https://doi.org/10.51408/issi2025_037

Measuring the Continuous Research Impact of a Researcher: The K_z Index

Kiran Sharma[1], Ziya Uddin[2]

[1]*kiran.sharma@bmu.edu.in*, [2]*ziya.uddin@bmu.edu.in*
School of Engineering and Technology, BML Munjal University, Gurugram, Haryana-122413 (India)

Abstract

The ongoing discussion regarding the utilization of individual research performance for academic hiring, funding allocation, and resource distribution has prompted the need for improved metrics. While traditional measures such as total publications, citations count, and the h-index, etc. provide a general overview of research impact, they fall short of capturing the continuous contribution of researchers over time. To address this limitation, we propose the implementation of the K_z index, which takes into account both publication impact and age. In this study, we calculated K_z scores for 376 research profiles. K_z reveals that the researchers with the same h-index can exhibit different K_z scores, and vice versa. Furthermore, we observed instances where researchers with lower citation counts obtained higher K_z scores, and vice versa. Interestingly, the K_z metric follows a log-normal distribution. To determine if the distribution of K_z is independent of subject discipline, we plotted the distribution for three different disciplines. Our analysis concluded that the distribution of $K_z$is indeed independent of the discipline. It highlights its potential as a valuable tool for ranking researchers and facilitating informed decision-making processes. By measuring the continuous research impact, we enable fair evaluations, enhance selection processes, and provide focused career advancement support and funding opportunities.

Introduction

Research impact (Penfield, Baker, Scoble and Wykes, 2014) is a crucial factor when evaluating the contributions of researchers (Reed, Ferre, Martin-Ortega, Blanche, Lawford-Rolfe, Dallimer and Holden, 2021). It plays a vital role in assessing the quality, significance, and reach of their work, which is instrumental in academic promotions, grant allocations, award selections, and overall career progression. Existing indices like the h-index and citation count are commonly used to measure research impact (Bornmann and Daniel, 2005, 2009; Egghe, 2010); however, it's important to recognize that citations may not provide a comprehensive representation of impact, especially in fields where citation practices differ or in emerging research domains with limited citation opportunities. Therefore, a more nuanced approach is necessary to capture the full extent of the research impact (de Saint-Georges and de la Potterie, 2013), considering multiple dimensions beyond traditional metrics.
Initially introduced in 2005 by Hirsch, the h-index is calculated based on the number of papers that have received at least h citations from other papers (Hirsch, 2005). The h-index has been subject to criticism due to its limitations in providing a comprehensive view of scientific impact (Costas and Bordons, 2007; Ding, Liu and Kandonga, 2020). It failed to capture the impact of highly cited papers (Bi, 2023).

Also, it does not take into account the number of authors in each publication (Schubert and Schubert, 2019).
However, since its introduction, the h-index has gained significant popularity in academia and has been commonly employed to evaluate the academic success of researchers in various areas, including hiring decisions, promotions, and grant acceptances. Despite efforts by researchers to propose alternative variants of the h-index (Egghe, 2006; Jin, Liang, Rousseau and Egghe, 2007; Zhang, 2009; Alonso, Cabrerizo, Herrera-Viedma and Herrera, 2010; Khurana and Sharma, 2022), the traditional h-index remains widely used as a performance metric in the assessment of scientists because of its simplicity. After the inception of h-index, many variants of h-index have been proposed to overcome its limitations (Alonso, Cabrerizo, Herrera-Viedma and Herrera, 2009; Batista, Campiteli and Kinouchi, 2006; Hirsch, 2019; Schreiber, 2008a, b; Todeschini and Baccini, 2016).
To overcome the limitations of h-index, Egghe in 2006, proposed g-index (Egghe, 2006) which is determined by the distribution of citations across their publications. It is determined by sorting the articles in decreasing order based on the number of citations they have received. The g-index is defined as the largest number g for which the top g articles collectively accumulate at least g^2 citations. This means that a researcher with a g-index of 10 has published at least 10 articles that collectively have received at least ($10^2 = 100$) citations. It's important to note that unlike the h-index, the citations contributing to the g-index can be generated by only a small number of articles. For example, a researcher with 10 papers, where 5 papers have no citations and the remaining five have 350, 35, 10, 2, and 2 citations respectively, would have a g-index of 10 but an h-index of 3 (as only three papers have at least three citations each).
Further, after recognizing the limitations of the h-index (Ding et al., 2020; Egghe, 2011), researchers have proposed various complementary measures to provide a more comprehensive assessment of research impact such as AR-index (Jin et al., 2007), e-index (Zhang, 2009), p-index (Prathap, 2010), h' -index (Zhang, 2013), h_c-index (Khurana and Sharma, 2022), etc. Van Leeuwen (2008) compared the h-index with various bibliometric indicators and other characteristics of researchers. Similarly, Rons and Amez (2009) proposed a new indicator named, impact validity indicator, in search of excellent scientists.
Further, the e-index proposed by Zhang in 2009 (Zhang, 2013), measure the excess citations received by an author's publications beyond the h-core. The e-index places a strong emphasis on highly cited papers, as it focuses on excess citations beyond the h-core. Jin et al. in 2007 proposed the AR-index (Jin et al., 2007) which is used to measure the citation intensity of the h-core (publications with at least h citations) while considering the age of publications. The limitation of using AR-index is that it focuses on the h-core without considering variations within it. Different publications within the h-core may have different citation counts, but the AR-index does not account for these variations. In 2010, a p-index introduced by Prathap

(Prathap, 2010), measure the productivity and impact by considering an author's h-index, total publications, and the number of citations received. The limitation of the index is that it does not consider the distribution of citations across an author's papers. It treats all papers equally and does not differentiate between highly cited and minimally cited papers.

In the study by Khurana et al. (Khurana and Sharma, 2022), an enhancement to the h-index is proposed to capture the impact of the highly cited paper. They introduced h_c which is based on the weight assigned to the highly cited paper. h_c has a greater impact on researchers with lower h-index values, particularly by highlighting the significance of their highly cited paper. However, the effect of h_c on established researchers with higher h-index values was found to be negligible. It is worth noting that the h_c focuses on the first highly cited paper and does not consider the impact of subsequent highly cited papers. This limitation again highlights the need for a more comprehensive measure that takes into account all the important factors contributing to research impact.

Another measure named, L-sequence, introduced by Liu et al. (Liu and Yang, 2014), computes the h-index sequence for cumulative publications while taking into account the yearly citation performance. In this approach, the L number is calculated based on the h-index concept for a specific year. Consequently, the impact of the most highly cited paper in that year may be overlooked, and papers with less than L citations are also not considered. Although the concept captures the yearly citation performance of all papers, it does not effectively capture the continuous impact of each individual paper. Also gathering data for the L-sequence can be challenging, as it requires delving into the citation history of each paper for every year.

Quantifying research impact is a multifaceted endeavour (Batista et al., 2006). There is no universally accepted metric till now to measure the continuous research impact of a researcher. Different stakeholders may prioritize different indicators, such as the number of publications, total citations, patents, etc. Measuring the continuous research impact of a researcher is crucial for granular assessment, differentiation among researchers, funding decisions, identification of emerging talent, etc. Determining an inclusive and comprehensive approach that captures the diverse dimensions of research impact remains a challenge.

Research objective

The primary objective of this study is to introduce a robust and reliable metric that can effectively capture the continuous research impact of a researcher. The aim of the proposed metric is to differentiate between two researchers who possess identical research parameters, for example; the number of publications or total citations or h-index, etc. In order to accomplish the stated objective, a newly introduced measure called the K_z-index has been proposed.

K_z-index

Based on the limitations of the h-index, especially h ignores the highly cited papers, the index h_c was proposed (Khurana et al., 2022). In index h_c a weight of the highest cited paper of an individual was computed. Following this study, the proposed K_z index serves as a tool to measure the continuous research impact of a researcher. It aims to capture the continuous and evolving contributions made by the researcher over time, considering factors such as total publications, citation count, h-index, and publication age.

Definition of K_z-index

K_z takes into account two important factors of research: paper impact and paper age.

1. **Impact** (k): The impact of a paper is calculated by considering two factors: the number of citations (*C*) it has received and its author's h-index.

The impact of the paper is calculated by using the following equation;

$$C \leq (h+1)^k \quad \ldots(1)$$

where $k \in R^+$ (positive real number).

2. **Age**(Δt): Δt represents the publication age in relation to the current year and can be calculated as

$$\Delta t = C_y - P_y \quad \ldots(2)$$

where C_y represents the current year and P_y represents the publication year.
Now, from equations (1) the value of "k" can be calculated and using equation (2), K_z can be calculated for every researcher as

$$K_z = \sum_{i=1}^{N} k_i' \quad \ldots(3)$$

where $k' = \frac{k}{\Delta t}$, and *N* is number of publications.
Equation (3) highlights the significance of K_z metric by incorporating essential research indicators, including the number of publications, total citations, year of publication, publication age, and h-index. This comprehensive approach ensures that all significant indicators of a researcher's work are considered, resulting in a more robust and holistic assessment of their research impact.

Advantages of K_z

Measuring the continuous research impact of a researcher is crucial for several reasons:

1. *Granular assessment:* Traditional matrices such as the number of publications, total citations, h-index, etc. present an overall research impact and do not have the capability to capture the ongoing progress and advancement of their work, whereas K_z can acquire a more nuanced and thorough comprehension of a researcher's contributions as they evolve over time.
2. *Differentiation among researchers:* Even if two researchers having the same h-index, the patterns of their research impact over time may vary significantly. Analysing their continuous research impact can uncover disparities in

productivity and can provide a more comprehensive understanding of their individual profiles. Hence, K_z allows for a more nuanced differentiation among researchers.

3. *Evaluation of long-term impact:* Researchers may experience fluctuations in their productivity and impact over their careers. Measuring continuous research impact enables the evaluation of long-term contributions. K_z has the capability of highlighting researchers who consistently generate influential work and have a lasting impact on their field.
4. *Career progression and funding decisions:* Many academic institutions, funding agencies, and hiring committees rely on research performance metrics to make decisions. K_z can provide more informed evaluations of researchers, enabling fairer assessments and enhancing the recognition of sustained excellence.
5. *Identification of emerging talent:* Continuous research impact measurement can help identify early-career researchers with promising trajectories. By recognizing their continuous growth and impact, further opportunities can be provided to nurture their potential.

Case studies of K_z

We conducted four case studies to explore the significance of K_z. Each case study involved two researchers, namely *R*1 and *R*2. The number of publications was kept constant across all cases, while the focus was on comparing the *h*-index and total citations (*TC*) of two researchers.

1. **Case I - Identical *h*-index and total citations:** Table 1 represents the first case study where we assumed that both researchers *R*1 and *R*2 have the same *h*-index and total citations count. However, despite sharing these characteristics, researcher *R*2 obtained a higher K_z score than *R*1. This difference in K_z scores can be attributed to the impact of the publication year, which played a dominant role in determining the continuous research impact of each researcher. It highlights the significance of considering the temporal aspect of research contributions when assessing the research impact on individuals.
2. **Case II - Identical *h*-index and different total citations:** In this case (Table 2), both researchers *R*1 and *R*2 have an equal number of publications and *h*-index, but they differ in their total citations count. Researcher *R*1 has one highly cited paper, while researcher *R*2 has multiple highly cited papers. Despite *R*1 having a higher total number of citations compared to *R*2, *R*2 obtains a higher K_z score. This indicates that the impact of having multiple highly cited papers outweighs the effect of a single highly cited paper in determining the continuous research impact.
3. **Case III(a) - Different *h*-index and total citations:** In this case (Table 3), both researchers have an equal number of publications but differ in their *h*-index, number of high impact papers, and total citations. Researcher *R*1 has a higher *h*-index but lower total citation count compared to *R*2. However, despite *R*1 having a lower total citation count, they obtain the highest K_z score. This highlights the importance of considering the continuous research impact captured by K_z, which

takes into account not only the number of citations but also the publication age and impact of publications.

4. **Case III(b) - Different h-index and total citations:** In this case (Table 4), we again considered two researchers with an equal number of publications but different h-index, high impact papers, and total citations. Researchers $R1$ had a higher h-index and total citation count compared to researcher $R2$. Surprisingly, despite these differences, it was researcher $R2$ who obtained the highest K_z score. This finding suggests that the K_z score takes into account factors beyond just h-index and total citations, emphasizing the importance of considering the continuous impact and temporal aspects of research contributions.

Table 1. Two researchers with identical h-index and total citations.

Case I	**Researcher *R*1, *h*=4**					**Researcher *R*2, *h*=4**				
S. No	P_y	**C**	$\boldsymbol{k}$	**dt**	$\boldsymbol{k'}$	P_y	**C**	$\boldsymbol{K}$	**dt**	$\boldsymbol{k'}$
1	2014	40	2.292	9	0.255	2014	2	0.43	9	0.048
2	2015	30	2.113	8	0.264	2015	3	0.682	8	0.085
3	2016	0	0	7	0	2016	3	0.682	7	0.098
4	2017	3	0.682	6	0.114	2016	40	2.292	7	0.327
5	2018	24	1.974	5	0.395	2017	1	0	6	0
6	2019	1	0	4	0	2018	30	2.113	5	0.423
7	2020	1	0	3	0	2019	22	1.92	4	0.48
8	2021	1	0	2	0	2020	0	0	3	0
9	2022	0	0	1	0	2021	1	0	2	0
10	2022	10	1.43	1	1.431	2022	8	1.292	1	1.292
	TC = 110, K_z = 2.459					TC = 110, K_z = 2.753				

Table 2. Two researchers with identical h-index and different total citations.

Case II	**Researcher *R*1, *h*=4**					**Researcher *R*2, *h*=4**				
S. No	P_y	**C**	$\boldsymbol{k}$	**dt**	$\boldsymbol{k'}$	P_y	**C**	$\boldsymbol{K}$	**dt**	$\boldsymbol{k'}$
1	2014	1000	4.292	9	0.477	2014	500	3.861	9	0.429
2	2015	4	0.861	8	0.108	2015	300	3.54	8	0.443
3	2016	0	0	7	0	2016	100	2.861	7	0.409
4	2017	4	0.861	6	0.144	2016	0	0	7	0
5	2018	5	1	5	0.2	2017	2	0.43	6	0.072
6	2019	1	0	4	0	2018	50	2.43	5	0.486
7	2020	1	0	3	0	2019	1	0	4	0
8	2021	1	0	2	0	2020	3	0.682	3	0.228
9	2022	0	0	1	0	2021	1	0	2	0
10	2022	0	0	1	0	2022	0	0	1	0
	TC = 1016, K_z = 0.929					TC = 957, K_z = 2.067				

Table 3. Two researchers with different h-index and total citations where R1 has higher h-index and lower total citations than R2.

Case III (a)	Researcher *R1*, *h*=5					Researcher *R2*, *h*=3				
S. No	P_y	**C**	***K***	**dt**	***k′***	P_y	**C**	***k***	**dt**	***k′***
1	2014	90	2.511	9	0.279	2014	250	3.982	9	0.443
2	2015	80	2.445	8	0.306	2015	2	0.5	8	0.063
3	2016	20	1.672	7	0.239	2016	2	0.5	7	0.071
4	2017	3	0.613	6	0.102	2016	82	3.178	7	0.454
5	2018	24	1.773	5	0.355	2017	2	0.5	6	0.083
6	2019	2	0.386	4	0.097	2018	110	3.39	5	0.678
7	2020	3	0.613	3	0.204	2019	1	0	4	0
8	2021	3	0.613	2	0.307	2020	2	0.5	3	0.167
9	2022	2	0.386	1	0.387	2021	2	0.5	2	0.25
10	2022	23	1.75	1	1.75	2022	0	0	1	0
	TC = 250, K_z = 4.026					TC = 453, K_z = 2.209				

Table 4. Two researchers with different h-index and total citations where R1 has higher h-index and total citations than R2.

Case III(b)	Researcher *R1*, *h*=6					Researcher *R2*, *h*=4				
S. No	P_y	**C**	***k***	**dt**	***k′***	P_y	**C**	***k***	**Dt**	***k′***
1	2014	200	2.722	9	0.303	2014	2	0.43	9	0.048
2	2015	150	2.575	8	0.322	2015	2	0.43	8	0.054
3	2016	5	0.827	7	0.118	2016	3	0.682	7	0.098
4	2017	10	1.183	6	0.197	2016	1	0	7	0
5	2018	35	1.827	5	0.365	2017	280	3.501	6	0.584
6	2019	1	0	4	0	2018	2	0.43	5	0.086
7	2020	33	1.796	3	0.599	2019	40	2.292	4	0.573
8	2021	1	0	2	0	2020	70	2.639	3	0.88
9	2022	2	0.356	1	0.356	2021	2	0.43	2	0.215
10	2022	32	1.781	1	1.781	2022	50	2.43	1	2.431
	TC = 469, K_z = 4.041					TC = 452, K_z = 4.969				

Empirical study

To calculate the continuous research impact (K_z) of researchers, the research profiles of 376 individuals affiliated with Monash University, Australia were obtained. Monash University is a public research institution located in Australia, and information about the researchers can be found on their webpage at https://research.monash.edu/ en/persons/. The webpage provides the researcher's research ID and Orcid ID, which facilitated the extraction of their publication details and citations from the Web of Science database. From a pool of 6316 researchers' profiles, we selected 376 profiles across different disciplines, ensuring a range of *h*-index values ($1 \leq h \leq 112$). The choice of databases was made based on data availability. For each researcher ID, information regarding the publication year and the corresponding citations received were extracted. For each researcher, the h-index, and K_z were computed. Additionally, the overall research age or career length of the researcher was determined by subtracting the year of his/her first publication from the current year.

Comparative analysis of K_z

By using equation (3), we calculated the K_z score of 376 researchers. In Figure 1, a scatter plot depicting the relationship between K_z and career length. Each dot on the plot represents an individual researcher. The horizontal dashed line represents the median of the axis, while vertical dashed lines are used to divide the plot into three zones based on the length of the researchers' careers: early career (≤ 10 years), mid-career ($10 < \text{years} \leq 20$) and advanced career ($> 20$ years). This visualization clearly differentiates between the star performer and average performer at different career stages.

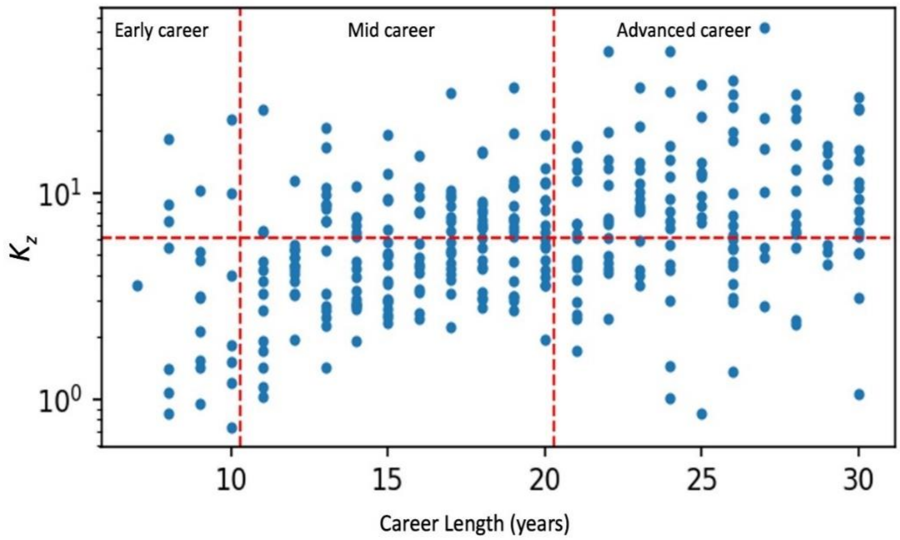

Figure 1. Scattered plot of K_z versus career length. Each dot corresponds to a researcher. The horizontal dashed line represents the median of the axis and vertical dashed lines divides the plots in three zones based on the researcher's career length.

Table 5 elucidates the significance of utilizing K_z over the h-index when researchers share the same h-index. The table presents details on career length (in years), the number of papers, total citations, h-index, and K_z score for a group of researchers who share a common h-index. Specifically, the table includes information for two sets of researchers: one set of researchers with h-index 25, labelled as $R1$-$R8$, and the other set of researchers with h-index 30 labelled as $R9$-$R16$. As highlighted earlier, the K_z score provides valuable differentiation between two researchers with the same h-index based on their continuous research impact. For instance, within Table 5, researchers $R4$ and $R6$ share an h-index of 25 and an identical number of papers (59) with total citations 2982 and 1530 respectively. However, crucially, they do not share

the same K_z score. Notably, despite $R4$ having a higher total citation count than $R6$, the former exhibits a lower K_z score.
Likewise, in the case of researcher $R13$ and $R16$, both share an h-index of 30. While $R13$ boasts a longer career length, a greater number of publications, and higher total citations compared to $R16$, it's noteworthy that $R16$ attains a higher K_z score. This scenario is just one among several instances of researchers depicted in the provided table. The presence of an identical h-index underscores its limitation in distinguishing the top-performing researcher from their peers, while K_z serves as a significant discriminator for identifying impactful researchers. This distinction emphasizes the varying impact levels among researchers. Similarly, in Table 6, profiles of researchers with the same career age are presented, yet their K_z scores differ. Consider researchers $R9$ and $R11$, who share the same career length. However, $R9$ has fewer publications and a higher number of citations and h-index compared to $R11$, resulting in a higher K_z score for $R11$.

Table 5. Comparative analysis among researchers having identical h-index.

S.No	Researcher ID	Career Length (Yrs)	#Papers	Total Citations	*h*-Index	K_z
R1	B-6419-2008	17	44	2415	25	5.24
R2	H-6054-2014	19	38	3433	25	6.76
R3	D-5776-2019	26	68	1984	25	6.828
R4	J-1532-2014	18	59	2982	25	7.896
R5	N-8153-2014	20	78	4217	25	9.156
R6	E-6623-2015	14	59	1530	25	10.618
R7	A-3854-2010	21	86	2034	25	11.224
R8	K-5277-2012	24	73	3783	25	11.912
R9	B-8486-2008	29	79	2851	30	4.487
R10	G-1412-2012	34	69	2816	30	5.517
R11	H-3196-2013	13	94	2538	30	8.684
R12	F-2273-2010	16	102	2627	30	10.446
R13	I-1956-2014	23	123	3797	30	11.05
R14	I-1738-2013	19	105	3306	30	11.309
R15	D-4239-2011	25	133	3343	30	12.475
R16	H-4935-2013	15	100	2945	30	18.97

Table 6. Comparative analysis among researchers having identical career length.

S.No	Researcher ID	Career Length (Yrs)	#Papers	Total Citations	*h*-Index	K_z
R1	K-5514-2018	10	9	32	4	1.043
R2	P-7354-2019	10	8	171	6	1.69
R3	I-9365-2017	10	20	287	10	3.823
R4	G-3877-2013	10	75	1189	18	9.813
R5	L-4481-2018	10	90	6012	28	22.385
R6	N-4364-2019	20	23	757	14	1.905
R7	A-4190-2009	20	32	832	14	3.795
R8	B-7556-2008	20	60	7144	27	6.847
R9	C-9764-2013	20	122	5917	42	10.995
R10	I-1587-2014	20	107	1127	18	12.88
R11	C-4319-2011	20	170	5080	39	19.088
R12	H-9193-2014	30	26	181	8	2.939
R13	P-8366-2016	30	98	5701	40	6.378
R14	B-9553-2008	30	91	6784	45	10.524
R15	H-5706-2014	30	171	4559	35	15.996
R16	A-5452-2008	30	283	26495	89	25.657
R17	I-6251-2012	30	280	58171	68	29.05

Furthermore, upon scrutinizing researchers $R5$ and $R14$ who have distinct career lengths, a noticeable disparity comes to light. Despite $R5$ being a younger researcher with a lower h-index than the more experienced $R14$, their research impact is effectively captured by K_z. Notably, $R5$ possesses a higher K_z score compared to $R14$.Therefore, K_z distinctly identifies impactful researchers, particularly in scenarios where researchers exhibit nearly identical numbers of publications, citations, and h-index.

In Table 7, we explored 11 comparative scenarios involving researchers with identical h-index and career length. One notable case is $S1$, where two researchers share an 8-year career length and an h-index of 12. Despite the similarities, the researcher with a higher total of publications and citations attains a superior K_z score. Conversely, in $S3$, where two researchers have a 13-year career and an h-index of 19, the one with fewer total publications but a higher citation count than the counterpart secures a higher K_z score. On the contrary, in $S7$, with a career length of 17 years and an h-index of 13 for both researchers, the one with more total publications but fewer citations than the other earns a higher K_z score.

This highlights that the K_z metric comprehensively considers relevant research indicators, including total publications, citation count, h-index, and publication age, to capture an individual's continuous impact. It's important to note that a higher K_z score cannot be solely attributed to either more total publications or higher citations count. Furthermore, one cannot conclusively assert that an individual with a higher h-index will always possess a higher K_z score. The K_z metric adopts a holistic approach, simultaneously considering multiple factors in the assessment of research impact.

Table 7. Comparative analysis among researchers having identical research career length (yrs) and h-index.

S.No	Researcher ID	Career Length (Yrs)	#Papers	Total Citations	h-Index	K_z
S1	F-9424-2013	8	37	1595	12	7.041
	O-7942-2018	8	34	454	12	5.291
S2	AAE-7279-2019	12	47	1529	15	11.122
	I-9929-2012	12	37	1236	15	4.321
S3	L-4989-2018	13	84	1875	19	20.182
	M-7607-2014	13	106	1130	19	8.26
S4	E-6431-2011	14	16	508	8	4.057
	N-1676-2017	14	14	726	8	2.771
S5	A-7222-2013	14	28	608	14	6.299
	L-1320-2019	14	23	875	14	3.264
S6	K-7419-2014	15	52	482	11	2.845
	G-1470-2011	15	36	351	11	4.741
S7	O-9174-2014	17	36	708	13	4.444
	J-5651-2016	17	16	857	13	2.173
S8	Q-9068-2018	18	47	2034	21	7.279
	H-4554-2014	18	53	1462	21	8.99
S9	F-6776-2014	18	159	1843	23	15.62
	H-8387-2012	18	78	1798	23	8.635
S10	F-4112-2014	22	18	617	13	2.402
	C-6296-2014	22	35	1456	13	4.842
S11	C-2440-2013	28	38	6087	27	2.401
	N-5018-2017	28	87	2588	27	7.02

Probability distribution of K_z

Figure 2 presents a graphical representation of the plot for $\log(K_z)$, which exhibits a mean value of μ and a standard deviation of σ. This plot is compared to the normal distribution with the same mean and standard deviation. The overlapping nature of the two plots suggests that the variable K_z follows a log-normal distribution.

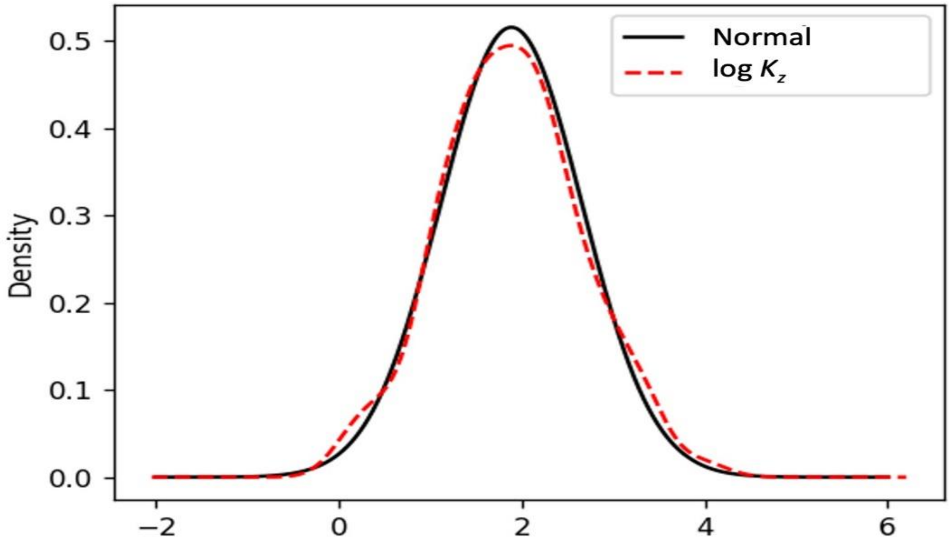

Figure 2. Distribution of $\log(K_z)$ (dashed) versus normal distribution (solid) with same μ and σ.

To confirm this observation, a "Goodness of Fit" test was conducted using the χ^2 distribution. The objective of the Goodness of Fit Test was to assess the suitability of the null hypothesis that states "the distribution of $\log(K_z)$ conforms well to a normal distribution." The test was executed in the following manner:
The logarithm of the values of K_z was computed, and these values were then classified into seven distinct classes, taking into account the mean ($\mu = 0.78787$) and standard deviation ($\sigma = 0.37448$). Subsequently, the observed frequencies (O_i) for each class were determined. To obtain the expected frequencies (E_i), the entire dataset consisting of 376 observations was subjected to calculations based on the normal distribution. The specific calculations and their results are provided in Table 8.

Table 8. Goodness of fit test.

Classes	**Observed Frequencies (O_i)**	**Expected frequencies (E_i) for $\mathcal{N}(\mu, \sigma)$**
$\log(K_z) < \mu - 1.5\sigma$	14	25
$\mu - 1.5\sigma \le \log(K_z) < \mu - \sigma$	34	35
$\mu - \sigma \le \log(K_z) < \mu - 0.5\sigma$	57	56
$\mu - 0.5\sigma \le \log(K_z) < \mu + 0.5\sigma$	157	144
$\mu + 0.5\sigma \le \log(K_z) < \mu + \sigma$	57	56
$\mu + \sigma \le \log(K_z) < \mu + 1.5\sigma$	29	35
$\log(K_z) \ge \mu - 1.5\sigma$	28	25
Total	376	

The χ^2 value was computed using the formula $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ and yielded a value of 7.466. As the calculated χ^2 value is smaller than the critical value $\chi^2{}_{(6,0.05)} = 12.592$, we cannot reject the null hypothesis at a significance level of 0.05. Therefore, we can conclude that $\log(K_z)$ is a suitable fit for the normal distribution.

Analysis of $\boldsymbol{K_z}$ distribution across Physical Science, Agriculture, and Engineering & Technology Domains

To compare the Kz score across different disciplines, authors profiles of 931 researchers from Physical Science, 432 from Agriculture, and 887 from Engineering & Technology domains has been analyzed. The Scopus ID of all authors has been extracted from Indian Scholars profile database, named VIDWAN (https://vidwan.inflibnet.ac.in/). Then the complete profile of authors with their publication details and citations has been extracted from Scopus database. Further, the corresponding $\boldsymbol{K_z}$ values are computed using equation (3). Hence the domain wise distribution of $\log(K_z)$ is plotted and shown in the Figure 3. Form the figure it is observed that $\boldsymbol{K_z}$ follows the log-normal distribution in all the three research domains.

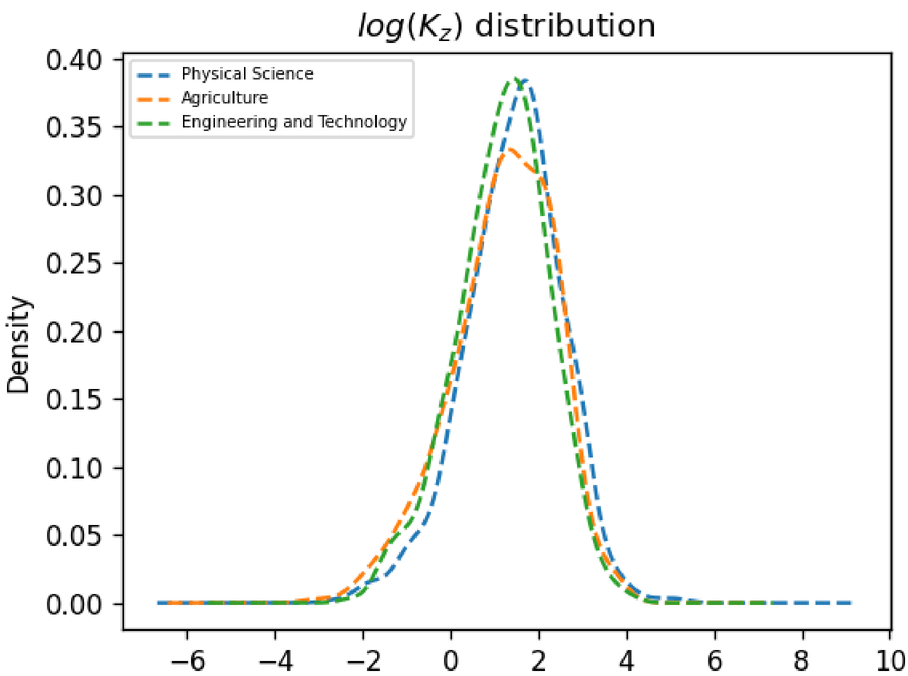

Figure 3. Distribution of $\log(K_z)$ for Physical Science, Agriculture, and Engineering & Technology domains.

It was also observed that the variance of $\log(K_z)$) is consistent across these disciplines. To statistically confirm this, Bartlett's test is applied as follows:

Null Hypothesis: *All variances are equal*

Alternative Hypothesis: *At least one variance is different*

Significance Level: *0.05*

The test resulted in a χ^2 value of 5.77 and a p-value of 0.056, supporting the equality of variances. With equal variances confirmed, ANOVA was used to determine if the mean of $\log(K_z)$ differs among the disciplines. The ANOVA results are shown in Table 9. The p-value being less than the significance level (0.05) indicates that the mean $\log(K_z)$ significantly differs across the three disciplines.

Table 9. Data Summary and ANOVA.

DATA SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Agriculture	432	224.6054	0.51992	0.258682
Engineering	887	453.1929	0.510928	0.212634
Physical Science	931	577.7204	0.620537	0.23162

ANOVA

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	6.232933	2	3.116466	13.58977	1.36E-06	2.99973
Within Groups	515.2919	2247	0.229324			
Total	521.5248	2249				

To further investigate if any two domains have similar mean of $\log(K_z)$ values, Tukey's Honest Significant Difference (HSD) and Fisher's LSD tests were applied. The summaries of these tests are provided in Tables 10-13. Both tests reveal that the mean $\log(K_z)$ is the same for Agriculture and Engineering & Technology domains, but differs for the Physical Science domain.

Tukey Pairwise Comparisons

Table 10. Grouping Information Using the Tukey Method and 95% Confidence.

Factor	**N**	**Mean**	**Grouping**	**Remark**
Physical Science	931	0.6205	A	*Means that do not share a letter are significantly different*
Agriculture	432	0.5199	B	
Engineering & Technology	887	0.5109	B	

Table 11. Tukey Simultaneous Tests for Differences of Means.

Difference of Levels	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
Engineering - Agriculture	-0.0090	0.0281	-0.32	0.945
Physical Sci - Agriculture	0.1006	0.0279	3.61	0.001
Physical Sci - Engineering	0.1096	0.0225	4.88	0.000

Fisher Pairwise Comparisons

Table 12. Grouping Information Using the Fisher LSD Method and 95% Confidence.

Factor	N	Mean	Grouping	Remark
Physical Science	931	0.6205	A	*Means that do not share a letter are significantly different*
Agriculture	432	0.5199	B	
Engineering & Technology	887	0.5109	B	

Table 13. Fisher Individual Tests for Differences of Means.

Difference of Levels	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
Engineering - Agriculture	-0.0090	0.0281	-0.32	0.749
Physical Sci - Agriculture	0.1006	0.0279	3.61	0.000
Physical Sci - Engineering	0.1096	0.0225	4.88	0.000

Identification of top contributors and low contributors

In the case of a normal distribution, the middle 50% of the data is encompassed within a range of +0.67 and -0.67 standard scores from the mean. Consequently, researchers in the top 25% satisfy the condition $K_z \geq e^{(\mu-0.67\sigma)}$, while researchers in the bottom 25% satisfy the condition $K_z \leq e^{(\mu-0.67\sigma)}$. Similarly, using the properties of normal distribution, the α% of top and bottom performers can be identified. Unlike previous indices such as the h, g, e, h_c, etc., the K_z-index allows for the identification of both top and bottom contributors. This categorization based on K_z scores can be beneficial for universities, scientific communities, and research funding agencies in identifying significant contributors.

Discussion and conclusion

In this study, we have discussed various research indicators, including total publications, citations count, h-index, etc., commonly used to measure the impact of research. While total publications, citation count, and h-index are commonly used indicators to assess research impact, they have some limitations when considered individually.

Some of the limitations when considering the research indicators alone are highlighted below:

1. *Total publications*: Relying solely on the number of publications can be misleading, as it does not consider the quality or impact of those publications. Quantity alone does not reflect the significance or influence of a researcher's work.
2. *Citations count*: While citation count is a useful indicator of the influence and visibility of a researcher's work, it can be influenced by factors such as the field of study, publication age, and citation practices within the research community. Additionally, self-citations can artificially inflate citation counts and impact assessments.
3. *h-index:* The *h*-index takes into account both the number of publications and their corresponding citations. However, it does not differentiate between highly cited publications and those with fewer citations. A researcher with a few highly influential papers can have the same h-index as someone with many moderately cited papers. Additionally, h-index ignores all the papers which are cited less than h.
4. *Temporal considerations:* Individual metrics may not capture the continuous progress and development of a researcher's work over time. They provide a snapshot of impact at a specific moment and may not reflect the long-term contributions or evolving research trajectory.

To overcome these limitations and capture the dynamic nature of research impact, it is essential to consider multiple indicators and employ comprehensive assessment approach. We attempted to address above mentioned issues and proposed an index named K_z index, which incorporates various factors to provide a more nuanced understanding of research impact. This study focuses on certain drawbacks of the h-index, particularly its exclusion of papers with citations below the h-index value and those exceeding it. To illustrate, if the h-index is 10, papers with citations below 10 are deemed to have no impact on the author's contribution and are consequently excluded from the h-index calculation. Moreover, whether a paper has 20, 30, or 100 citations, they all contribute equally to the h-index value, which remains fixed at 10. In contrast, our proposed index, K_z, considers all papers regardless of citations being higher or lower than the author's h-index. The paper explicitly delineates scenarios where a high h-index alone may not necessarily indicate an active researcher. Additionally, we take into account the time of publication and the popularity of papers over both long and short periods to gauge the author's contribution to the research community. The distribution of K_z is field independent as well as takes into account the temporal aspect of the work. Unlike other research indicators, K_z takes into account not only the total publications and citations count but the age of the publications too. Our results demonstrate how K_z can effectively differentiate between two potential researchers who may have the same *h*-index, citations count, or career length. By incorporating K_z into the evaluation process, we can better assess the research dynamics of an individual and gain insights into their continuous impact over time.

To conclude, K_z holds the potential to serve as a superior measure for capturing the impact of individuals, institutions, or journals. Its comprehensive consideration of various research indicators (total citations, total publications, *h*-index, etc.) allows a more nuanced assessment of research impact. Further K_z can be utilized as a ranking method to evaluate and rank researchers within an institution based on their research impact. Similarly, institutions and journals can be compared and ranked according to their research impact. This information can be valuable in decision-making processes, as funding agencies, research award committees and hiring bodies can leverage the power of K_z to rank potential candidates within a specific field. It provides a standardized tool to assess and compare the impact of research entities, facilitating more informed decisions and promoting recognition based on research excellence.

There are some challenges and limitations associated while computing the K_z metric too.

1. *Data availability and accuracy:* Different databases may have variations in the coverage of publications and citations, potentially leading to incomplete or inconsistent data. Obtaining accurate and comprehensive data from various sources can be a challenge.
2. *Data quality and reliability:* The accuracy and reliability of the data gathered from different data sources, used for computing K_z are crucial as inaccurate or incomplete data can result in misleading assessments of research impact.
3. *Self-citation manipulation:* The issue of self-citation manipulation, where researchers excessively cite their own work to inflate their impact metrics, can pose a challenge as detecting such manipulations requires careful scrutiny and data filtering techniques.
4. *Special case for citation 0 or 1:* As mentioned earlier, K_z works fine for all the cases of papers with more than 1 citation. As K_z is computing the continuous research impact of an author, therefore the papers with zero and one citation have been considered as having no impact. The proposed index K_z is the summation of all the individual *k* values divided by the time interval, therefore the papers with zero citation or 1 citation do not seem to have much significance in the continuous research impact of a particular author. In previously published studies by Khurana et al. (Khurana and Sharma, 2022), the authors have shown such cases as limiting cases.
5. *Fractional citations:* For multiauthor publication, the proposed index does not provide the fractional weightage to citations (Bi, 2023). At present, each

individual in the multiauthor publication received the full citation while computing the K_z-score.

As discussed, it can be inferred that the K_z index is a comprehensive mathematical function that considers multiple factors to assess the impact of a researcher. These factors include the researcher's total publications, the citation count of each paper, the researcher's h-index, and the age of publication. The K_z index recognizes influential papers which often receive citations at a faster rate, indicating a greater impact, and therefore assigns them higher weight in impact evaluation. By considering these aspects, the K_z index tends to yield higher values in cases where a researcher has made significant contributions that have garnered substantial citations.

Data Availability Statement

The datasets generated during and/or analysed during the current study along with python codes are available from the corresponding author on reasonable request.

Conflict of interest

The author declares no conflict of interest.

References

Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., Herrera, F., 2009. h-index: A review focused in its variants, computation and standardization for different scientific fields. Journal of informetrics 3, 273–289.

Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., Herrera, F., 2010. hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices. Scientometrics 82, 391–400.

Batista, P.D., Campiteli, M.G., Kinouchi, O., 2006. Is it possible to compare researchers with different scientific interests? Scientometrics 68, 179–189.

Bi, H.H., 2023. Four problems of the h-index for assessing the research productivity and impact of individual authors. Scientometrics 128, 2677– 2691.

Bornmann, L., Daniel, H.D., 2005. Does the h-index for ranking of scientists really work? Scientometrics 65, 391–392.

Bornmann, L., Daniel, H.D., 2009. The state of h index research: is the h index the ideal way to measure research performance? EMBO reports 10, 2–6.

Costas, R., Bordons, M., 2007. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. Journal of informetrics 1, 193–203.

Ding, J., Liu, C., Kandonga, G.A., 2020. Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. Scientometrics 122, 1303–1322.

Egghe, L., 2006. Theory and practise of the g-index. Scientometrics 69, 131–152.

Egghe, L., 2010. The hirsch index and related impact measures. Annual review of information science and technology 44, 65–114. Egghe, L., 2011. A disadvantage of h-type indices for comparing the citation impact of two researchers. Research Evaluation 20, 341–346.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proceedings of the National academy of sciences 102, 16569– 16572.

Hirsch, J.E., 2019. h α: An index to quantify an individual's scientific leadership. Scientometrics 118, 673–686.
Jin, B., Liang, L., Rousseau, R., Egghe, L., 2007. The r-and ar-indices: Complementing the h-index. Chinese science bulletin 52, 855–863.
Khurana, P., Sharma, K., 2022. Impact of h-index on author's rankings: an improvement to the h-index for lower-ranked authors. Scientometrics 127, 4483–4498.
Liu, Y., Yang, Y., 2014. Empirical study of l-sequence: The basic h-index sequence for cumulative publications with consideration of the yearly citation performance. Journal of Informetrics 8, 478–485.
Penfield, T., Baker, M.J., Scoble, R., Wykes, M.C., 2014. Assessment, evaluations, and definitions of research impact: A review. Research evaluation 23, 21–32.
Prathap, G., 2010. The 100 most prolific economists using the p-index. Scientometrics 84, 167–172.
Reed, M., Ferre, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M., Holden, J., 2021. Evaluating impact from research: A methodological framework. Research Policy 50, 104147.
Rons, N., Amez, L., 2009. Impact vitality: an indicator based on citing publications in search of excellent scientists. Research Evaluation 18, 233–241.
de Saint-Georges, M., de la Potterie, B.v.P., 2013. A quality index for patent systems. Research Policy 42, 704–719.
Schreiber, M., 2008a. A modification of the h-index: The hm-index accounts for multi-authored manuscripts. Journal of Informetrics 2, 211–216. Schreiber, M., 2008b. To share the fame in a fair way, hm modifies h for multi-authored manuscripts. New Journal of Physics 10, 040201.
Schubert, A., Schubert, G., 2019. All along the h-index-related literature: a guided tour, in: Springer handbook of science and technology indicators. Springer, pp. 301–334.
Todeschini, R., Baccini, A., 2016. Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research. John Wiley & Sons.
Van Leeuwen, T., 2008. Testing the validity of the hirsch-index for research assessment purposes. Research Evaluation 17, 157–160. Zhang, C.T., 2009. The e-index, complementing the h-index for excess citations. PLoS One 4, e5429.
Zhang, C.T., 2013. The h'-index, effectively improving the h-index based on the citation distribution. PloS one 8, e59912.