scopus are definitely recognized in the scientific world. Indicators calculated with their tools are used in global university rankings and in everyday academic activities. However, using these scientometric platforms to assess publication activity and academic productivity of Russian researchers doesn't always provide correct (or comprehensive) information for two main reasons: 1) these citation indexes only take into account publications in English; and 2) natural science journals are traditionally selected.

On top of that, many Russian researchers in humanities and social sciences find the global citation indexes and the whole scientometrics inappropriate, as the object of study is most often an article in an academic journal.

> The specific nature of historical and philological sciences consists in that they are targeted not only at acquiring new knowledge but also at supporting cultural traditions of the society, at preserving and developing its cultural heritage. Fundamental research has a predominantly monographic character here, while one of the most important forms of scientific effort is creation and renewal of basic resources required to maintain the culture and the humanitarian knowledge: multi-volume academic dictionaries, monuments of classical literature and folklore, commented publications of historical and ancient written records, catalogues of archeological materials, fundamental Internet resources (e. g. the Russian National Corpus), scientific expeditions, reference books and map data sources, scientific reports on expedition activities [Department of Historical and Philological Sciences of the Russian Academy of Sciences, 2013].

This argument also refers to social sciences, in particular to educational research, where monographic works are alternated with publications on methodology.

Russian Science Citation Index
A national science citation index should have become a substantive response to this and a number of other questions. National indexes are developed, in particular, in countries using non-English alphabets, like logographic systems or the Cyrillic script. The Russian national index was ordered by the Ministry of Education and Science of the Russian Federation. The project was launched in 2005, and in 2010

the Russian Science Citation Index (hereinafter the RSCI) came into effect officially and fully. From that moment, information on publication activity obtained with this index has been used in reports of the Ministry of Education and Science, in ranking and monitoring indicators, and in grant programs.

At first, the RSCI only included Russian scientific journals and articles published in them. These were journals from the *List of Russian Academic Journals Where the Key Scientific Results of Ph.D. and Sc.D. Theses Should Be Published* and scientific periodicals submitted to the platform of the Scientific Electronic Library (the content of which is used by the RSCI for indexation and calculation) by publishers themselves under the agreement with the Scientific Electronic Library (SEL). Later, abstracts of Scopus articles written by Russian authors were also added to the E-Library.

> These make over 500,000 bibliographic records about papers with at least one Russian author published in foreign journals, plus over 1,000,000 articles citing these papers. The data borrowed from Scopus embraces over 15 years, from 1996 until now. <...> In agreement with Web of Knowledge developers, Web of Science (WoS) citation indexes are available online for each RSCI article — of course, if the latter is included in WoS [Arefyev, Yeremenko, Glukhov, 2012. P. 67].

Naturally, problems of authors in humanities and social sciences were not resolved that way, as indexed content was again restricted to journals, while publications from global citation indexes increased indicators of researchers in natural sciences by adding the data that they had already reflected in their indicators through Scopus and Web of Science.

Following the demands of Russian researchers, the RSCI base was completed with bibliographic descriptions of author's abstracts and theses (over the past decade), as well as with books (monographs, textbooks, collections of articles, conference proceedings). Both bulks of documents belong to the catalogue of the Russian State Library (RSL). Publishers were enabled to submit structured data about any types of publications to the RSCI under the agreement with the SEL. All types of an author's publications cited in articles included in the Scientific Electronic Library are also reflected in the author's publications page.

At the moment, the RSCI includes [Yeremenko]:

- 2,800,000 articles from over 3,500 Russian scientific journals (since 2006), SEL;
- 680,000 Scopus articles by Russian scientists (since 1996), Elsevier;
- 780,000 theses (since 1983), RSL;

- 700,000 monographs and collections of articles (since 2003), RSL;
- 500,000 patents (since 1994), Federal Institute of Industrial Property;
- 3,000 reports on government contracts under Federal Special-Purpose Research and Technology Programmes (since 2007), Ministry of Education and Science;
- 100,000 diverse publications added by organizations, SEL.

"All in all, the RSCI comprises over 5,700,000 publications by Russian scientists. About 3,000 publications are added every day" [Ibid].

Thus, problems of Russian authors were partially solved. Large amounts of added information affected the linear indicators of researchers, i.e. total publications and total citations.

> The RSCI has turned out to be the most effective—and I believe will continue to do so—in enhancing the domestic visibility of humanities journals. Before this, our science used to be absolutely inaccessible, except <...> nine Scopus journals. Natural sciences had been propagated to some extent, but humanities did make a serious leap when our index appeared [Pislyakov, 2011].

The *h*-index data was also influenced by the bulks of added information. In most cases, inaccuracy of the *h*-index results could be tolerated and attributed to the period of RSCI establishment and adjustments if the index hadn't become part of official documents affecting scientific activities.

> The *h*-index is one of the scientometric indicators (excluding total publications and total citations) that have been recently treated as certain criteria of research paper (or thesis) quality and have become increasingly important, specifically in the context of various inspections initiated by the Ministry of Education and Science of the Russian Federation, being taken into consideration in mathematical modeling of quality management systems, in building social motivation, in applying social partnership principles in the labor market, in improving the quality of teacher working conditions, and in managing the organizational culture of universities [Nazarenko, 2013. P. 149].

The *h*-index and the Russian Scientific Citation Index in research in the Russian language

Problems associated with using the *h*-index are reflected in publications of Russian researchers. According to the Scientific Electronic Library, Russian Scientific journals have published around 200 papers, reviews and articles on the *h*-index over the past five years. An overwhelming number of works belongs to M. Nazarenko (Moscow State Institute of Radio-Engineering Electronics and Automation, Dubna branch), who

provides background information on the *h*-index and gives an overview of related research papers. A number of studies investigate how the *h*-index is applied to assess collective performance of a university or scientific institution. We should mention here articles by O. Mikhaylov and T. Mikhaylova from Kazan State Technological University [Mikhaylov, Mikhaylova, 2010, 2011; Mikhaylov, 2013]. Industry-specific research is also carried out; e. g. using the *h*-index in biology is discussed in papers by Y. Mokhnachyova and T. Kharybina [2013a, 2013b]. The *h*-index is examined closely in terms of practical application of scientometric indicators in Russian academic environment by some distinguished experts: V. Pislyakov, O. Moskalyova, Y. Granovsky, P. Arefyev. An important role is assigned to the *h*-index in articles of RSCI developers G. Yeremenko and V. Glukhov. A detailed analysis of the *h*-index and its use can be found in publications and reports prepared for scientific and practical conferences by researchers from Russian offices of Web of Science (P. Kasyanov, O. Utkin, S. Paramonov, V. Bogorov) and Scopus (V. Sobolev, G. Yakshonok). A significant importance is given to scientometric indicators and the *h*-index calculation in oeuvres by executives and employees of the National Electronic Information Consortium (NEICON), the longstanding performer of government contracts on providing science and education with electronic scientific information for the Ministry of Science and Education of the Russian Federation (A. Kuznetsov, I. Razumova, Y. Polnikova, etc.).

The abovementioned studies are often based on the bulk of RSCI publications and on the performed calculations accepted 'as is', with a proviso that RSCI data might not reflect all the publications. Thus, O. Mikhaylov and T. Mikhaylova say: "The *h*-index of an average researcher at our university is, frankly speaking, very low. <…> It's no use comparing this value to those of the world's leading universities" [Mikhaylov, Mikhaylova, 2011. P. 341]. However, the authors do not mention that the RSCI *h*-index may currently distort the actual publication activity of a specific scientist.

Scientometric research often prefers to apply global citation indexes, while Russian researchers, particularly those in humanities and social sciences, wait for the answers to questions associated with the RSCI and its indicators, which make a significant part of report-, competition-, ranking-, and grant-related documents of the Ministry of Education and Science and of other institutional bodies. Thus, in the two conferences organized by the RSCI SEL in 2013, only five (Science Online, May 2013) and two (Science Index, December 2013) speakers touched upon practical application of the RSCI (except employees of the Scientific Electronic Library), whereas global citation indexing data were used in eighteen and six reports, respectively[3].

---

[3] Science Online (electronic resources for science and education: proceedings of international conferences). Available at: http://elibrary.ru/project_scienceonline.asp

Having explored a number of his own RSCI indexes, A. Orlov posed the following question:

> In order to achieve accuracy of bibliographic descriptions and scientometric indicators applied, we should correct the information, line by line, on the basis of a preliminary research on properties of scientometric databases. Is it worth the time? [Orlov, 2013. P. 40].

It is the desire to answer this question that was the cause of our research.

Analysis of calculating the *h*-index in the Russian Scientific Citation Index

Let's consider some examples of using the *h*-index in the National Citation Index to prove the relevance of the problem we see as controversial.

The methodical approach is borrowed from a study on applying the *h*-index in compilation of rankings [Aleskerov et al., 2012]. The calculation was made as follows:

> The *h*-index calculation algorithm is rather simple: we sort all articles of an author (institution) from the highest to the lowest number of citations and go down the list until the position number of an article is higher than the number of its citations. The number of all the preceding articles is the *h*-index [Bedny, Sorokin, 2012. P. 26].

To understand special aspects of calculating the *h*-index in the RSCI, we should take into account the following facts about applying this scientific citation index.

- In the RSCI, the *h*-index is calculated only based on the information uploaded to the Scientific Electronic Library platform and validated, i. e. of all author's publications, the index counts only those with bibliographic descriptions in the RSCI SEL.
- Validated publications also include the ones "found in bibliographies"[4]. Criteria of selecting publications from the works cited are ambiguous. They are not stated on the RSCI website, and the RSCI customer support service says publications are selected based on the quality of reference description. However, there are some incomplete, imprecise bibliographic descriptions among works selected from bibliographies to author's publications lists, while bibliographically accurate references often remain in the works cited.
- The *Science Index. Organization* analytical superstructure, announced and presented in full format in 2013, allows responsi-

---

[4] Russian Scientific Citation Index. Available at: http://elibrary.ru/project_risc.asp.

ble representatives of institutions to add and modify information about publications of authors working in the respective institutions, so the calculations given below are not abstract anymore, and providing the results obtained in the experiment becomes possible through amending the bulks of bibliographic information.

• Calculation of the *h*-index in the RSCI is based on selective (incomplete) use of bibliographic data, even if it is included on the RSCI SEL platform (with no uniquely determined selection criteria), not to mention the data that is not included. That is why the reference group consisted of the authors whose publications and citations amounted to dozens or hundreds.

The experiment was built around three authors, employees of Herzen State Pedagogical University of Russia (Saint Petersburg). Selection was based on the following criteria:

• All authors are professors, employed at the present time, with their works having been actively published and cited since a long time ago, and ever more now.

• Each author represents one of the three domains of science: natural sciences, pedagogical sciences, and humanities. Russian authors in humanities and social sciences are poorly covered in the global citation indexes; the RSCI was designed to neutralize this inequality by providing publications in all fields of knowledge to the fullest extent.

• The authors were selected based on similar publication indexes: their number of publications is from 30 to 70, while their number of citations is from 300 to 400. Therefore, all the three authors were ranked the same (with allowance for publication standards of respective sciences) in these two linear indexes at the beginning of calculation.

We are not naming the authors here, but all the numerical data presented reflects their incumbent RSCI indicators.

Prior to starting the experiment, we suggested that only part of bibliographic data (publications and citations) reflected on the platform is taken into account while calculating the *h*-index in the RSCI. Discarded are:

• Data on author's publications referred to in article bibliographies but not included in the author's publications list.

• Citations of author's publications that are not matched with the bibliographic description in the author's profile during automated processing of data files, although 'manual' processing reveals they are comparable and do not contain any critical errors that would prevent matching the description in bibliography of the

Table 1

Indicator	Author A (natural sciences)	Author B (pedagogical sciences)	Author C (humanities)
Number of the author's RSCI publications	15	63	28
Number of the author's publications including articles found in bibliographies	35	65	30
Number of citations of the author's RSCI publications	26	24	5
Number of citation of the author's publications including articles found in bibliographies	87	30	7
Cumulative number of the author's citations	359	309	383
h-index	5	3	1
Cited publications taken into account in the h-index calculation (publications—citations, sorted from the highest to the lowest number of citations, publications with the same number of citations grouped together)	1–15 1–11 1–8 1–7 1–6 1–4 5–3 6–2 9–1	1–5 1–4 6–2 5–1*	1–2 5–1
Number of publications cited at least once	26	13	6
Number of publications with zero citation index	9	51	23

* For this author, the total number of citations in this cell doesn't match the number of citations including articles found in bibliographies, as it should. All figures are taken from the RSCI. The inconsistence may be caused by recalculation of data on the RSCI platform that was not yet completed at the moment of research.

citing article with that in the author's publications list (all citations containing critical errors or inconsistent with standard bibliographic descriptions are excluded from calculations to form a separate database).

The following operations were performed without using automated data processing facilities:

- We amended the authors' profiles before the experiment: we 'tied' (the RSCI term) all 'untied' publications and references and deleted publications not belonging to the authors.
- We analyzed the bulk of publications and citations of each of the three selected authors.
- We compared the information about publications and citations provided on the RSCI SEL platform.
- We made the comprehensive list of all the publications and citations presented on the platform.
- We suggested the potential number of publications and citations for each author, with due account of all the RSCI data analyzed.

Table 2

Indicator	Author A (natural sciences)	Author B (pedagogical sciences)	Author C (human-ities)
Number of publications including articles found in bibliographies	79	105	84
Number of citations of publications that had been attributed to the author's profile before the experiment*—in case if had been matched with all citations from the List of Citations of the Author's Works (the current indicator calculated in the RSCI is given in parentheses)	261 (87)	97 (30)	24 (7)
Number of citation of the author's publications including articles found in bibliographies	335	279	357
Number of citations with errors in the key components of bibliographic description (publication name, year or publication, source name), preventing their validation with the author's publication	24	30	26
Cumulative number of the author's citations**	359	309	383
h-index	10	9	9
Cited publications taken into account in the h-index calculation (publications —citations, sorted from the highest to the lowest number of citations, publications with the same number of citations grouped together)	1–53 1–33 1–19 3–15 2–14 1–13 3–10 2–8 2–7 2–6 2–5 4–4 12–3 12–2 23–1	1–34 1–21 1–19 2–13 1–12 1–11 1–10 1–9 4–8 3–6 3–5 2–4 7–3 11–2 21–1	1–154 1–34 1–18 1–12 1–11 2–10 1–9 1–8 2–5 2–4 5–3 13–2 32–1
Number of publications with zero citation index	8	45	21

* Number of the author's publications including the articles found in bibliographies.
** The cumulative number of citations remains the same, as it also includes both references with precise description and those with errors in the key components of bibliographic description.

- We suggested the hypothetical h-index each author would have had if the calculation had been based on the entire bulk of data available.

Table 1 shows the data reflected in the authors' RSCI profiles at the start of research.

Table 2 shows the results obtained after comparing the analyzed authors data manually.

Conspicuous is the difference between the initial indicators for the authors analyzed (Table 1): prior to manual comparison, the au-

thor in natural sciences had in his profile more descriptions of publications from cited works than the two others (20 descriptions for Author A and only two descriptions for Authors B and C). As more publications from bibliographies were added (which brought about more links "publication—citations" for Author A), the *h*-index calculation was also affected. Before the experiment was launched (which is equivalent to the current state of things in the RSCI), the *h*-index was calculated for Authors A, B, and C on the basis of the following publications—citations: 26–87, 13–30, 6–7, respectively. After all the citations contained in the authors' profiles were validated, the pattern looked like this: 26–261, 13–97, 6–24.

An analysis of data on the authors' citations revealed publications containing all the key components of bibliographic description and no critical errors that would have made these publications unidentifiable with automated data processing algorithms. That way, while the number of the author's publications including articles found in bibliographies was 35, 65 and 30 for Authors A, B, and C respectively prior to manual data comparison, it rose to 79, 105 and 84 after the manual comparison.

We also discovered references to the authors' works that had critical errors in their descriptions or lacked some of the key components of bibliographic description (publication name, year of publication, source name): 24 references for Author A, 30 for Author B, and 36 for Author C. Indeed, these references cannot be processed accurately in the automatic mode, but they can be easily validated with the cited publication by 'manual' data processing using the *Science Index. Organization* module tools.

Having matched all the data in the author's publications list and the data in the list of references to author's publications, we have formed the following bulk of data (all correct descriptions of references to the authors' works that had been missing in the authors' publications lists are defined as publications here) that can be used to calculate the h-index: Author A—79 publications, 335 citations; Author B—105 publications, 279 citations; Author C—84 publications, 357 citations.

Using the calculation algorithm described above (publications sorted from the highest to the lowest number of citations are given in Table 2), we obtained the following h-indexes: Author A—10 (used to be 5, doubled), Author B—9 (used to be 3, tripled), Author C—9 (used to be 1, multiplied by 9).

The indexes can be even higher if we manually correct the errors in the key components of reference descriptions and add data that was not indexed by the RSCI (provided that if new data on publications is added, data on references to such publications should be included in the Electronic Library).

Summary

The difference in indexes of authors in natural sciences versus humanities and social sciences, which is a current feature of the RSCI, is due not only to peculiarities of publication activity and the citation culture in different domains, but also to different results of the RSCI data processing. We may suggest that this discrepancy is caused by different structuredness of data downloaded from Scopus (which provided for most of the data on publications in natural sciences in the RSCI) and other sources. The suggestion has the following basis: Author A's RSCI publications included 6 works from Scopus, together accounting for 24 citations. Authors B and C didn't have any Scopus-indexed publications. Perhaps, Author A's profile included more publications 'retrieved from bibliographies' because all the references were made in academic journals with strict format requirements, so the accuracy of algorithm processing was higher.

Having compared large groups of authors through the example of the Herzen University (indexes were taken from the *Science Index. Organization*), we established that the *h*-index range is 8–21 for natural science faculties, 5–14 for pedagogical faculties, and 4–9 for humanitarian faculties. Applying manual data processing methods may neutralize these differences.

Unlike in global citation indexes, developers of the RSCI include works found in bibliographies in author's publications list (for example, in the Web of Science, references to non-indexed publications can only be found with the Cited Reference Search, and author's publications list doesn't include cited publications that have not been indexed yet), but this process is not finished yet, while selection criteria are sometimes confusing.

Creators of the RSCI SEL got it right when they introduced the *Science Index. Organization* analytical tool. With its help, all the hypothetical calculations provided in our experiment can be implemented in real life, thus evening the authors' odds of presenting accurate data on their publication activity in official documents.

The *Science Index. Organization* module is used on the commercial basis, as an annual subscription. The principle of the tool is that 'manual' operations are partially delegated to institutions themselves. An official representative of an institution has the right to edit and validate bibliographic descriptions existing on the RSCI SEL platform (e. g. those found in bibliographies), whereupon these descriptions earn the status of being officially included in the RSCI SEL, and the publications are automatically taken into account while calculating the *h*-index for the author. Besides, an institution representative may also add bibliographic descriptions of publications the RSCI SEL has no information about yet. The institution guarantees credibility and accuracy of the information added, which is then validated by an RSCI operator.

Automated data processing is inevitably associated with errors on scientometric platforms. This is true both for global citation index-

es, where selected bulks of data often have to be further processed 'manually' for accurate results and feedback service is used to report about errors, and for the RSCI.

> The RSCI budget makes it impossible for operators to always process manually all the data submitted. That is why such procedures as analyzing references or tying publications and references to authors, institutions and journals are automated in the RSCI. Quite naturally though, a lot of data is hard to analyze, in particular due to the poor culture of constructing bibliographies in most Russian journals [Russian Scientific Citation Index]5.

The RSCI is a young citation index. Unlike the global scientometric platforms, it is open and available to the public, which improves the dynamics of changes significantly. What has already been done on this scientometric platform allows us to hope for many of the abovementioned problems to get solved very soon.

In conclusion, we should note that quantitative methods to assess scientific impact are never absolutely accurate, and so is the *h*-index. Quantitative indicators should not become the exclusive assessment criterion. "As soon as you begin using a formal index to assess a content-related process, the efforts soon become focused on increasing the index by any means, instead of developing the process",—says A. Parshin, full member of the Russian Academy of Sciences[6].

An accurate, unbiased approach to scientific impact assessment with due account of opinion polarity and drawbacks of each individual procedure can only be provided by a combination of methods, including peer review, scientometric indicators, discrimination between indexes for different fields of knowledge, engagement of scientists in discussion of every new assessment method, testing and endorsement of each technique, and development of efficient assessment tools (such as the RSCI) —not as one-time projects but as multidimensional scientific impact assessment tools for which appropriate support an funding are provided.

References

1. Aleskerov F., Katayeva Y., Pislyakov V., Yakuba V. (2013) Otsenka vklada nauchnykh rabotnikov metodom porogovogo agregirovaniya [Using the Threshold Aggregation Method to Assess Researchers' Contributions]. *Large-Scale Systems Control: A Collection of Oeuvres*, no 44, pp. 172–189.
2. Alonso S., Cabrerizo F., Herrera-Viedma E., Herrera F. (2010) Hg-Index: A New Index to Characterize the Scientific Output of Researchers Based on the Hand G-Indices. *Scientometrics*, vol. 82, no 2, pp. 91–400.

---

5 Available at: http://elibrary.ru/project_risc.asp.

6 Science in figures. Available at: http://www.gazeta.ru/science/2013/11/11_a_5745593.shtml

3. Arefyev P., Yeremenko G., Glukhov V. (2012) Rossiyskiy indeks nauchnogo tsitirovaniya—instrument dlya analiza nauki [The Russian Science Citation Index as a Science Analysis Instrument]. *Bibliosfera*, no 5, pp. 66–71.

4. Bedny B., Sorokin Y. (2012) O pokazatelyakh nauchnogo tsitirovaniya i metodakh ikh primeneniya [On Scientific Citation Indices and Methods of Their Application]. *Vysshee obrazovanie v Rossii*, no 3, pp. 17–28.

5. Bredikhin S., Kuznetsov A., Shcherbakova N. (2013) *Analiz tsitirovaniya v bibliometrii* [Citation Analysis in Bibliometrics]. Novosibirsk, Moscow: Institute of Computational Mathematics and Mathematical Geophysics SB RAS, NEICON.

6. Bredikhin S., Kuznetsov A. (2012) *Metody bibliometrii i rynok elektronnoy nauchnoy periodiki* [Bibliometric Methods and the Electronic Marketplace of Scientific Periodicals]. Novosibirsk: Institute of Computational Mathematics and Mathematical Geophysics SB RAS, NEICON.

7. Bureau of the History and Philology Department, Russian academy of Sciences (2013) *Ob otsenke nauchnoy deyatelnosti institutov gumanitarnogo profilya* [On Assessing Research Activities of Humanities Institutes]. Available at: http://www.saveras.ru/archives/3691 (accessed 30 January 2014).

8. Egghe L. (2006) An Improvement of the H-Index: The G-Index. *ISSI Newsletter*, no 2 (1), pp. 8–9.

9. Hirsch J. E. (2005) An Index to Quantify an Individual's Scientific Research Output. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 15, pp. 16569–16572. Available at: http://arxiv.org/abs/physics/0508025 (accessed 23 February 2014).

10. Ivin A., Nikiforov A. (1997) *Slovar po logike* [Logics Vocabulary]. Moscow: Tumanit, VLADOS.

11. Jin B. H. (2007) The AR-Index: Complementing the H-Index. *ISSI Newsletter*, vol. 3, no 1, pp. 6. Available at: http://sci2s.ugr.es/hindex/pdf/Jin2007.pdf (accessed 23 February 2014).

12. Mikhaylov O. (2013) Razmyshleniya ob otsenke nauchnoy deyatelnosti [Speculations on Assessing Research Performance]. *Large-Scale Systems Control: A Collection of Oeuvres*, no 44, pp. 144–160.

13. Mikhaylov O., Mikhaylova T. (2010) Indeks Khirsha v otsenke deyatelnosti uchyonogo v natsionalnom issledovatelskom universitete [The *h*-Index in Assessing Researchers' Activities in a National Research University]. *Vestnik Kazanskogo tekhnologicheskogo universiteta*, no 11, pp. 485–487.

14. Mikhaylov O., Mikhaylova T. (2011) "Khirshemetriya" v Kazanskom natsionalnom issledovatelskom tekhnologicheskom universitete [Hirsch Metrics in the Kazan National Research Technological University]. *Vestnik Kazanskogo tekhnologicheskogo universiteta*, no 18, pp. 338–341.

15. Moskalyova O. (2012) *Poverit indeksom nauku: lektsii* [Measuring Science with an Index. A course of lectures]. Available at: http://www.gazeta.ru/science/2012/12/19_a_4896245.shtml (accessed 30 January 2014).

16. Mokhnachyova Y., Kharybina T. (2013a) Nauchnaya produktivnost rossiyskikh uchyonykh v oblasti biologii, nauk ob okruzhayushchey srede i ekologii v period 2002–2001 gg. po baze dannykh Web of Science [Scientific Productivity of Russian Researchers in Biology, Environmental Sciences, and Ecology in 2002–2001 based on the Web of Science Database]. *Informatsionnye resursy Rossii*, no 2 (132), pp. 7–13.

17. Mokhnachyova Y., Kharybina T. (2013b) Publikatsionnaya aktivnost rossiyskikh uchyonykh v oblasti biologii, nauk ob okruzhayushchey srede i ekologii v period 2002–2001 gg. [Publication Productivity of Russian Researchers in Biology, Environmental Sciences, and Ecology in 2002–2001]. *Vestnik Rossiyskoy akademii nauk*, vol. 83, no 10, pp. 867.

18. Nazarenko M. (2013) Indeks Khirsha liderov rossiyskogo indeksa nauchno-go tsitirovaniya po chislu publikatsiy [The *h*-Index of Russian Science Citation Index Leaders by the Number of Published Works]. *Mezhdunarod-ny zhurnal prikladnykh i fundamentalnykh issledovaniy,* no 6, pp. 149–150.

19. Orlov A. (2013) Dva tipa metodologicheskikh oshibok pri upravlenii nauch-noy deyatelnostyu [Two Types of Methodology Mistakes in Scientific Management]. *Large-Scale Systems Control: A Collection of Oeuvres*, no 44, pp. 32–54.

20. Pislyakov V. (2011) *Nauka cherez prizmu statey* [Science through the Prism of Research Papers]. Available at: http://polit.ru/article/2011/12/21/pislya-kov_2011/ (accessed 30 January 2014).

21. President of the Russian Federation (2012) O merakh po realizatsii gosu-darstvennoy politiki v oblasti obrazovaniya i nauki. Ukaz ot 7 maya 2012 g. № 599 [On Measures to Implement the State Education and Science Policy. Decree No 599 of May 7, 2012]. *Rossiyskaya gazeta,* May 9, 2012. Available at: http://www.rg.ru/2012/05/09/nauka-dok.html (accessed 30 January 2014).

22. Shtovba S., Shtovba Y. (2013) Obzor naukometricheskikh pokazateley dlya otsenki publikatsionnoy deyatelnosti uchyonogo [A Review of Scientometric Indicators Used to Assess Publication Activities of Scientists]. *Large-Scale Systems Control: A Collection of Oeuvres*, no 44, pp. 262–278.

23. Shtovba S., Shtovba Y. (2011) Sh-indeks—novaya drobnaya modifikatsiya indeksa Khirsha [sh-Index as a New Fractional Modification of the h-Index]. *Nauchnye trudy Vinnitskogo natsionalnogo tekhnicheskogo universiteta*, no 3. Available at: http://archive.nbuv.gov.ua/e-journals/VNTU/2011_3/2011–3_ru.files/ru/11sdsmoh_ru.pdf (accessed 30 January 2014).

24. Yeremenko G. (2013) *RINTs i SCIENCE INDEX: novye vozmozhnosti dlya avtorov, izdateley i nauchnykh organizatsiy* [RSCI & SCIENCE INDEX: New Opportunities for Authors, Publishers, and Scientific Organizations]. Available at: http://elibrary.ru/projects/science_index/conf/2013/presenta-tions/eremenko.pdf (accessed 30 January 2014).

25. Zhang C.-T. (2009) The E-Index, Complementing the H-Index for Excess Citations. *PLoS ONE*, vol. 4, no 5. Available at: http://www.plosone.org/arti-cle/info%3Adoi%2F10.1371%2Fjournal.pone.0005429 (accessed 30 January 2014).