ARTICLE

https://doi.org/10.1057/s41599-020-0521-5

OPEN

Check for updates

In the race for knowledge, is human capital the most essential element?

Laura Sinay[®] ^{1,2⊠}, Rodney William (Bill) Carter² & Maria Cristina Fogliatti de Sinay³

Clarivate Analytics, managers of Web of Science, publishes an annual listing of highly cited researchers. The opening sentence of the 2019 report asks "Who would contest that in the race for knowledge, is human capital that is most essential?". They go on to state that "talent—including intelligence, creativity, ambition, and social competence (where needed) outpaces other capacities such as access to funding and facilities". These contradict previous findings, according to which other factors are possibly more influential than human capital. Using Clarivate Analytics' database for 2018, we investigated which factors are most relevant in development of scientific knowledge. Rather than human capital alone, we found that language, gender, funding, and facilities introduce bias to assessments and possibly prevent talent and discoveries from emerging. We also found that the profile of the highly cited scholars, as established by Clarivate Analytics, is so narrow that it may compromise the validity of scientific knowledge, because it is biased towards the perception and interests of male scholars affiliated with very-highly developed countries where English is commonly spoken and of their sponsors. This highly cited scholars accounted for 76% of the random sample analyzed, absent were women from Latin-America, Africa, Asia, and Oceania, and scholars affiliated with institutions in low-human-developed countries. Also, 98% of the published research came from institutions located in very-highly developed countries. These findings provide evidence that challenges the view that 'talent is the primary driver of scientific advancement'. This is important because search engines, such as Web of Science, can modify their algorithms to ensure the work of scholars that does not fit the currently dominant profile can have their importance increased so that their findings can more equitably contribute to knowledge development. This, in turn, will increase the validity of scientific inquiry.

¹ Federal University of the Rio de Janeiro State—UNIRIO, Rio de Janeiro, RJ, Brazil. ² University of the Sunshine Coast, Sunshine Coast, QLD, Australia. ³ Private consultant, Rio de Janeiro, RJ, Brazil. ^{Se}email: laura.sinay@unirio.br

Introduction

very year Clarivate Analytics, the company that manages Web of Science, lists highly cited researchers from analysis of their database of published peer-reviewed articles. Early in the 2019 report, they asked "Who would contest that in the race for knowledge, is human capital that is most essential?" (Clarivate Analytics, 2018a, p. 5). They amplified this question by stating that "talent—including intelligence, creativity, ambition, and social competence (where needed)—outpaces other capacities such as access to funding and facilities" (Clarivate Analytics, 2018a, p. 5). This understanding contradicts previous findings that suggest other elements used in algorithms by automated search engines such as gender, language and funding, might be significantly restricting the development of scientific knowledge (Sinay et al., 2019a; Angermuller and Hamann, 2019).

Using a randomly selected group of scholars listed on the Clarivate Analytics' database for 2018, we explored whether other factors might be as, or more important than human capital in the race for knowledge development. We hypothesized that if human capacity and talent are the most essential criteria for knowledge development, then the profile of prominent scholars that emerge from databases should be equally distributed among gender, level of country development, access to funding and languages spoken in the country where scholars are affiliated.

Other elements affecting the race for knowledge

While Clarivate Analytics states that talent is the fundamental element in the race for knowledge (Clarivate Analytics, 2018a, 2018b), other scholars have identified additional factors that may as well be affecting science development. Among those, gender is one of the most frequently described elements (Ceci and Williams, 2011; Moss-Racusin et al., 2012; Nielsen, 2016; Cooper et al., 2019). Female scholars' influence on science have been found to be negatively affected by stereotypes (Cooper et al., 2019), family commitments (Ceci and Williams, 2011), implicit favoritism of academic decision makers for promoting men over women (Moss-Racusin et al., 2012) and by females' lower research productivity in comparison to males (Nielsen, 2016; Mairesse and Pezzoni, 2015; Mayer and Rathmann, 2018). In this context, currently and worldwide <30% of scholars are women (UNESCO, 2019).

Another element frequently cited as an impediment for knowledge development is language (Angermuller and Hamann, 2019). As English has been stated as the language of science (American Society for Cell Biology, 2012), those scholars that are not fluent in English tend to be disadvantaged in the race for knowledge development (American Society for Cell Biology, 2012; Drubin and Kellogg, 2017). A frequent complaint is "that manuscript reviewers often focus on criticizing their English, rather than looking beyond the language to evaluate the scientific results and logic of a manuscript" (Drubin and Kellogg, 2017, p. 2).

Access to funding has also been found to influence the race for knowledge (Jacob and Lefgren, 2011; Vlăsceanu and Hâncean, 2015; Cattaneo et al., 2016; Hottenrott and Lawson, 2017). The degree of this influence, however, seems to vary according to the amount and extent of funding (Kem, 2010; Rosenbloom et al., 2015).

The algorithm currently used by automated search engines is also believed to influence the development of science (Adam, 2002; American Society for Cell Biology, 2012; Hicks et al., 2015; Sinay et al., 2019a). This algorithm was proposed initially in 1955 by Prof. Eugene Garfield as a tool to disseminate and retrieve scientific literature (Garfield, 2007) and to recognize authorship (i.e. who influenced who in the scientific world) (Garfield, 1956). It was, then, a procedure developed to systematize the existing scholarly literature, but now serves many different purposes (Garfield, 2007). For example, it currently guides library-purchasing policy for journals, author assessments on where to publish, measurments of scientific productivity and for determining research funding and tenure of scholars and, of course, it is also used to order the results of scholarly performance (Garfield, 1965, 2007; Adam, 2002; Baneyx, 2008; Hall, 2015; Sinay et al., 2019a).

Garfield's algorithm was introduced to the public in 1964 by the Institute for Scientific Information via the Web of Knowledge platform (later Web of Science) (Clarivate Analytics, 2018b). Today, it underpins most of the scholarly search engines. Some of the fundamental assumptions of the algorithm of science, however, are believed to influence the development of knowledge by overestimating the importance of some profiles of scholars over others (Hicks et al., 2015; Bol et al., 2018; Sinay et al., 2019a).

One of these assumptions relates to Bradford's law (Bradford, 1934), according to which a "small percentage of journals account for a large percentage of the articles published in a specific field of science" (Garfield, 1965, p. 112). Due to this understanding, when using Web of Science, scholars can only access "research literature linked to a rigorously selected core of journals" (Clarivate Analytics, 2019b, p. 3). A similar approach is used by Scopus (Elsevier, 2019). Google Scholar may be the exception as, in 2018, their webpage explicitly stated that all academic contributions from 'sensible' websites, including 'gray' literature, were incorporated on their database (Sinay et al., 2019b). However, in 2020 they specifically list their sources: "articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites", making no reference to 'gray' literature (Google Scholar, 2019, 2 paragraph 2).

In the context of Bradford's law, when a researcher makes an automated search, results tend to only include articles published in high-impact journals. Hence, despite their quality, papers in journals with moderate to low-impact factors, often in the arts and social sciences, are excluded from consideration by the algorithm (Clarivate Analytics, 2018a; Elsevier, 2019). As most high-impact journals are in English (Bortolus, 2012), also excluded are works published in other languages (Amano et al., 2016).

Another issue is that the algorithm of science is set to estimate scholars' productivity based on the number of papers they publish in high-impact journals (Noorden and Chawla, 2019; Google Scholar, 2019; Clarivate Analytics, 2020a). While this omits innovative work published in languages other than English, it also disadvantages scholars from less developed countries, usually with more lecturing responsibilities and less time for developing research (Boyer, 1990; American Society for Cell Biology, 2012). It also introduces gender bias with women tending to publish less than men (Mairesse and Pezzoni, 2015; Nielsen, 2016; Mayer and Rathmann, 2018). More importantly, it ignores that in the race for 'productivity', 'top' scholars are publishing an implausible number of works, reaching as high as 3000 scholarly publications per author (Sinay et al., 2019a).

The algorithm measures a scholar's success based on the number of times their work has been cited within the related database (Garfield, 1970; Clarivate Analytics, 2020b), ignoring negative citations and introducing potential bias towards well-established schools of thought that reinforce a paradigmatic position rather than innovation and discovery (Sinay et al., 2019a; Agnieszka et al., 2019). This is particularly controversial when works are being authored by thousands of scholars, which fosters cross-citations (Noorden and Chawla, 2019).

Lastly, the algorithm is based on Merton's technical norms of science (Merton, 1942; Merton and Garfield, 1979), which states "race, nationality, religion, class, and personal qualities are as such irrelevant" for science development (Merton, 1942, p. 53). While we wish this were so, we are well aware that the perspective of scholars limits the "problems that are studied, the methods and technologies that are applied and, in the case of social sciences, the moral values from where the observer considers the research" (Sinay et al., 2019a, p. 553). While this assumption has been denied for long and by many (Brightman, 1939; Ihde, 2002; Angermuller, 2017), it remains in the algorithm of science by default through lack of buffering of the influence of personal characteristics (Clarivate Analytics, 2020b).

There is an obvious divergence between Clarivate Analytics' understanding of the most impactful elements on the race for knowledge and what has been published in the scholarly literature. The first states talent is the key element; the later asserts gender, language, funding and the algorithm of science largely define who is likely to win 'the race'.

Method

A random sample collected from the Clarivate Analytics database of Highly Cited Researchers for 2018 was used in this research to explore other factors that might be as, or more important than human capital in the race for knowledge development. The database is organized in 22 areas of knowledge¹ and comprises 3539 authors, listed alphabetically.

A random number generator² was used to arbitrarily select three numbers between 1 and 26, corresponding to the 26 letters of the English alphabet. The numbers retrieved were 1, 11, and 15, which correspond to the letters A, K, and O, respectively. For determining the sample, within each area of knowledge, we selected the first three scholars whose last name starts with each of these three letters. Scholars whose first name was not available or only the initials were available were excluded from the sample because identifying gender would not have been possible. These scholars were substituted by the next one on the list (by doing so, it was possible to determine the gender of all the scholars on the dataset). Also, there were cases in which there were less than three scholars per area of knowledge whose last name started with the selected letters; in these cases, too, the sample was completed with the next scholar on the list. By this process, we randomly selected 198 scholars. Thus, the chosen scholars are not necessarily the 'top' scholars among the Clarivate Analytics' dataset. They just represent a random selection.

This research focuses on exploring the influence of gender, access to funding (sponsorship), journal of publication, number of co-authors, language and level of development of country of primary affiliation in the race for knowledge development, because these were identified in previous research as being relevant. Other factors, such as religion and ethnicity, are likely to also be relevant, yet information on these is not available.

Information regarding country of primary affiliation of scholars was gathered on the Clarivate Analytics dataset and was "specified by the Highly Cited Researchers themselves" (Clarivate Analytics, 2019a, p. 19). These data were analyzed based on the Human Development Report 2019 (United Nations Development Program, 2020) to determine the level of human development of the countries where top-scholars work. It was also used to analyze the likelihood of top-scholars being fluent on English, as previous research indicates that scholars fluent on English are more likely to be cited (Drubin and Kellogg, 2017). In this context and in this research, it was considered that English is commonly spoken in countries where at least 30% of the population speaks the language. This analysis was based on the World Fact Book (Central Intelligence Agency, 2020) and, when additional information was necessary, on Wikipedia.

Gender of the sample was discovered using Google Photos. Information on the percentage of male and female scholars per country is based on UNESCO's Institute for statistics (UNESCO, 2020).

The latest 'highly cited publication' up to December 2018 of each scholar was used to collect data about number of citations, of co-authors and of sponsors, and journal of publication. These publications were found using Web of Science, which automatically highlights the 'highly cited publications'; i.e. those that "reflect the top 1% of papers by field and publication year" (Clarivate Analytics, 2020a, 2020b, p. 6). Information regarding number of citations was collected in August 2019.

Information related to sponsorship was gathered from publications as stated by the authors. Where authors did not disclosure funding, research was considered to have been developed without sponsorship.

Descriptive statistic methods were used to analyze gender, access to funding (sponsorship), journal of publication, number of co-authors, language and level of development of country of primary affiliation. This included estimating average, minimum (min) and maximum (max) sample values, standard deviation (SD), median and mode for distribution of parameters analyzed.

Results

Scholar profiles of the analyzed sample (N = 198) indicate that about half (49.5%) of the authors are affiliated with institutions in the USA. The remaining 100 scholars are affiliated with 29 other countries. Therefore, of the 192 countries recognized by the United Nations, 83% were not represented in the sample.

The same percentage (i.e. 83%) of scholars in the sample are male. UNESCO has information on gender distribution for 21 of the 30 countries to which top-scholars in the sample are primarily affiliated (Fig. 1). The average percentage of male scholars in these countries is 67 (SD = 8; min = 51; max = 85). Within the analyzed sample, 90 scholars are primarily affiliated to the 21 countries for which UNESCO (2020) has gender distribution data. Within this sample (N = 90), 80% are male scholars. That is, not only are there fewer women than men in science, but it is also more difficult for female scholars to reach the top.

Of the analyzed scholars, 98% are primarily affiliated with 'very-highly developed' countries.³ Within the sample, only one scholar is affiliated with a country ranked as having 'medium-human development' (Pakistan). However, this scholar is also affiliated with the University of Pretoria (University of Pretoria, 2020) in South Africa, which is classified by UNESCO as a high-developed country. No scholars within the analyzed sample are primarily affiliated with countries classified as having 'low-human development'.

Among the countries analyzed, China, which is classified as having a 'high' development level, has the highest number of scholars (Fig. 2). We infer therefore that level of development is, at least for the analyzed sample, more influential in the race for knowledge than the total number of scholars.

Language follows a similar tendency with 93% of the analyzed scholars being affiliated with countries where English is spoken by at least 30% of the population. Of the remaining scholars, five are affiliated with Spain, four with Japan, two with China, two with Turkey and one with Brazil. China and Japan are among the three countries with the highest number of scholars (Fig. 2). In comparison, the UK has less than half the total number of scholars of Japan and about 80% less than China. Still, there are 15 scholars affiliated to the UK within the sample. Therefore, English fluency considerably affects knowledge development.



Fig. 1 Average gender distribution among scholars (based on UNESCO, 2020). UNESCO's (2020) data were used to specify values of blue and read bars. The dashed yellow line represents the ideal situation in which half of the scholars are female and the straight yellow line represents percentage of male scholars within the complete analysed sample. The green dashed line represents the average percentage of male scholars at the countries considered in the dataset but based on UNESCO's data's (2020) and the straight green line represents the same percentage but based on the reduced analyzed sample.

The influence of language and level of development is even more severe when we only consider the women's sample (N=33). All 33 female scholars in the sample are primarily affiliated with 'very-highly developed' countries where English is commonly spoken. There are no female scholars affiliated to Latin-American or African countries, and only two are affiliated to countries in Asia, both from South Korea. That is, language and level of development are key filters on female scholars' influence in the race of knowledge.

The number of authors per publication varied between 1 and 2834. Considering the whole sample (N = 198), the average number of authors per publication is 56.3 (SD = 229; median 11; mode 4). The number of authors per publication of 1% of the sample is greater than three times the value of the standard deviation (outliers). If we exclude these from the sample (N = 196), then the average number of authors per publication

drops to 36.5, but still with a high standard deviation (SD = 84). That is, top scholars tend to publish with many authors, which facilitates increasing the number of publications and reinforcement of bias caused by cross-citation.

For sponsorship, 90% of the authors identified who sponsored their research. Among the 23 authors that did not report sponsorship, only one (0.5% of the sample) is not affiliated to a very-highly developed country. That is, with this exception, scholars in the sample either had access to '1st world' facilities or were sponsored.

The number of sponsors per research paper varied between 0 and 221. If we consider the whole sample (N = 198) and that non-identification of sponsors means no sponsorship, which may not be always the case, then the average number of sponsors per publication is 7.26 (SD = 22.2; median 11; mode 1). The number of sponsors of 1% of the sample is greater than three



Number of top-scholars in the sample
Estimated number of researchers per country according to UNESCO

Fig. 2 Number of total scholars and of top-scholars on the analyzed sample organized per country. UNESCO's (2020) data were used to discover the number of scholars per million inhabitants and Worldometer's (2020) data to determine the total population of each country.

times the standard deviation (outliers). If we exclude these from the sample (N = 196), then the average number of sponsors is 5.08 (SD = 5.3).

The 198 papers analyzed were published in 130 academic journals. All papers and all journals were published in English. The average number of publications per journal is 1.5 (SD = 1.55, median 1; mode 1). Yet, the number of publications of 23% of the sample is greater than three times the standard deviation (outliers). If we exclude these (N = 152), then the average number of publications per journal is 1.23 (SD = 0.57). The two most represented journals were Science with 5.1% of the publications (10 articles), and Nature and its family of journals having 17.6% of the publications considered (11 articles in Nature plus 24 in its family journals). While there are at least 40,000 scholarly journals actively publishing, nearly half (45%) of the papers of the sample were published on just 10 families of journals (Table 1). That is, papers published in languages other than English or in less impactful journals, independent of their scientific contribution, are less frequently cited. Also, the race for knowledge development is being influenced by a very reduced percentage of the active journals.

Table 1 Journals most used for publication by most-cited scholars.

Journal	Number of publications
Nature	35
Science	10
IEEE	11
Lancet	6
Nucleic Acids Research	6
New England Journal of Medicine	5
JAMA—Journal of the American Medical	4
Association	
Astrophysical Journal Supplement Series	5
Food Chemistry	3
Blood	3

Is human capital the most influential ingredient for 'top scholars'?

Seven peculiarities of the sample are critical for considering whether human capital is indeed the most essential characteristic of 'top scholars' and most essential in the race for knowledge. (1) Of the 198 scholars that compose the analyzed sample, there are no women from Latin-America, Africa, Asia or Oceania. (2) The database has no scholars affiliated with institutions in 'low-human-development' countries, and only one scholar is affiliated with a 'medium-development' country. (3) There are no papers or journals written in languages other than English. (4) At least 90% of the research outputs analyzed have been sponsored. (5) Of the research, 94% has been developed in institutions located in 'very-highly-developed' countries. (6) Of the scholars, 83% are males affiliated with 'very-highly-developed' countries where English is commonly spoken. (7) Almost half of the most visible researchers are authors affiliated with institutions in the USA.

Consequently, either scholars affiliated with less developed countries (especially women and those that do not have English as their first language) are less talented, intelligent, creative, ambitious and have less social competence than men affiliated with institutions located in 'very-highly developed' countries where English is commonly spoken, or language, gender, funding and facilities play a major role in the race for knowledge. As there is no scientific evidence to support that talent is related to gender, level of development of country of affiliation or languages spoken, one can conclude that the opening statement of Clarivate Analytics is misleading. Instead, the analyzed data demonstrates that gender, language, funding (which is related to sponsorship) and facilities (which largely depends on the level of development of country of primary affiliation) are the essential factors in the race for knowledge, which corroborates previous findings.

Why does it matter to prove that talent is not the essential element in the race for knowledge?

"In a time when the influence of fake news prevails over the influence of facts, science becomes, more than ever, the most reliable source of knowledge" (Sinay et al., 2019a, p. 549). Yet, not everybody believes in science. According to Butts (2016), "despite extensive efforts at public science education, polling over the past 30 years has consistently shown that about 40-45 percent of [USA] Americans believe that humans were supernaturally created in the past 10,000 years... [and] Similar results have been found for beliefs regarding anthropogenic climate change" (Butts, 2016, p. 286). While some (non) believers have negligible impact on development, others can cause real harm. Denial of climate change, for example, can threaten life itself (Butts, 2016), while "the high-profile anti-vaccination movement has become influential in parts of the US and Europe" (Bedford, 2019, p. 16), potentially bringing back fatal illnesses that were nearly extinct in the recent past (Roberts, 2019).

People's belief in science is related to trust in the scientific method used to achieve verifiable results, and this trust is somewhat dependent on people's perception of (lack of) predisposition and convergence of understandings. As such, knowledge is expected to be developed based upon multiple lines of evidence with different sponsors and in diverse settings, as this may mitigate potential biases (Denzin and Lincoln, 1994) while amplifying the possible questions and valid answers (Sinay, 2008). Amplifying questions means studying the same issue from different perspectives and with different values, technologies and interests.

Despite its importance, this diversity does not seem to be occurring in science. Instead, a very restricted profile of scholars (men fluent in English and affiliated with institutions located in very-high developed countries) is directing knowledge, along with their sponsors. Hence, issues such as, say, crop production in Africa or infant attention deficit hyperactivity disorder (ADHD), tend not to be studied by those closer to the problem, in these cases, Africans, mothers, and teachers. The restricted profile of scholars can be easily explained: knowledge is power and, since the beginning of modern science, power has largely been in the hands of the men of the wealthiest nations. Hence, these men have historically directed knowledge development. While the limited profile of scholars can be historically explained, it diminishes people's confidence in science. If the number of people not believing in science were small, this would not be a problem, but the number of people ignoring scientific advice in the context of the SARS-CoV-2 and the consequent fatalities cannot be ignored. If society wants to be ready to react to threats such as the COVID-19 pandemic, the number of people trusting and following science needs to increase.

Increasing the number of scholars with different profiles on knowledge development remains an enormous challenge, but women and men of less developed countries are already doing science and their contribution needs to be acknowledged. Therefore, it is now just a matter of adjusting the algorithms used in scholarly search engines (Adam, 2002; Baneyx, 2008; Hall, 2015; Sinay et al., 2019a), which are based on misleading assumptions that neglect most of the work published by female scholars and scholars affiliated with less developed countries especially where English is not commonly spoken (Sinay et al., 2019a). If algorithms are amended and the databases are augmented, the discoveries made by a heterogenic group of scholars will start to influence the development of knowledge (Sinay et al., 2019a).

Reconsidering the question posed in this paper, it matters to prove that Clarivate Analytics' assumption is misleading, because they manage the algorithm of Web of Science; hence, have the power (and the knowledge) to update it. This, in turn, will increase the validity of science.

Data availability

2018 Clarivate Analytics dataset is available at: https://hcr. clarivate.com/resources/archived-lists/. The sample used for this work is available at: http://research.usc.edu.au/vital/access/ manager/Repository/usc:29846.

Received: 12 November 2019; Accepted: 16 June 2020; Published online: 08 July 2020

Notes

- 1 The 22 areas of knowledge: Agricultural Sciences Biology & Biochemistry Chemistry; Clinical Medicine Computer Science Economics & Business Engineering; Environment/Ecology Geosciences; Immunology; Materials Science Mathematics; Microbiology; Molecular Biology & Genetics; Neuroscience & Behavior; Pharmacology & Toxicology; Physics; Plant & Animal Sciences; Psychiatry/Psychology; Social Sciences; Space Science and Cross-field.
- 2 The random number generator is available at https://www.random.org
- 3 UNESCO classifies levels of development as low, medium, high and very-high.

References

- Adam D (2002) News feature: the counting house. Nature 415:726–729 Agnieszka G, Siudem G, Gagolewski M (2019) Should we introduce a dislike
- button for academic articles? J Assoc Inf Sci Technol 71:221–229 Amano T, González-Var JP, Sutherland WJ (2016) Languages are still a major barrier to global science. PLoS Biol 14:1–8
- American Society for Cell Biology (2012) San Francisco dEclaration On Research Assessment. https://sfdora.org/read/. Accessed 3 Apr 2019
- Angermuller J (2017) Accumulating discursive capital, valuating subject positions. From Marx to Foucault. J Crit Discourse Stud 15:414–425
- Angermuller J, Hamann J (2019) The celebrity logics of the academic field. The unequal distribution of citation visibility of Applied Linguistics professors in Germany. J Discourse Stud 1:77–93

- Baneyx A (2008) "Publish or Perish" as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history. Arch Immunol Ther Exp 56:363–371
- Bedford H (2019) Measles: how a preventable disease returned from the past. BBC News. https://www.bbc.com/news/health-47800438. Accessed 3 Apr 2019
- Bol T, Vaan MM, Rijt AD(2018) The Matthew effect in science funding Proc Natl Acad Sci USA 115:4887–4890
- Bortolus A (2012) Running like alice and losing good ideas: on the quasicompulsive use of English by non-native English speaking scientists. Ambio 41:769-772
- Boyer E (1990) Scholarship reconsidered: priorities of the professoriate. The Carnegie Foundation for the Advancement of Teaching, New York
- Bradford S (1934) Sources of information on specific subjects. Engineering 137:85-86
- Brightman R (1939) The social function of science. Nature 24:179-89
- Butts C (2016) Why I know but don'tbelieve. Science 354:286-187
- Cattaneo M, Meoli M, Signori A (2016) Performance-based funding and university research productivity: the moderating effect of university legitimacy. J Technol Transf 41:85–104
- Ceci S, Williams W (2011) Underÿstanding current causes of womenas underrepresentation in science Proc Natl Acad Sci USA 68:3157–3162
- Central Intelligence Agency (2020) The World Factbook. https://www.cia.gov/ library/publications/the-world-factbook/. Accessed 3 Aug 2019
- Clarivate Analytics. (2018a) Highly cited researchers 2018: identifying top talent in the sciences and social sciences. https://clarivate.com/webofsciencegroup/ blog/highly-cited-researchers-2018-identifying-top-talent-in-the-sciencesand-social-sciences/. Accessed 3 Aug 2019
- Clarivate Analytics. (2018b) The concept of citation indexing: a unique and innovative tool for navigating the research literature. https://clarivate.com/ webofsciencegroup/essays/concept-citation-indexing/. Accessed 3 Aug 2019
- Clarivate Analytics. (2019a) Highly cited researchers: identifying top talent in the sciences and social sciences. https://clarivate.com/webofsciencegroup/wpcontent/uploads/sites/2/dlm_uploads/2019/11/WS370932093-HCR-Report-2019-A4-RGB-v16.pdf. Accessed 3 Aug 2019
- Clarivate Analytics. (2019b) Web of Science Platform. https://clarivate.com/ webofsciencegroup/solutions/webofscience-platform/. Accessed 3 Aug 2019
- Clarivate Analytics (2020a) Essential science indicators. https://clarivate.com/ webofsciencegroup/solutions/essential-science-indicators/. Accessed 3 Aug 2019
- Clarivate Analytics. (2020b) InCites benchmarking & analytics: understanding the metrics. https://clarivate.libguides.com/incites_ba/understanding-indicators. Accessed 3 Aug 2019
- Cooper T, Noa A, Bar-Ilan J, Rabin MS (2019) Women in academia: a bibliometric perspective. Inf Res http://www.informationr.net/ir/24-4/colis/colis1926. html. Accessed 3 Aug 2019
- Denzin N, Lincoln Y (1994) Handbook of qualitative research. Sage Publications, Thousand Oaks
- Drubin D, Kellogg D (2017) English as the universal language of science: opportunities and challenges. Mol Biol Cell 23:1399
- Elsevier (2019) Content policy and selection. https://www.elsevier.com/solutions/ scopus/how-scopus-works/content/content-policy-and-selection. Accessed 3 Aug 2019
- Garfield E (1956) Citation indexes: new paths to scientific knowledge. Chem Bull 43:1112
- Garfield E (1965) Science Citation Index—answers to frequently asked questions. Rev Int Doc 32:112–116
- Garfield E (1970) Citation indexing for studying science. Nature 227:669-671
- Garfield E (2007) The evolution of the Science Citation Index. Int Microbiol 10:65–69
- Google Scholar (2019) Stand on the shoulders of giants. https://scholar.google.com/ intl/en/scholar/about.html. Accessed 3 Aug 2019
- Hall C (2015) A citation analysis of tourism recreation research. Tour Recreat Res 35:305–3009
- Hicks D, Wouters P, Waltman L, Rijcke S, Rafols I (2015) Bibliometrics: the Leiden manifesto for research metrics. Nature 520:429–431
- Hottenrott H, Lawson C (2017) Fishing for complementarities: research grants and research productivity. Int J Ind Organ 51:1-38
- Ihde D (2002) How could we ever believe science is not political? Technol Soc 24:179–89
- Jacob B, Lefgren L (2011) The impact of research grant funding on scientific productivity. J Public Econ 95:1168–1177
- Kem S (2010) Analytic model for academic research productivity having factors, interactions and implications. Cancer Biol Ther 12:949–956
- Mairesse J, Pezzoni M (2015) Does gender affect scientific productivity? Rev Écon 66:65–113

- Mayer S, Rathmann J (2018) How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. Scientometrics 79:1663–1693
- Merton RK (1942) A note on science and democracy. J Legal Political Sociol 1:115-126
- Merton RK, Garfield E (1979) Citation indexing: its theory and application in science, technology, and humanities. ISI Press, Philadelphia
- Moss-Racusin C, Dovidio J, Brescoll V, Graham M, Handelsman J (2012) Science faculty as subtle gender biases favor male students Proc Natl Acad Sci USA 109:16474–16479
- Nielsen M (2016) Gender inequality and research performance: moving beyond individual-meritocratic explanations of academic advancement. Stud Higher Educ 41:2044–2060
- Noorden RV, Chawla DS (2019) Hundreds of extreme self-citing scientists revealed in new database. Nature 572:578-579
- Roberts M (2019) Vaccines: low trust in vaccination 'a global crisis'. BBC News. https://www.bbc.com/news/health-48512923. Accessed 3 Aug 2019
- Rosenbloom JL, Ginther DK, Juhl T, Heppert JA (2015) The effects of research & development funding on scientific productivity: academic chemistry, 1990–2009. PLoS ONE 10:1–23
- Sinay L (2008) Modelling and forecasting cultural and environmental changes. PhD Dissertation, University of Queensland
- Sinay L, Sinay MCF, Carter RW, Martins A (2019a) Reviewed assumptions for the algorithm of Garfield. Rev Admin Univ São Paulo 54:548–549
- Sinay L, Sinay MCF, Carter RW, Martins A (2019b) Who is writing the science behind the management of protected areas?. Ann Ecol Environ Sci 3:18–24
- UNESCO (2019) Women in science. Institute for Statistics of UNESCO. http://uis. unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en. pdf. Accessed 3 Aug 2019
- UNESCO (2020) How much does your country invest in R&D? http://uis.unesco. org/apps/visualisations/research-and-development-spending/. Accessed 3 Aug 2019
- United Nations Development Program (2020) Human development report 2019 beyond income, beyond averages, beyond today: inequalities in human development in the 21st century. http://hdr.undp.org/sites/default/files/ hdr2019.pdf. Accessed 3 Aug 2019
- University of Pretoria (2020) Prof M (Mujahid) Abbas. https://www.up.ac.za/ mathematics-and-applied-mathematics/article/1942189/prof-m-mujahidabbas. Accessed 3 Aug 2019
- Vläsceanu L, Hâncean MG (2015) Policy incentives and research productivity in the Romanian higher education. An institutional approach. In: Curaj MLA (ed) The European higher education area. Springer, Cham
- Worldometer (2020) Countries in the world by population. https://www. worldometers.info/world-population/population-by-country/. Accessed 3 Aug 2019

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2020