Scholarly Assessment Reports

Research Performance Assessment Issues: The Case of Kazakhstan

RESEARCH

Levy Library Press

G. ALIBEKOVA
S. ÖZÇELIK
A. SATYBALDIN
M. BAPIYEVA
T. MEDENI

**Author affiliations can be found in the back matter of this article*

ABSTRACT

The study deals with research performance assessment issues as important aspects of research management and research quality. The case of Kazakhstan clearly demonstrates the impact of the prevailing bibliometric-centered approach. The aim of this study is to suggest an inclusive scale of individual research performance assessment. The method used is a quantitative study of the opinions of researchers and academics on the range of research related activities they traditionally carry out. The study expands the knowledge base on academic human resource management, and can be of high relevance for substantiating the criteria of performance assessment of researchers by HR managers of universities and public research institutions. The research results can be helpful for setting and complying with individual and institutional criteria for research performance evaluation.

POLICY HIGHLIGHTS

- A Survey among 264 researchers in Kazakh universities and public research institutions (response rate: 63%) asked them to rate their activities in five groups: supervising activity, professional advancement, publications, public recognition, and scientific & organizational activities.
- The results demonstrate that their priorities correspond to national and international priorities: publishing papers in local and international peer-reviewed journals indexed in WoS and Scopus, and monographs.
- Findings revealed that "Supervising" and "Professional Advancement" activities have the highest importance among all criteria groups.
- Respondents gave the highest preference to participating at overseas and international conferences, seminars and workshops, thus expressing their desire to disseminate their research findings internationally and to build international links.

CORRESPONDING AUTHOR:
G. Alibekova
Institute of Economics of the Ministry of education and science of Kazakhstan, Almaty, Kazakhstan
galibekova77@gmail.com

KEYWORDS:
performance assessment; bibliometrics; peer review; research performance; research evaluation policy; research performance criteria

TO CITE THIS ARTICLE:
Alibekova, G., Özçelik, S., Satybaldin, A., Bapiyeva, M., & Medeni, T. (2021). Research Performance Assessment Issues: The Case of Kazakhstan. *Scholarly Assessment Reports*, 3(1): 5, pp. 1–13. DOI: *https://doi.org/10.29024/sar.37*

- Scientific & organizational activities are the core activities which correlate to all other activities. And the role of S&O criteria is definitely underestimated in the performance assessment of researchers.
- The current research performance evaluation system in Kazakhstan is dominantly based on bibliometrics and is one-sided and biased. An inclusive scale of individual research performance assessment needs to be developed, considering researchers' ideas and preferences.

INTRODUCTION

Performance evaluation plays an important role in management. Performance evaluation is presented as a process of quantification of the performance and effectiveness of actions (Hansen 2017). It plays a key role in building institutional strategies, funding schemes, recruitment and career promotion issues.

Research management is quite a creative process. There are no exact standards that allow an objective and precise assessment of research performance. The assessment of individual researchers remains a deliberately difficult process without a standard solution.

Research performance evaluation has a range of goals. There are numerous studies on the impact of research performance evaluation. Some studies state that employees benefit from performance evaluation which leads to greater motivation (Terlicka 2013, Tuytens & Devos 2012). However, others conclude that without clear goals the performance appraisal can cause employee dissatisfaction and even a reduction in organisational commitment (Maley 2013). Fair performance assessment has a significant impact on organizational excellence and competitiveness in the long run (Arbab & Abaker 2018). When considering different views the link between performance evaluation and employee motivation tends to become blurred (Idowu 2017).

Nowadays, in research performance evaluation the two main approaches used are peer review and bibliometrics. The latter tends to be more popular due to a growing volume of open information published on the internet. The Web of Science and Scopus citation databases are the main tools available for bibliometric assessment. Bibliometric measures derived from these databases became target indicators for the performance assessment of researchers, organizations and countries. Global ranking agencies explore the opportunities of bibliometrics which are used to serve as evaluation criteria. One of the criteria of the Global Competitiveness Index is the number of publications and their citations in the Scopus database, while the Global Innovation Index is based on citation indices derived from the Web of Science. The world university rankings (for instance, QS, Times Higher Education, ARWU) also push the research institutions and universities to publish in sources indexed in Web of Science or Scopus. Bibliometrics has become entrenched in the individual, team and institutional assessment of research performance.

Obviously, inappropriate indicators create perverse incentives (Wilsdon et al. 2015). Excessive bibliometric pressure increases publications in the predatory journals.[1] In the case of Australia, the bibliometric approach of research performance evaluation policy showed a drop in the relative quality of publications along with an increase in publication productivity (Butler 2003). In Turkey it led to uncertainty, and to research ethics and quality issues (Yucel & Demir, 2018). Similarly, Kazakhstan appeared in the list of the most affected countries (Macháček & Srholec, 2021). Such misuse of bibliometrics led to the publication of the Leiden Manifesto which comprises ten principles that guide best practice in metrics-based research assessment (Hicks et al., 2015). Unfortunately, most governments' policies do not comply with the Leiden manifesto principles and bibliometric targets continue to prevail and lead to low-quality publications and research ethics violations.

Moreover, researchers do not only carry out publishing, they engage in a lot of other activities: reviewing, organizing conferences, writing proposals, supervising students research, etc. Despite

1 According to *https://predatoryjournals.com*: "Predatory Journals take advantage of authors by asking them to publish for a fee without providing peer review or editing services".

their high significance, these kinds of activities are not taken into account when measuring research performance. While promoting "inclusive metrics of success and impact to dismantle a discriminatory reward system in science", Davies et al. (2021) distinguish between a narrow and an inclusive view of scientific impact. The narrow view is focused on citation in evaluating scientific impact, while the inclusive view captures all activities: education, collaboration, science communication, community engagement, etc.

The current research adopts the inclusive view: the main issue is to recommend inclusive performance assessment criteria through assessing the perceived significance of each type of activity. The inclusive performance assessment motivates high quality of research, high level of productivity and strong performance in all areas.

Therefore, this research aims to assess the importance of different types of activities and metrics from the point of view of researchers in order to recommend an inclusive set of research performance criteria.

MAIN METHODS OF RESEARCH PERFORMANCE ASSESSMENT

The two main approaches of research performance evaluation are peer review and bibliometrics. The debates over the pros and cons of both approaches have continued for about three decades. Here we discuss the recent research findings on both methods.

Peer review is a qualitative tool. The main drawbacks of peer review that have been observed over time are subjectivity and expensiveness. Similarly, the current peer review system has been criticized for failures in representing the opinions and interests of non-peer clients (Frijters & Torgler 2019). Researchers still find biases in peer review processes (Santos et al., 2021; Haffar et al., 2019). Some studies confirm the contribution of the peer review system to gender inequality (Sato et al., 2021). Tamblyn et al. (2018), for instance, showed that female applicants were found to receive consistently lower reviewer scores than male applicants.

Despite the inconsistency of the h-index and the inappropriateness of the journal impact factor for research performance assessment (Waltman & van Eck, 2012; van Leeuwen, 2012), the bibliometric approach to research assessment has become popular. It is most frequently performed in the Nordic countries, the Netherlands, Italy and the United Kingdom (Jappe, 2020). Bibliometrics are considered more attractive than peer review because they demand much less time and effort. Moreover, some studies confirm that they are consistent with peer review results. Thus, policymakers increasingly use these metrics in addition to, and in some cases instead of, expert assessment of research results (Belter, 2015).

However, the stand-alone use of bibliometrics brings distorted results. The emphasis on publications in high impact factor journals diverts the attention of researchers from the real contribution to knowledge production. They shift their focus to tactical actions to match established reward and promotion systems (Mahmood & Shah 2016). Evaluation using bibliometric indicators can be also complicated and unreliable due to the high probability of misuse. The interpretation of these indicators should be done by field experts. Otherwise, the use of metrics can lead to the demotivation of researchers and a decrease in the quality of research and even education. This is because unlike scientific productivity, the *quality of research* does directly affect the quality of education (Cadez et al., 2017). Also, bibliometric indicators can be manipulated, causing problems in measuring societal impact (Moed, 2017). Some authors claim that "citations and impact factors metrics are flawed and biased against already marginalized groups and fail to accurately capture the breadth of individuals' meaningful scientific impacts" (Davies et al., 2021). Moreover, the "quick-and-dirty" use of metrics can introduce inaccurate results into the assessment of academic research performance (Moed, 2020).

The dominance of the bibliometric approach in research performance evaluation has both short and long-term consequences such as encouraging plagiarism, "salami slicing", demotivation of young researchers to write papers due to slim chances of publication in high-impact-factor journals, etc. (Lawrence, 2003; Cherubini, 2008; Notkins, 2008). Among the long-term consequences Brischoux & Cook (2009) mentioned a reduction of the amount of information available to the scientific community because of the failure to publish in high-impact-factor journals and the suffering of aspects of research such as core creativity if scientists only focus on their publication records.

DOI: 10.29024/sar.37

To apply bibliometric methods it is essential to ensure that international journals are a major means of communication in each particular field (van Raan 2014). If the major scientific communication is through national journals, the bibliometric approach is not applicable. In the bibliometric approach, one should take into account that industry research has a lower impact than public research (Abramo, 2021). Hence, it must be evaluated in a different way than public research. Evaluators also should take into account that correlations between technological and scientific impact indicators are mostly low or moderate (Guerrero-Bote, Moed & De-Moya-Anegón, 2021). This shows that they measure different aspects of research performance. As the use of the h-index alone causes a loss of citation information, Zhang proposes the introduction of the e-index (Zhang 2009). When applying the bibliometric approach the phenomenon of "delayed recognition", also known as the "Sleeping Beauty phenomenon", should be taken into account. "Sleeping Beauties" are publications which may attract a lot of attention and become cited after many years (van Raan, 2004). It has been noted that publication in open access journals positively influence citation impact (Moed, 2007). Therefore, bibliometrics cannot indicate anything if the scientific field is not represented sufficiently well in open information space or if information policy prohibits posting publications in open access directories. For bibliometric analysis in the social sciences and humanities, a broader range of both publications (including non-ISI journals and monographs) and citation indicators (including non-ISI reference citation values) is needed (Nederhof, 2006).

Based on the literature review, it should be noted that the methodology of bibliometric assessment has a number of advantages over the method of expert assessment: it is inexpensive, non-invasive, easy to implement, provides both rapid updates and intertemporal comparisons, is based on objective quantitative and qualitative data, and provides a high degree of representativeness. Citation metrics can make the process more efficient and cost-effective, but expert evaluation should remain a central element in any research evaluation process (Butler, 2007). The challenge is to combine these two methodologies so that they work effectively together (Moed, 2007). Peer review supported by advanced bibliometric methods can make the process more objective and transparent (van Raan, 2014). To avoid the subjectivity of peer review-based evaluation 'disciplined peer review' has been suggested (Besselaar & Sandström, 2020).

One instance where the bibliometric approach can replace peer review is in assessing groups (Ball, 2020). This is possible due to bibliometric methods based on advanced application tools, which have become high quality, reliable and very informative instruments in research performance evaluation. However, evaluating individual research performance is a complex task that would ideally examine productivity, scientific impact, and research quality, a task that metrics alone have been unable to fulfill (Sahel, 2011).

The literature review on approaches to research performance assessment shows that both the bibliometric and expert review approaches have their advantages and disadvantages. Recent studies show that there is a trend to combine these two approaches to achieve synergy in research performance evaluation results.

RESEARCH PERFORMANCE EVALUATION POLICY IN KAZAKHSTAN

The science evaluation policy reforms in Kazakhstan started in 2011 when the country joined the Bologna education system and announced new PhD thesis requirements of publication in WoS or Scopus indexed journals. That led to the boosting of publications in the predatory journals.

Later, Kazakhstan continued, step by step, its policy of introducing bibliometrics into research project evaluation, settling requirements for research principles, evaluating candidates for the conferring of academic titles, etc. As a result of the policy, Kazakhstan has entered the list of the countries most affected by a reliance on bibliometrics (Macháček and Srholec, 2021).

Table 1 presents the number of publications per year made by authors from institutions located in Kazakhstan during 2000–2020. The table shows that the number of publications increased rapidly in 2015 compared to 2010. This increase took place in 2013. While during 2000–2012 Kazakhstan's rank in the Scimago list was in the range 90–100, as from 2013 it increased to 70, and in 2014 to 66, remaining more or less stable up until 2020.

YEAR	NUMBER OF CITABLE DOCUMENTS	RANK
2000	240	86
2005	371	90
2010	472	98
2015	2479	67
2020	5339	68

Table 1 Number of publications of Kazakhstan per year in Scimago Journal Rank.

Source: Scimago, 2021. *https://www.scimagojr.com*.

When tracking changes to the national policy of research performance evaluation for different purposes (conferring PhD degree, academic titles, granting funds for projects, etc) we can confirm that the bibliometric approach has become more sophisticated year on year.

The policy related to conferring PhD degrees is still changing and its history resembles the overall policy trends in research performance evaluation. At the initial stage the policymakers put general publication requirements such as publication in Web of Knowledge journals with a non-zero impact factor or indexed in Scopus (***Table 2***). In the second stage (in 2016), they started to differ between research fields. The third stage of amendments differentiates between ranks of journals (quartiles and percentiles) in which PhD candidates have to publish.

Table 2 demonstrates similar changes in the Expert Council membership requirements. In 2018 a requirement on having publications and citations in Web of Science or Scopus was introduced. Another option is having publications in local journals approved by the national authority. This confirms that the local authorities realize that not all candidates to the Expert Council can have publications in Web of Science or Scopus but can nevertheless be considered experts. There are no such options for conferring PhD degrees.

YEAR	REQUIRED INTERNATIONAL PEER-REVIEWED PUBLICATIONS
2011	At least 1 paper in the journals indexed in ISI Web of Knowledge with a non-zero impact factor or indexed in Scopus.
2016	*Science, engineering, medicine, agriculture:* at least 1 paper in journals indexed in Web of Science with a non-zero impact factor or indexed in Scopus, Pubmed, zbMath, MathScinet, Agris, Georef, Astrophysical journal. *Other fields:* at least 1 paper in journals indexed in Web of Science with a non-zero impact factor or indexed in Scopus, JSTORE. *Expert Council membership requirements:* Specialists with a scientific degree, having at least 5 publications over the last 5 years in peer-reviewed international scientific journals or copyright certificates, patents, certificates of intellectual property, or 10 publications in journals recommended by the authorized body.
2018	*Science, Mathematics, Statistics; ICT; Engineering, manufacturing and construction; Agriculture and bioresources; Veterinary Sciences; Healthcare (medicine), Services:* At least 1 paper in journals with a non-zero impact factor according to Journal Citation Reports (Clarivate Analytics) OR journals of at least the 25th percentile of Scopus CiteScore or in zbMath, MathScinet, Astrophysical journals. *Other fields:* at least 1 paper in journals with a non-zero impact factor according to Journal Citation Reports (Clarivate Analytics) OR indexed in Web of Science Core Collection (Arts and Humanities Citation Index), Science Citation Index Expanded, Social Sciences Citation Index) OR journals of at least 25th percentile of Scopus CiteScore, OR included into JSTOR. *Expert Council membership requirements:* At least 1/3 part of members must have at least h-index –2 in Web of Science or Scopus; AND/OR must have publications in Q1-Q3 of Web of Science or Scopus; AND/OR at least 10 publications in local journals approved by the national authority.

Table 2 History of changes in required international peer-reviewed publications for conferring PhD degree in Kazakhstan.

When the requirements for principal positions in research projects are analyzed it can be seen that they differ not only between research fields but also between types of research. For basic research, the publication requirements are higher than they are for applied research.

DOI: 10.29024/sar.37

In natural, technical, life sciences and medicine, the requirements for project principal positions are stronger than they are for other sciences. Within the last 5 years candidates must have:

Basic research: at least 2 articles and/or reviews in peer-reviewed scientific journals included in Q1-Q3 of the Web of Science and (or) have a Scopus Cite Score percentile of at least 50, of which at least one as the corresponding or the first author.

Applied research: at least 2 articles and/or reviews in peer-reviewed scientific journals included in Q1-Q4 of the Web of Science and (or) have a Scopus Cite Score percentile of at least 35, of which at least one as the corresponding or the first author.

Expectations on output of research grants in natural, engineering, life sciences and medicine also differ between types of research and are higher than they are for other sciences:

Basic research: at least 3 articles and/or reviews in peer-reviewed scientific journals of Web of Science Q1-Q3 and/or have a Scopus CiteScore percentile of at least 50; or at least 2 articles and/or reviews in peer-reviewed scientific journals included in Web of Science Q1-Q2 and (or) with a Scopus CiteScore percentile of at least 65; OR at least 1 article and/or review in a peer-reviewed scientific journal with a percentile in the Journal Citation Reports of more than 90 or a Scopus CiteScore percentile of at least 95;

Applied research: at least 2 articles and/or reviews in peer-reviewed scientific journals, included in Web of Science Q1-Q3 and (or) have a Scopus CiteScore percentile of at least 50; – OR at least 1 article and/or review in a peer-reviewed scientific journal included in Web of Science Q1-Q3 and (or) have a Scopus CiteScore percentile of at least 50, and at least 1 foreign or international patent included in the Derwent Innovation database.

In social, humanities and military sciences the expected output includes at least 1 article and/or review in a peer-reviewed scientific journal of Q1-Q3 Web of Science, and (or) included in the Social Science Citation Index or Arts and Humanities Citation Index, and (or) having a Scopus CiteScore percentile of at least 35. The lower requirements for social and humanities sciences are due to their bias in favor of English language journals. (Moed, 2007).

Thus, the research performance evaluation system in Kazakhstan tends to strengthen the bibliometric-centered approach. The recent improvements have been related to differentiations in field and types of research.

METHODOLOGY

Researchers have indicated two approaches to studying and assessing research performance: evaluative and explanatory (Avital & Collopy 2008). Explanatory performance assessment has an objective of predicting performance, it is proactive and its rationale is to improve productivity.

Our study can be considered as an explanatory performance study as it has the objective to build the knowledge society by constructing an inclusive scale for performance assessment in the academic and research sector. This approach is predictive due to the embedded feature of motivating researchers to build capacity through supervising students and growing professionally.

Our research is conducted in two stages:

> First, an initial list of all types of research-related activities performed by researchers has been formed. The activities are classified into five groups: supervising, professional advancement, publications, public recognition, scientific & organizational activities.
>
> Second, these types of research activities are assessed through the Likert scale used to measure respondents' attitudes to a particular activity according to the following values: 1-Unimportant, 2-Not very important, 3-Moderately important, 4-Important, 5-Very important.

There are 264 people included in the sample. The response rate is 63%. The statistical analyses are performed through the IBM SPSS Statistics 22 package program. Frequency distributions (number, percentage) are given for numerical variables while descriptive statistics (mean, standard deviation, median, minimum, maximum) are given for categorical variables (eg, gender). Normality assumptions of numerical variables are examined with the coefficients of skewness and kurtosis and these coefficients are found to be within a ± 2 range. Therefore, parametric statistical methods were used in the study. The differences between two independent groups (e.g. gender) are examined by an Independent Sample t Test. The differences between more than two independent groups (e.g. age groups) were analyzed by the One Way Analysis of Variance (ANOVA). Tukey's multiple comparison test was used to determine from which group the difference arises as a result of a one-way ANOVA. The statistical significance level was set to $p < 0.05$ and the analysis was completed at a 95% confidence level.

DOI: 10.29024/sar.37

The analysis has shown the following make-up of respondents (***Figure 1***)

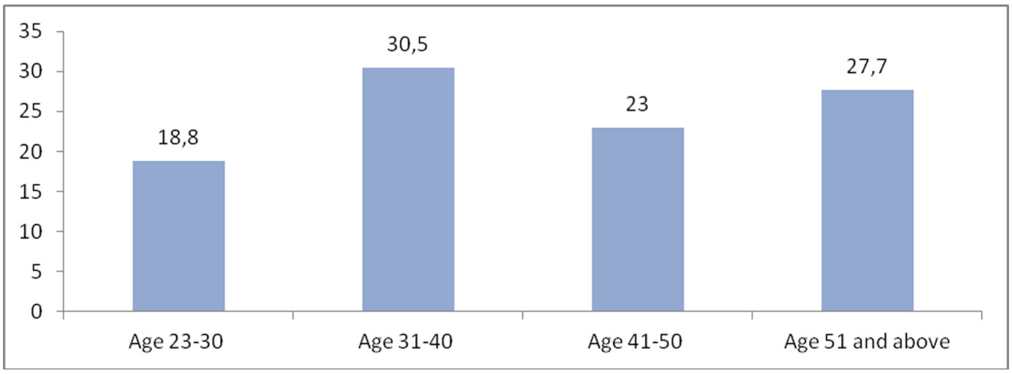

Figure 1 Respondents grouped by the criteria of age, %.

In addition, the average age of the respondents is 43.4 years, the youngest is 23 years old, and the oldest is 84 years old. 63.7% of them are female. In addition, 55.7% are from the Al-Farabi KazNU faculty and 44.3% are from various public research institutes (Institute of Economics, Institute of maths, Institute of mechanics and machine science, Institute of oriental sciences, Institute of ICT, Institute of language sciences, Institute of philosophy, political sciences and religion studies).

RESULTS

The activities surveyed are classified into five groups: Publications, Public Recognition, Scientific & Organizational activities, Professional Advancement, Supervising. The content of each group is seen in Appendix 1.

The overall results of the survey are given in ***Figure 2***.

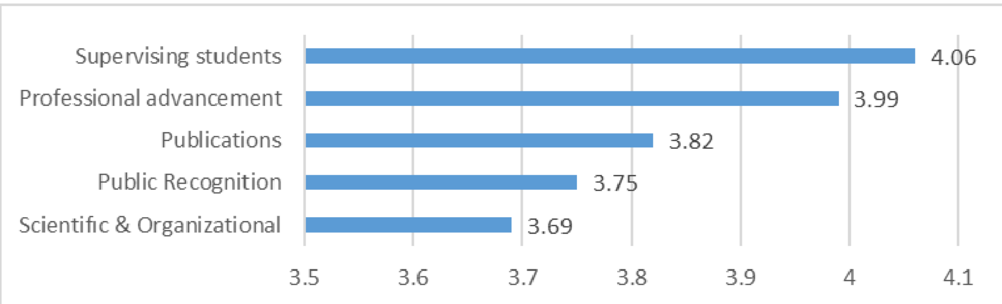

Figure 2 Average scores of activity groups by respondents (Likert Scale, 1 = unimportant, 5 = very important).[2]

According to the survey, Supervising (4.06) is perceived as the most "important" by the respondents. One can question this outcome since 55% respondents come from the university and this cannot be true for researchers of public research institutions (PRI). According to ***Table 3***, there is no statistically significant difference between the institutions in terms of activity scores as a result of the Independent Sample t Test ($p > 0.05$).

2 The reliability of these five activity scores is examined with the Cronbach Alpha coefficient and all of them were found to have high reliability levels ($\alpha > 0.700$). The skewness and kurtosis coefficients of the activity scores were in the range of ± 2. Therefore, the normal distribution assumption is satisfied.

ACTIVITY	ORGANIZATION	COUNT	MEAN	STANDARD DEVIATION	T	P
Scientific & Organizational	University	147	3,70	0,79	0,441	0,660
	PRI	117	3,66	0,68		
Professional Advancement	University	147	3,98	0,76	–0,142	0,887
	PRI	117	4,00	0,76		
Public Recognition	University	146	3,83	0,92	1,574	0,117
	PRI	115	3,65	0,96		
Supervising	University	143	4,16	1,02	1,806	0,072
	PRI	112	3,93	1,01		
Publications	University	145	3,85	0,81	0,736	0,462
	PRI	117	3,78	0,75		

Table 3 Analysis of Differences between Activity Scores, by Organization.

t: Independent sample t Test.

Professional Advancement (3.99) can be interpreted as "important" as well for both university faculty staff and PRI researchers. The group of Scientific & Organizational activities has the lowest value as seen in ***Figure 2***.

One can doubt if Public Recognition can be referred to as an activity since this group is a result of an activity or an achievement rather than an activity itself. But we are studying attitudes of researchers to different criteria. Therefore awards (e.g. public recognition indicators) can be considered as criteria as well. ***Figure 2*** shows it is perceived as less important.

Table 4 shows that there are no statistically significant differences between age groups in terms of activity scores except Public Recognition acitivity. More specifically, one way the ANOVA test reveals statistically significant differences among the age groups is in terms of Public Recognition scores ($p < 0.05$). For example, the Public Recognition scores for the 51 and over age group are significantly lower than those for the 23–30 and 31–40 age groups.

ACTIVITY	AGE	COUNT	MEAN	STANDARD DEVIATION	F	P
Scientific & Organizational	1. Age 23–30	48	3,55	0,68	0,882	0,451
	2. Age 31–40	78	3,74	0,69		
	3. Age 41–50	59	3,74	0,82		
	4. Age 51 and over	71	3,72	0,69		
Professional Advancement	1. Age 23–30	48	4,12	0,55	2,500	0,060
	2. Age 31–40	78	4,15	0,72		
	3. Age 41–50	59	3,92	0,87		
	4. Age 51 and over	71	3,87	0,68		
Public Recognition	1. Age 23–30	47	4,01	0,78	4,424	**0,005* Difference: 1–4 2–4**
	2. Age 31–40	76	3,86	0,93		
	3. Age 41–50	59	3,80	0,88		
	4. Age 51 and over	71	3,45	0,97		
Supervising	1. Age 23–30	45	3,99	0,96	0,559	0,642
	2. Age 31–40	76	4,18	0,84		
	3. Age 41–50	56	4,10	1,02		
	4. Age 51 and over	70	4,00	1,14		
Publications	1. Age 23–30	47	3,77	0,74	0,526	0,665
	2. Age 31–40	77	3,92	0,79		
	3. Age 41–50	59	3,77	0,84		
	4. Age 51 and over	71	3,83	0,70		

Table 4 Analysis of Differences between Activity Scores, by Age.

*: $p < 0,05$ (statistically significant), F: One way Variance Analysis (ANOVA), "Tukey" test for Differences between groups.

To check the relationships between the activities we applied correlation analysis, which demonstrated a moderate positive linear correlation between the criteria scores (***Table 5***).

		1	2	3	4	5
1. Scientific & Organizational	r	1	**,558****	**,640****	**,600****	**,515****
	p		**0,000**	**0,000**	**0,000**	**0,000**
	n	264	**264**	**261**	**255**	**262**
2.Professional Advancement	r		1	**,592****	**,482****	**,635****
	p			**0,000**	**0,000**	**0,000**
	n		264	**261**	**255**	**262**
3.Public Recognition	r			1	**,554****	**,566****
	p				**0,000**	**0,000**
	n			261	**254**	**260**
4.Supervising	r				1	**,518****
	p					**0,000**
	n				255	**255**
5.Publications	r					1
	p					
	n					262

Table 5 Analysis of relationships between activities.

r: Pearson Correlation Coefficient *: p < 0,05 **: p < 0,01.

More specifically, there is a moderate positive correlation between the scores of Scientific & Organizational (S&O) activity and other groups.

Also, a statistically significant positive correlation exists between: 1) Public Recognition group scores and Supervising and Publications group scores; 2) Supervising and Publications criteria scores (***Table 5***).

Table 5 demonstrates that S&O activities correlate to all other activity groups. Moreover, Professional Advancement helps to enhance S&O activities, and vice versa. Hence, the results indicate that S&O activities are the core activities which correlate with all other activities.

One more logical finding is the correlation between Publications and Supervising activities. Supervising PhD students helps their research supervisors to publish more papers. And vice versa – if the researchers have a high publication activity, this indicates their capacity, and hence they are assigned as a research supervisor to more students. By way of comparison, the Higher Education Council of Turkey's criterion for higher education institutions is "the number of students in master's and doctoral programs, the number of doctoral specialities, and the number of doctoral graduates".

The correlation between Publication and Public Recognition criteria seems plausible and clear as well. According to the national rules for awarding academic degrees and ranks, one should publish a certain number of papers to be nominated for the academic rank of professor or associate professor. To apply for a PhD thesis defence, doctorates must have at least one publication in journals of a certain quartile/percentile in the Web of Science Core Collection or the Journal Citation Reports or Scopus. In the current year, 2021, a new amendment has been added to the national rules for doctorates, according to which there is an option to skip writing a PhD thesis if a PhD student has two original research papers in the 1–2 quartile journals indexed in the Journal Citation Reports.

The One Way ANOVA test reveals no statistically significant differences between age groups in the various types of activity, except in Public Recognition activity ($p < 0.05$). For example, Public Recognition scores for the 51 and higher age group are significantly lower than those for the 23–30 and 31–40 age groups.

An independent Sample t Test (p < 0.05) reveals no statistically significant differences between females and males in terms of activity scores except in Public Recognition activities. Public Recognition scores of females are significantly higher than those of males (***Table 6***).

ACTIVITY	SEX	COUNT	MEAN	STANDARD DEVIATION	T	P
Scientific Organizational	Female	158	3,67	0,75	0,326	0,745
	Male	90	3,64	0,76		
Raising Qualification	Female	158	4,04	0,78	1,213	0,226
	Male	90	3,91	0,73		
Public Recognition	Female	157	3,89	0,85	3,040	**0,003***
	Male	88	3,49	1,05		
Supervising	Female	152	4,09	1,07	0,529	0,597
	Male	87	4,01	1,00		
Publications	Female	157	3,87	0,80	1,158	0,248
	Male	89	3,75	0,74		

Table 6 Analysis of Differences between Activity Scores, by Sex.

*: p < 0,05 ((statistically significant) t: Independent sample t Test.

The activities within each group are also assessed by respondents. ***Figure 3*** demonstrates the activities which scored as "important".

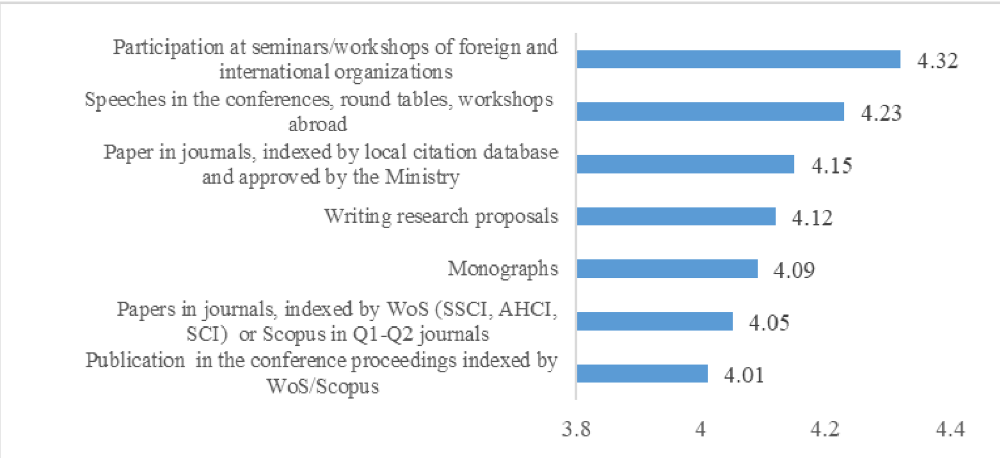

Figure 3 Types of activities scored as "important" within the different groups.

The results of the survey demonstrate that the priorities of local researchers correspond to national and international priorities (publishing papers in local and international peer-reviewed journals indexed in WoS and Scopus, monographs). But respondents gave the highest preference to participation at overseas and international seminars and workshops, speeches at overseas conferences, seminars, and round tables. It indicates a desire of researchers to disseminate their research findings internationally and to build international links. By way of comparison, the Higher Education Council of Turkey encourages the internationalization of research by applying the criteria of "joint projects", which is estimated through the number and budget of projects with international cooperation (both ongoing and completed).

The respondents are also aware of the importance of publishing articles in journals indexed by the Web of Science or Scopus. According to the analysis the average Likert score of the activity "publishing in first and second quartile (Q1-Q2) journals" is slightly higher (4.05) than for "publishing in Q3-Q4 journals" (3.94).

CONCLUSIONS

The theoretical and practical analysis of various concepts and tools for assessing research performance shows the growing importance of bibliometrics. Bibliometric criteria remain popular in quantifying the impact of a scientific article (Aragón, 2013).

The performance evaluation system which is dominantly based on bibliometrics is one-sided and biased. Such a system harms science itself as it inclines a scientist to make more publications and citations instead of doing high-quality research. Quality of research should be the primary concern when practising science and the publication of its results whilst important is a secondary concern.

Thus, peer review remains the main method of evaluating the researcher although it may be biased due to subjective factors such as conflicts of interest, lack of research competence, and superficial expertise.

However, one should keep in mind that the production and dissemination of scientific knowledge should primarily provide social and economic benefits through innovations that are not directly reflected in bibliometric scores. It is also essential to take into account that "Sleeping Beauties" can suddenly be awoken many years after publication (van Raan, 2004).

Our findings revealed that "Supervising" and "Professional Advancement" activities have a higher importance among all criteria groups. However, within the "Publication" activities, publishing papers in privileged local and international peer-reviewed journals and conference proceedings indexed in WoS and Scopus, and monographs are of a high level of importance as well.

"Publication", "Public Recognition" and "Scientific & Organizational" activities are assessed as moderately important. Unsurprisingly, the "Public Recognition" score is for older groups significantly lower it is for younger groups.

Surprisingly for the Asian continent the Public Recognition score for females is much higher than for males. One possible reason for this could be that there is an underestimation of females in conferring ranks, titles and other awards. This hypothesis can be used for future research.

The interesting and specific point is that S&O activities correlate to all the other activity groups. There is a moderate positive correlation between the scores of Scientific & Organizational (S&O) activity and other groups. We can suppose that S&O activities are the core activities which correlate to all other activities. And the role of S&O criteria is definitely underestimated in the performance assessment of researchers. This confirms the earlier conclusion about the necessity to take S&O criteria into account when evaluating researchers' performance. It is supposed that they lead to a more broad and intensive dissemination of research findings. This issue could be a subject of future research as well.

Analysis of sub-activities within each group showed the importance of building scientific communications internationally. High importance of "Participation at foreign and international seminars and workshops" and "Speeches at foreign conferences, seminars, round tables" indicates the welcomed willing of researchers to disseminate their research findings internationally and to build international links.

The study expands the knowledge base on the academic human resource management aspects, and can be of high relevance for substantiating the performance assessment of researchers criteria by HR managers of universities and public research institutions. The research results can be helpful for complying and coinciding individual and institutional criteria for research performance evaluation. This study suggests an inclusive scale to measure the performance of researchers. The inclusive scale for research performance assessment can motivate researchers to achieve organizational goals in a less resource-consuming way. Balanced intentions of university policies and researchers can allow the enhancement of research stimulus and increase research quality.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendix 1.** Questionnaire. DOI: *https://doi.org/10.29024/sar.37.s1*

COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR AFFILIATIONS

G. Alibekova orcid.org/0000-0003-3498-7926
Institute of Economics of the Ministry of education and science of Kazakhstan, Almaty, Kazakhstan

S. Özçelik orcid.org/0000-0001-9946-4785
Ankara Yildirim Beyazit University, Ankara, Turkey

A. Satybaldin orcid.org/0000-0001-7421-4472
Institute of Economics of the Ministry of education and science of Kazakhstan, Almaty, Kazakhstan

M. Bapiyeva orcid.org/0000-0001-7680-6674
National Center for Science and Technology Expertise, Almaty, Kazakhstan

T. Medeni orcid.org/0000-0002-2964-3320
Ankara Yildirim Beyazit University, Ankara, Turkey

REFERENCES

Abramo, G., D'Angelo, C. A., & **Di Costa, F.** (2021). The scholarly impact of private sector research: A multivariate analysis. *Journal of Informetrics, 15*(3). DOI: *https://doi.org/10.1016/j.joi.2021.101191*

Aragón, A. M. (2013). A measure for the impact of research. *Scientific Reports*, *3*, 1649. DOI: *https://doi.org/10.1038/srep01649*

Arbab, A. M. H., & **Abaker, M. O. S. M.** (2018). Human resources management practices and organizational excellence in public orgnizations. *Polish Journal of Management Studies*, *18*, 9–21. DOI: *https://doi.org/10.17512/pjms.2018.18.2.01*

Avital, M., & **Collopy, F.** (2008). "Assessing Research Performance: Implications for Selection and Motivation". *All Sprouts Content*, *13*. *https://aisel.aisnet.org/sprouts_all/13*

Ball, R. (2020). Handbook Bibliometrics. Berlin, Boston: De Gruyter Saur, 2020. DOI: *https://doi.org/10.1515/9783110646610*

Belter, C. W. (2015). Bibliometric indicators: opportunities and limits. *Journal of the Medical Library Association: JMLA*, *103*(4), 219–221. DOI: *https://doi.org/10.3163/1536-5050.103.4.014*

Brischoux, F., & **Cook, T. R.** (2009). Juniors seek an end to the impact factor race. *Bioscience*, *59*(8), 638–639. DOI: *https://doi.org/10.1525/bio.2009.59.8.2*

Butler, L. (2003). Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Research policy*, *32*(1), 143–155. DOI: *https://doi.org/10.1016/S0048-7333(02)00007-0*

Cadez, S., Dimovski, V., & **Groff, M.** (2017). Research, teaching and performance evaluation in academia: the salience of quality. *Studies in Higher Education*, *42*(8), 1455–1473. DOI: *https://doi.org/10.1080/03075079.2015.1104659*

Cherubini, P. (2008). Impact factor fever. *Science, 322*, 191. DOI: *https://doi.org/10.1126/science.322.5899.191b*

Davies, S., Putnam, H., Ainsworth, T., Baum, J., Bove, C., Crosby, S., Côté, I., Duplouy, A., Fulweiler, W., Griffin, A., Hanley, T., Hill, T., Humanes, A., Mangubhai, S., Metaxas, A., Parker, L., Rivera, H., Silbiger, N., Smith, N., & **Bates, A.** (2021). Promoting inclusive metrics of success and impact to dismantle a discriminatory reward system in science. *PLOS Biology*, *19*, e3001282. DOI: *https://doi.org/10.1371/journal.pbio.3001282*

Frijters, P., & **Torgler, B.** (2019). Improving the peer review process: a proposed market system. *Scientometrics, 119*, 1285–1288. DOI: *https://doi.org/10.1007/s11192-019-03076-1*

Guerrero-Bote, V. P., Moed, N., & **De-Moya-Anegón, F.** (2021). A further step forward in measuring journals' technological factor. *Profesional De La Información*, *30*(4). DOI: *https://doi.org/10.3145/epi.2021.jul.06*

Haffar, S., Bazerbachi, F., & **Murad, M. H.** (2019). Peer review bias: A critical review. *Mayo Clinic Proceedings*, *94*(4), 670–676. DOI: *https://doi.org/10.1016/j.mayocp.2018.09.004*

Hansen, M. B. (2017). "Performance Management and Evaluation." In B. Greve (ed.), *Handbook of Social Policy Evaluation*. London: Edward Elgar (pp. 223–241). Available at *http://vbn.aau.dk/files/237038407/160513_Performance_Management_and_Evaluation_final_allchangeacc.pdf* (last accessed 20 May 2017). DOI: *https://doi.org/10.4337/9781785363245.00019*

Hicks, D., Wouters, P., Waltman, L., et al. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*, 429–431. DOI: *https://doi.org/10.1038/520429a*

Idowu, A. (2017). Effectiveness of Performance Appraisal System and its Effect on Employee Motivation. *Nile Journal of Business and Economics, 3*. DOI: *https://doi.org/10.20321/nilejbe.v3i5.88*

Jappe, A. (2020). Professional standards in bibliometric research evaluation? A meta-evaluation of European assessment practice 2005–2019. *PLoS ONE, 15*(4), e0231735. DOI: *https://doi.org/10.1371/journal.pone.0231735*

Lawrence, P. (2003). The politics of publication. *Nature, 422*, 259–261. DOI: *https://doi.org/10.1038/422259a*

Macháček, V., & **Srholec, M.** (2021). Predatory publishing in Scopus: evidence on cross-country differences. *Scientometrics, 126*, 1897–1921. *https://doi.org/10.1007/s11192-020-03852-4*. Retraction Note at Scientometrics (2021). DOI: *https://doi.org/10.1007/s11192-021-04149-w*

Mahmood, K., & **Shah, S.** (2016). Validation of Journal Impact Metrics of Web of Science and Scopus. *Pakistan Journal of Information Management & Libraries*, *18*, 58–74. DOI: *https://doi.org/10.47657/2016181970*

Maley, J. (2013). Hybrid purposes of performance appraisal in a crisis. *Journal of Management Development*, *30*(10), 1093–1112. DOI: *https://doi.org/10.1108/JMD-03-2012-0036*

Moed, H. F. (2007). The effect of "Open access" on citation impact: an analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology, 58*, 2047–2054. DOI: *https://doi.org/10.1002/asi.20663*

Moed, H. (2017). Applied Evaluative Informetrics. DOI: *https://doi.org/10.1007/978-3-319-60522-7*

Moed, H. F. (2020). Appropriate Use of Metrics in Research Assessment of Autonomous Academic Institutions. *Scholarly Assessment Reports*, *2*(1), 1. DOI: *https://doi.org/10.29024/sar.8*

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: a review. *Scientometrics, 66*, 81–100. DOI: *https://doi.org/10.1007/s11192-006-0007-2*

Notkins, A. L. 2008. Neutralizing the impact factor culture. *Science, 322*, 191. DOI: *https://doi.org/10.1126/science.322.5899.191a*

Sahel, J.-A. (2011). Quality Versus Quantity: Assessing Individual Research Performance. *Sci. Transl. Med. 3*, 84cm13. DOI: *https://doi.org/10.1126/scitranslmed.3002249*

Santos, A., Morris, D. S., Rattan, R., & **Zakrison, T.** (2021). Double-blinded manuscript review: Avoiding peer review bias. *Journal of Trauma and Acute Care Surgery*, *91*(1), e39–e42. DOI: *https://doi.org/10.1097/TA.0000000000003260*

Sato, S., Gygax, P. M., Randall, J., & **Schmid Mast, M.** (2021). The leaky pipeline in research grant peer review and funding decisions: Challenges and future directions. *Higher Education*, *82*(1), 145–162. DOI: *https://doi.org/10.1007/s10734-020-00626-y*

Tamblyn, R., Girard, N., Qian, C., & **Hanley, J.** (2018). Assessment of potential bias in research grant peer review in Canada. *Canadian Medical Association Journal*, *190*, E489–E499. DOI: *https://doi.org/10.1503/cmaj.170901*

Terlicka, K. (2013). Managing human resources in a public institution – case study. *Polish Journal of Management Studies*, *8*, 250–257.

Tuytens, M., & **Devos, G.** (2012). Importance of system and leadership in performance appraisal. *Personnel Review*, *41*(6), 756–776. DOI: *https://doi.org/10.1108/00483481211263692*

Van den Besselaar, P., & **Sandström, U.** (2020). Bibliometrically Disciplined Peer Review: on Using Indicators in Research Evaluation. *Scholarly Assessment Reports*, *2*. DOI: *https://doi.org/10.29024/sar.16*

van Leeuwen, T. N. (2012). Discussing some basic critique on journal impact factors: revision of earlier commentsen. *Scientometrics, 92*, 443–455. DOI: *https://doi.org/10.1007/s11192-012-0677-x*

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics, 59*, 461–466. DOI: *https://doi.org/10.1023/B:SCIE.0000018543.82441.f1*

van Raan, A. F. J. (2014). Advances in bibliometric analysis: Research performance assessment and science mapping. In book: *Bibliometrics: Use and Abuse in the Review of Research Performance Edition: Wenner-Gren International Series* Vol. 87.

Waltman, L., & **van Eck, N. J.** (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology, 63*, 406–415. Open access pre-print version: *http://arxiv.org/abs/1108.390120*. DOI: *https://doi.org/10.1002/asi.21678*

Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. DOI: *https://doi.org/10.4135/9781473978782*

Yucel, A. G., & **Demir, S. B.** (2018). Academic Incentive Allowance: Scientific Productivity, Threats, Expectations. *International Online Journal of Educational Sciences*, *10*, 39–55. DOI: *https://doi.org/10.15345/iojes.2018.01.003*

Zhang, C.-T. (2009). The *e*-Index, Complementing the *h*-Index for Excess Citations. *PLoS ONE, 4*(5), e5429. DOI: *https://doi.org/10.1371/journal.pone.0005429*

TO CITE THIS ARTICLE:
Alibekova, G., Özçelik, S., Satybaldin, A., Bapiyeva, M., & Medeni, T. (2021). Research Performance Assessment Issues: The Case of Kazakhstan. *Scholarly Assessment Reports*, 3(1): 5, pp. 1–13. DOI: *https://doi.org/10.29024/sar.37*

Submitted: 18 April 2021
Accepted: 06 October 2021
Published: 19 October 2021

Scholarly Assessment Reports is a peer-reviewed open access journal published by Levy Library Press.