DOI: 10.59519/mper7101

Rethinking Research Metrics: The Performance Impact Score (PIS) as a Novel Alternative to the H-index

Haris Memisevic

University of Sarajevo – Faculty of Educational Sciences

Abstract

Providing an objective ranking of scientists based on their merit is a rather challenging task. Numerous factors complicate this endeavor, raising difficult questions about how such evaluations should be conducted. The goal of the present paper is to introduce a new measure for evaluating researchers' performance, the Performance Impact Score (PIS), which combines both productivity and research impact. In this study, I compared the PIS with h-index scores for 108 researchers at the University of Sarajevo. More specifically, I examined the overlap between the top 20 researchers according to the PIS and the top 20 researchers according to Web of Science, Google Scholar, and SCOPUS h-indexes. The concordance rate was 65% for Google Scholar and Web of Science, and 55% for SCOPUS. The analysis highlights the importance of considering both productivity and impact when evaluating researchers' rankings. It is also evident that the top 20 researchers vary across different metrics. The results further demonstrate that creating a fair and just ranking system requires going beyond the data available in bibliometric databases, particularly in cases where researchers rank highly by one metric but perform poorly when evaluated by another.

Article history

Received: 5.1.2025.

Accepted: 23.1.2025

Keywords:

scientometrics;

bibliography;

researchers;

evaluation

DOI: 10.59519/mper7101

Introduction

The field of scientometrics, which involves the quantitative analysis of scientific research, is increasingly being applied to support the recognition of researchers and the allocation of grants. Stakeholders and grant authorities now use scientometric benchmarks to evaluate research performance when ranking applicants for academic positions or determining eligibility for funding. This approach allows for the quantification of research productivity and citation impact, providing a numeric assessment of a researcher's publication achievements (Friedmacher et al., 2018). As a result, evaluating and valuing research output has become a crucial task, driven by the demands of funding agencies, promotion committees, and employers. The measurement of academic achievements has thus become an integral part of academic life (Kelly & Jennions, 2006; Lane, 2010). However, the question remains what factors should be taken into consideration in the evaluation of researchers. For instance, some authors emphasize that evaluations based on criteria such as quality, productivity, visibility, and impact are essential for assessing a scientist's contributions to their field (Sternberg, 2018). At present, there is no internationally accepted standard procedure for evaluating researchers based on their scientific output (Memisevic et al., 2017). To date, the H-index (Hirsch, 2005) has been widely regarded as the gold standard for evaluating researcher performance. Proposed by Jorge E. Hirsch in 2005, the H-index has become a widely accepted metric for assessing both the scientific output and impact of researchers. It is defined as the highest number of publications that have at least that many citations. For example, a researcher has an h-index of 20, if 20 of her papers have at least 20 citations. This dual consideration of both quantity (number of publications) and quality (citations per publication) makes the H-index a valuable tool for assessing individual researchers' contributions to their fields. Studies have demonstrated that the h-index correlates strongly with other measures of scientific quality, indicating its reliability as an indicator of a researcher's influence within their discipline (Bornmann & Daniel, 2005; Bornmann et al., 2008; Memisevic et al., 2022). Specifically, Bornmann et al. (2008) found that the h-index aligns with both objective and subjective assessments of scientific quality across different fields. This consistency across various evaluative frameworks has contributed to its widespread adoption as a "gold standard" in academic evaluation.

However, despite its widespread acceptance, the h-index is not without its criticisms. Critics argue that it can be an unreliable metric for certain fields or types of research, especially those with lower citation rates or where publication practices vary significantly (Akhtar, 2024). The h-index can be influenced by factors such as collaboration patterns and publication practices, which may not accurately reflect an individual's contributions (Costas & Bordons, 2007; Hirsch, 2019). For example, the h-index does not account for the number of authors on a paper, potentially skewing the perceived impact of a scientist's work (Rong et al., 2021). Additionally, the h-index can be easily inflated through excessive self-citations, which distorts its accuracy in reflecting a researcher's true scientific impact. Also, the h-index does not account for the number of co-authors,

DOI: 10.59519/mper7101

making it less reliable for assessing individual contributions in multi-author publications. As a result, it tends to underestimate the contributions of researchers who publish independently or with a small number of co-authors (Barnes, 2017).

Recently, several new metrics have been proposed to address some of the shortcomings of the H-index. One such metric is the z-score (Zerem, 2017), which eliminates discrepancies in evaluating scientific output by combining the Author Score and Author Citation Score, preventing discrimination against new publications, and discouraging the inclusion of authors with minimal or no contribution. Some authors have criticized the z-score due to oversimplification of author's contribution, thus making it unsuitable for all academic disciplines (Bates, 2017). Another proposed measure as an alternative to h-index is a g-index (Egghe, 2006). The g-index is a metric designed to assess a researcher's scientific impact by emphasizing the total citation performance of their most cited papers, rather than focusing on a specific threshold like the h-index. Although its purpose was to capture the influence of highly cited papers, the g-index can be inflated in cases where a single paper, authored by many researchers, receives an extraordinarily high number of citations. For example, a paper with 100 co-authors and 100,000 citations could significantly boost the g-index, even if the individual researcher had minimal contribution, making the g-index less reliable for evaluating individual impact in such cases.

The mission of finding an improved alternative to H-index is a continuous task. Given the limitations of existing metrics, I would like to propose a new metric, the Performance Impact Scale (PIS) that might remove some of the limitations of existing metrics. The PIS is designed to evaluate researchers' overall contributions by combining both their productivity (quantity of work) and impact (quality of work). It integrates the two critical dimensions of academic performance: the number of publications and the influence those publications have within the scientific community, as indicated by H-index. Thus, h-index is part of the PIS formula, but productivity is a new dimension, which has a particular formula. The productivity of a researcher is calculated by considering the total number of publications and the author's contribution to those publications. Specifically, for each researcher, I calculated the percentage of papers in which they are either the first author or the corresponding author (whichever percentage is higher). This is important because being the first or corresponding author often signifies a more substantial role in research.

The productivity formula is as follows:

Productivity = Number of Publications × Percentage of Papers as First or Corresponding Author (whichever number is higher, if a researcher has not a single paper in which she is the first or corresponding author, a value of 1 will be assigned for calculation)

The PIS is than calculated according to following formula:

PIS = Productivity x H-index

DOI: 10.59519/mper7101

The goal of the present paper is to compare the performance of the PIS with h-indexes from three databases, Web of Science (WOS), Google Scholar (GS), and SCOPUS. I also wanted to determine the amount of overlap in the top 20 researchers from the University of Sarajevo (UNSA) according to these different indices. Lastly, I wanted to examine the outliers more closely, and determine what do they tell us about concrete profile of the researcher.

Method

To begin, I selected a sample of 108 top researchers from the University of Sarajevo (UNSA), based on their profiles in Google Scholar (GS). For each researcher, I gathered data on their H-index scores from three major bibliometric databases: Google Scholar, Web of Science (WOS), and SCOPUS. In addition to these traditional metrics, I calculated the Performance Impact Score (PIS) for each researcher. The PIS is a novel metric that incorporates both productivity and research impact, with productivity being particularly reliant on the researcher's role as the first or corresponding author. This information is readily available through the WOS database, which automatically provides the percentage of papers in which the researcher is listed as the first or corresponding author. However, this feature is not directly available through Google Scholar or SCOPUS, making it easier to calculate the PIS using WOS data alone. By compiling these metrics, I aimed to conduct a comprehensive evaluation of the researchers' academic performance across different dimensions.

Statistical analysis

I calculated correlations between the PIS, GS H-index, WOS H-index, and SCOPUS H-index. Multivariate outliers were identified by examining the Mahalanobis distance values. Any values greater than three were considered significant outliers, indicating data points that deviate substantially from the multivariate distribution.

Results

I first present correlations between the PIS and h-indexes in Google Scholar, WOS, and SCOPUS databases.

Table 1. *Correlations between the PIS and h-indexes*

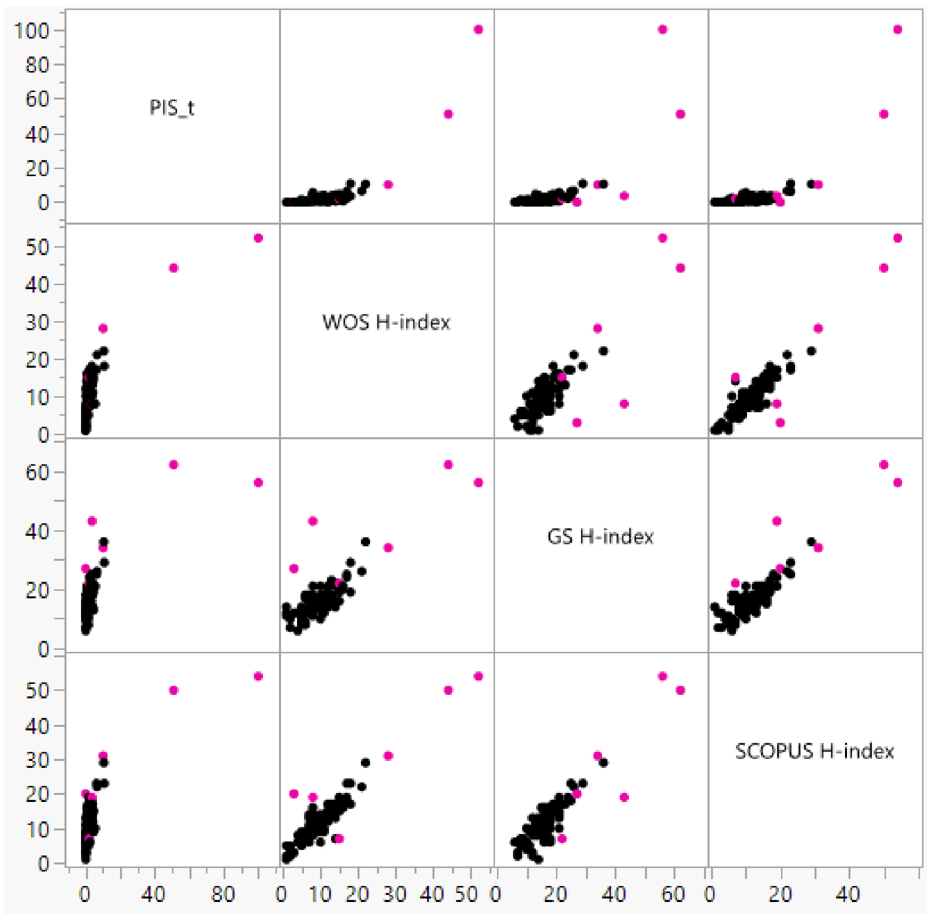
	PIS	**GS H-index**	**WOS H-index**	**SCOPUS H-index**
PIS	1	-	-	-
GS H-index	.74	1	-	-
WOS H-index	.80	.83	1	-
SCOPUS H-index	.78	.90	.92	1

Note. all p's < .001.

DOI: 10.59519/mper7101

As expected, the highest correlation of the Performance Impact Score (PIS) was observed with the WOS h-index, given that the WOS h-index is an integral part of the WOS formula. Notably, the shared variance between PIS and the WOS h-index is 64%, indicating a strong overlap. However, 36% of the PIS score is explained by the Productivity Index, highlighting the unique contribution of productivity in assessing researcher performance beyond traditional citation metrics.

Although the size of the correlations is high and statistically significant, there were six significant outliers (Mahalanobis distances >3) that require additional inspection. They are identified by the different color in the Figure 1.



Note. I used min-max normalization for the PIS

Figure 1. *Scatterplot of Correlations between the PIS and H-indexes*

Closer inspection of these outliers provides a better understanding of the data trends. Two outliers were identified as they both had high PIS and high h-indexes according to all databases (much higher than the rest of the sample). Two outliers exhibited relatively low h-index scores in the Web of Science (WOS) database compared to their higher h-index scores in SCOPUS and Google

DOI: 10.59519/mper7101

Scholar (GS). In one case, the reason for this discrepancy was clear: the researcher primarily published in journals indexed by SCOPUS but not by WOS. In the second case, the WOS profile was incomplete and lacked several articles, particularly those authored by large consortium groups (e.g., EUROASPIRE) in which the researcher was involved. Upon manual review, I discovered two separate WOS profiles for the researcher. Furthermore, one of these profiles had incorrectly merged the researcher's publications with those of an international collaborator from the same consortium group (whose name appeared in brackets). However, that profile was inaccurate as it contained both, the articles of UNSA researcher and international researcher thus providing inaccurate profile. One outlier had low SCOPUS h-index score in comparison with high GS and WOS h-index score. Further investigation into this unusual discrepancy revealed that the SCOPUS profile of this outlier was fragmented into two separate profiles under the same name, which had not yet been merged in the SCOPUS database. Last outlier had low PIS as compared to h-indexes. This likely stems from the fact that the author's productivity score was low (very few articles in which the researcher was the first or corresponding author). A total of eight researchers ranked in the top 20 according to all indices (the PIS and H-indexes), while 34 researchers were ranked in the top 20 based on at least one of the indices.

Discussion

The development of new scientometric indices is motivated by the need to overcome the limitations of existing metrics, such as the h-index and g-index, which may fail to fully capture the complexities of scientific output and impact. The goal of the present study was to introduce and test a new measure for evaluating researcher performance, the Performance Impact Score (PIS), which combines both the productivity and impact of a researcher. By comparing the PIS with traditional bibliometric measures like the h-index from Google Scholar, Web of Science (WOS), and SCOPUS, the study aimed to determine how well the new metric correlates with existing ranking systems. I found moderate to strong correlations among all indices, a finding similar to previous studies (Memisevic et al., 2019). As expected, the PIS was most strongly correlated with the WOS h-index, as the latter forms an integral part of the PIS formula. Notably, 64% of the PIS variance was explained by the WOS h-index, while 36% of the PIS score was attributed to the productivity index, which reflects the number of publications and the researcher's role in those publications. The h-index has notable limitations, particularly its failure to account for a researcher's productivity, which is an important factor that should not be overlooked in evaluations. Researcher productivity, alongside impact, is often correlated with both academic prestige and teaching effectiveness, as well as funding and institutional recognition (Chua et al., 2002) and as such should be taken into account in researcher evaluations.

The use of productivity as a criterion for evaluating researchers can be critiqued for potentially disadvantaging those who, despite making significant contributions to research, do not have publications where they are the first or corresponding author. This issue can arise particularly for researchers involved in large collaborative projects, where leadership roles may not always be

DOI: 10.59519/mper7101

distributed in a way that allows every contributor to assume these prominent authorship positions. In such cases, assessing productivity solely based on these roles may fail to fully recognize their valuable contributions. This is especially relevant in interdisciplinary or multi-author studies, where roles are often more collaborative and less hierarchical. By focusing on first or corresponding authorship, we might risk underappreciating the intellectual contributions of researchers who play important roles in other aspects of the research, such as conceptualization, data analysis, or writing.

However, despite these concerns, the inclusion of productivity as a measure can be defended by emphasizing that the first and corresponding authorship roles often signify greater responsibility and leadership within a research project. Encouraging researchers to actively seek these positions serves as a motivation to take ownership of their work, which is an essential aspect of academic development. These roles represent a commitment to driving research forward, ensuring quality, and contributing meaningfully to the academic community. By incorporating first and corresponding authorship into the evaluation process, the system incentivizes researchers to take initiative, fostering a more engaged and accountable research environment. While not all researchers may have had the opportunity to lead studies in the past, this evaluation criterion underscores the importance of taking on such responsibilities in shaping their future academic careers, ultimately leading to a more proactive approach to scientific contribution.

Other authors have also recognized the challenge of multiauthor papers and their evaluation. The introduction of the hα-index seeks to extend the h-index by incorporating the number of co-authors into the calculation, thereby providing a more equitable assessment of a researcher's impact in collaborative settings. This index recognizes that the presence of multiple authors can affect the citation dynamics of a paper and attempts to adjust for this factor (Leydesdorff et al., 2019). Another way to alleviate this challenge is through detailed description of each author’s contribution to the paper. A clear understanding of each author's role in a publication is essential for accurate assessments of productivity and impact (Sathian et al., 2014). However, this becomes increasingly challenging when there are many authors involved, especially in large collaborative research efforts. While fields such as experimental physics may provide some exceptions where authorship is more clearly defined due to the nature of the work, in many other disciplines, it can be difficult to pinpoint the exact contributions of each author. A common but imprecise solution is to assign equal contributions to all authors, but this method often fails to accurately reflect the actual distribution of work. Thus, although not ideal, I believe the production index proposed in this paper might serve as a good proxy for actual author’s contribution.

Science, along with many other fields of human’s endeavor, is a highly competitive field (Lüscher, 2018). In this study I wanted to determine who are the top 20 scientists at the University of Sarajevo. The answer depends on which database you consult. The overlap analysis revealed that the top 20 researchers according to the PIS were not always identical to those ranked in the top 20 by the other indices, showcasing that different metrics can produce varying rankings. Only eight

DOI: 10.59519/mper7101

researchers were among the top 20 regardless of the metrics used. On the other hand, there were 34 researchers who were among the top 20 according to one or more metrics used. This again highlights the importance of considering both productivity and impact when evaluating researchers, as these metrics can offer different insights into a researcher's overall contribution.

While the PIS offers a more nuanced approach by accounting for productivity alongside citation impact, the study also identified several outliers that pointed to discrepancies between the metrics. For example, discrepancies between WOS, SCOPUS, and GS profiles were observed in several cases, such as incomplete or inaccurate WOS profiles or researchers with multiple profiles in SCOPUS. These findings reinforce the notion that bibliometric databases need to be used carefully and that a combination of metrics is essential for more accurate evaluations. Incorrect scientific profiles and errors were already reported in the literature (Memisevic et al., 2019; Selivanova et al., 2019). However, it is mostly up to the researchers to keep their profiles updated and accurate, although some errors might happen independent of the authors.

Furthermore, this study underscores the limitations of traditional bibliometric indicators such as the h-index. While widely accepted, the h-index has been critiqued for its inability to fully capture the scope of a researcher's impact, especially in fields with lower citation rates or when publication practices significantly differ. The number of citations is not a measure of research quality per se. Probably we all are aware of some highly cited, controversial papers, which are based on inaccuracies and controversies and not on scientific merit. Even when such papers are retracted, they still receive citations. This is the best refutation argument of a g-index as a measure of scientific excellence.

We are certainly not even close to end the quest for "best scientometric index", and the PIS certainly does not have that ambition. The PIS, with its dual consideration of productivity and impact, addresses some of these shortcomings by providing a more holistic evaluation metric. Future studies might improve the PIS metric by incorporating factors such as peer review activity (Južnič et al., 2010), scientific awards (Meho, 2020), and altmetrics (Nath & Jana, 2021). In conclusion, while the PIS presents a promising approach for assessing researcher performance, it is important to continue refining evaluation metrics to account for the complexities of academic work.

Conclusions

I proposed a novel measure for evaluating researchers' performance called the Performance Impact Scale (PIS). It is composed of performance and impact scores. The PIS accounts for approximately 35% of the independent variance not shared with the WOS H-index, highlighting its unique contribution beyond traditional citation metrics. As research output becomes an increasingly crucial factor in funding and promotion decisions, the development of comprehensive, fair, and transparent ranking systems is more important than ever. Further studies are needed to validate the PIS across different fields and contexts, and to examine how it can be adapted to more accurately reflect individual researchers' contributions in various academic disciplines.

DOI: 10.59519/mper7101

References

Akhtar, M. K. (2024). The H-Index Is an Unreliable Research Metric for Evaluating the Publication Impact of Experimental Scientists. https://doi.org/10.31219/osf.io/wh4ft

Barnes, C. (2017). The h-index Debate: An Introduction for Librarians. *The Journal of Academic Librarianship*, *43*(6), 487-494. https://doi.org/https://doi.org/10.1016/j.acalib.2017.08.013

Bates, D. W. (2017). The ranking of scientists based on citations. *Journal of Biomedical Informatics*, *75*, 128. https://doi.org/https://doi.org/10.1016/j.jbi.2017.10.012

Bornmann, L., & Daniel, H. D. (2005). Does the H-Index for Ranking of Scientists Really Work? *Scientometrics*, *65*(3), 391-392. https://doi.org/10.1007/s11192-005-0281-4

Bornmann, L., Mutz, R., & Daniel, H. D. (2008). Are There Better Indices for Evaluation Purposes Than the <b><i>h</I></B> Index? A Comparison of Nine Different Variants of the <b><i>h</I></B> Index Using Data From Biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830-837. https://doi.org/10.1002/asi.20806

Chua, C., Cao, L., Cousins, K., & Straub, D. W. (2002). Measuring researcher-production in information systems. *Journal of the Association for information Systems*, *3*(1), 145-215. https://doi.org/10.17705/1jais.00026

Costas, R., & Bordons, M. õ. (2007). The H-Index: Advantages, Limitations and Its Relation With Other Bibliometric Indicators at the Micro Level. *Journal of Informetrics*, *1*(3), 193-203. https://doi.org/10.1016/j.joi.2007.02.001

Egghe, L. (2006). Theory and Practise of the G-Index. *Scientometrics*, *69*(1), 131-152. https://doi.org/10.1007/s11192-006-0144-7

Friedmacher, F., Pakarinen, M. P., & Rintala, R. J. (2018). Congenital diaphragmatic hernia: a scientometric analysis of the global research activity and collaborative networks. *Pediatric Surgery International*, *34*(9), 907-917. https://doi.org/10.1007/s00383-018-4304-7

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569-16572. https://doi.org/10.1073/pnas.0507655102

Hirsch, J. E. (2019). Hα: An Index to Quantify an Individual’s Scientific Leadership. *Scientometrics*, *118*(2), 673-686. https://doi.org/10.1007/s11192-018-2994-1

Južnič, P., Pečlin, S., Žaucer, M., Mandelj, T., Pušnik, M., & Demšar, F. (2010). Scientometric indicators: peer-review, bibliometric methods and conflict of interests. *Scientometrics*, *85*(2), 429-441. https://doi.org/10.1007/s11192-010-0230-8

Kelly, C. D., & Jennions, M. D. (2006). The h index and career assessment by numbers. *Trends in Ecology & Evolution*, *21*(4), 167-170. https://doi.org/10.1016/j.tree.2006.01.005

Lane, J. (2010). Let's make science metrics more scientific. *Nature*, *464*(7288), 488-489. https://doi.org/10.1038/464488a

Leydesdorff, L., Bornmann, L., & Opthof, T. (2019). Hα: The Scientist as Chimpanzee or Bonobo. *Scientometrics*, *118*(3), 1163-1166. https://doi.org/10.1007/s11192-019-03004-3

Lüscher, T. F. (2018). Measuring the unmeasurable: assessing the quality of science and scientists. *European Heart Journal*, *39*(20), 1765-1769. https://doi.org/10.1093/eurheartj/ehy295

Meho, L. I. (2020). Highly prestigious international academic awards and their impact on university rankings. *Quantitative Science Studies*, *1*(2), 824-848. https://doi.org/10.1162/qss_a_00045

DOI: 10.59519/mper7101

Memisevic, H., Ajnadzic, N., & Memisevic, M. (2022). Measuring Scholarly Impact in Bosnia and Herzegovina: The Picture Beyond the Numbers. *Journal of Scientometric Research*, *11*(1), 112-117.

Memisevic, H., Pasalic, A., Mujkanovic, E., & Memisevic, M. (2019). In Search of a Silver Bullet: Evaluating Researchers' Performance in Bosnia and Herzegovina. *Journal of Scientometric Research*, *8*(3), 125-130.

Memisevic, H., Taljic, I., & Hadziomerovic, A. M. (2017). Making Use of H-index: the Shape of Science at the University of Sarajevo. *Acta Informatica Medica*, *25*(3), 187-190. https://doi.org/10.5455/aim.2017.25.187-190

Nath, A., & Jana, S. (2021). A Scientometric Review of Global Altmetrics Research. *Science & Technology Libraries*, *40*(3), 325-340. https://doi.org/10.1080/0194262X.2021.1918607

Rong, Z., Chang, R., Lu, L., Wang, X., & Zhang, Z. (2021). Research on the Index for Evaluating Academic Activeness of Researchers. *Proceedings of the Association for Information Science and Technology*, *58*(1), 881-882. https://doi.org/10.1002/pra2.596

Sathian, B., Simkhada, P., & Sreedharan, J. (2014). Importance of the Authorship Order and Number of Co Authors in a Publication for Evaluation: A Necessary Enquiry. *Nepal Journal of Epidemiology*, *4*(4), 384-385. https://doi.org/10.3126/nje.v4i4.11207

Selivanova, I. V., Kosyakov, D. V., & Guskov, A. E. (2019). The Impact of Errors in the Scopus Database on the Research Assessment. *Scientific and Technical Information Processing*, *46*(3), 204-212. https://doi.org/10.3103/S0147688219030109

Sternberg, R. J. (2018). Evaluating Merit Among Scientists. *Journal of Applied Research in Memory and Cognition*, *7*(2), 209-216. https://doi.org/10.1016/j.jarmac.2018.03.003

Zerem, E. (2017). The ranking of scientists based on scientific publications assessment. *Journal of Biomedical Informatics*, *75*, 107-109. https://doi.org/https://doi.org/10.1016/j.jbi.2017.10.007